

# Cryptic speciation shapes the biogeographic history of a northern distributed moss

DENNIS ALEJANDRO ESCOLÁSTICO-ORTIZ<sup>1,2,3,\*</sup>, LARS HEDENÄS<sup>4</sup>, DIETMAR QUANDT<sup>5,6</sup>, DÖRTE HARPKE<sup>5,6</sup>, JUAN LARRAÍN<sup>7</sup>, MICHAEL STECH<sup>8</sup> and JUAN CARLOS VILLARREAL A.<sup>1,2,3</sup>

<sup>1</sup>Département de Biologie, Université Laval, Québec, G1V 0A6, Canada

<sup>2</sup>Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, G1V 0A6, Canada

<sup>3</sup>Centre d'études nordiques (CEN), Université Laval, Québec, QC, Canada

<sup>4</sup>Department of Botany, Swedish Museum of Natural History, Stockholm, Sweden

<sup>5</sup>Nees Institute for Biodiversity of Plants, University of Bonn, Bonn, Germany

<sup>6</sup>Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

<sup>7</sup>Centro de Investigación en Recursos Naturales y Sustentabilidad (CIRENYS), Universidad Bernardo O'Higgins, Avenida Viel 1497, Santiago, Chile

<sup>8</sup>Naturalis Biodiversity Center, Section National Herbarium of the Netherlands, Leiden University, Leiden, The Netherlands

Received 5 October 2021; revised 23 March 2022; accepted for publication 27 April 2022

Increasing evidence indicates that wide distributed bryophyte taxa with homogeneous morphology may represent separate evolutionary lineages. The evolutionary histories of these cryptic lineages may be related to historical factors, such as the climatic oscillations in the Quaternary. Thus, the post-glacial demographic signatures paired with cryptic speciation may result in complex phylogeographic patterns. This research has two aims: to determine whether the widespread moss *Racomitrium lanuginosum* represents cryptic molecular taxa across the Northern Hemisphere and to infer the effects of Quaternary glaciations on spatial genetic diversity. We used the internal transcribed spacer (ITS) marker to resolve the phylogeographic history of the species and single nucleotide polymorphisms (genotyping-by-sequencing) to infer the genetic structure and demographic history. Finally, we assessed the historical changes in the distribution range using species distribution models. *Racomitrium lanuginosum* comprises distinct molecular lineages sympatrically distributed in the Northern Hemisphere. We also uncovered long-distance dispersal from eastern North America to Scandinavia and potential *in situ* survival in northern Scandinavia. Due to the genetic signatures, the Alaska Peninsula could be considered a glacial refugium. The species experienced post-glacial expansion northwards in the Northern Hemisphere, mainly from the Alaska Peninsula. Our results exemplify the complex phylogeographic history in cold environments and contribute to recognizing evolutionary patterns in the Northern Hemisphere.

**ADDITIONAL KEYWORDS:** biogeography – bryophytes – dispersal – ecology – genetics – glacial refugia – molecular evolution – phylogeography – systematics.

## INTRODUCTION

Wide spatial distribution ranges spanning different continents are characteristic of spore-dispersed organisms, such as plants and lichens (e.g. [Alors et al., 2017](#); [Hernández-Rojas et al., 2020](#)). Among plants, bryophytes exhibit worldwide and disjunct

distributions as evidence of their vagility ([McDaniel & Shaw, 2005](#); [Pisa et al., 2014](#); [Kyrkjeeide et al., 2016a](#); [Biersma et al., 2017](#); [Patiño & Vanderpoorten, 2018](#); [Vanderpoorten et al., 2019](#)). The extensive distribution range of some bryophyte species [e.g. *Racomitrium lanuginosum* (Hedw.) Brid., *Sphagnum magellanicum* Brid.] has led to reconsideration of whether they represent single taxonomic entities or species complexes. Integrative approaches using

\*Corresponding author. E-mail: [escolasticodennis@gmail.com](mailto:escolasticodennis@gmail.com)

morphological, molecular and ecological data have recognized distinct bryophyte species in taxa with wide distribution ranges (Medina *et al.*, 2013; Patiño *et al.*, 2017; Köckinger & Hedenäs, 2017; Sim-Sim *et al.*, 2017; Renner, Heslewood & Heinrichs, 2018; Vigalondo *et al.*, 2019; Hedenäs, 2020b). However, in some cases, bryophyte disjunct populations have little morphological variation and lack distinctive traits for species recognition, often resulting in the circumscription of one single taxon (Shaw, 2001; Renner, 2020). In a general sense, cryptic taxa are genetically differentiated lineages that do not exhibit distinctive phenotypic traits (Struck *et al.*, 2018). Several cryptic taxa have been identified in widely distributed bryophytes using data from molecular markers (McDaniel & Shaw, 2003; Hedenäs, 2008, 2010; Fuselier *et al.*, 2009; Kyrkjeeide *et al.*, 2016b; Lewis *et al.*, 2017; Alonso-García *et al.*, 2020; Šlipiko *et al.*, 2020). Consequently, species once thought of as one taxon have been revealed to be groups of well-differentiated molecular lineages.

Bryophyte cryptic taxa may occur allopatrically at a continental or regional scale, as exemplified by the genetic differentiation observed in Northern Hemisphere populations from the Amphi-Atlantic zone and Beringia (Hedenäs, 2012; Kyrkjeeide *et al.*, 2016b). In other cases, distinct lineages have sympatric distributions (Piñeiro *et al.*, 2012; Kyrkjeeide *et al.*, 2016a). The latter situation may be more complex to describe because those lineages may have diverged sympatrically or allopatrically and then dispersed into other areas producing sympatric distributions. These patterns are strongly related to the historical factors that influenced the genetic divergence of cryptic taxa. In this regard, the climatic oscillations during the Quaternary are recognized to have affected the evolutionary and demographic history of organisms, notably for plants with northern distributions (Hewitt, 2004). Specifically, during the Last Glacial Maximum (LGM; 26.5–19 kya; Clark *et al.*, 2009), the advance of ice sheets in North America and Europe restrained species distribution areas, causing population bottlenecks, local extinctions and isolation of populations (Abbott *et al.*, 2000; Abbott & Brochmann, 2003; Westergaard *et al.*, 2019). Like other plants, bryophyte populations survived in ice-free areas, *in situ* in microrefugia or under the ice sheet, and after the retreat of glaciers, species dispersed to deglaciated areas (Kyrkjeeide *et al.*, 2014; Désamoré *et al.*, 2016; Ledent *et al.*, 2019). Hence, the demographic signatures of glaciations combined with cryptic speciation may result in intricate phylogenetic relationships (e.g. hybridization, admixture) that obscure the biogeographic history of northern cryptic taxa (Szövényi *et al.*, 2008; Piñeiro *et al.*, 2012; Stubbs *et al.*, 2020).

Among bryophytes, the moss genus *Racomitrium* Brid. (*s.l.*) is widely distributed in both hemispheres and displays a complex phylogenetic history translated into many taxonomic generic changes (Ochyra & Bednarek-Ochyra, 2007; Larrain *et al.*, 2013). In the genus, the species *R. lanuginosum* is distributed only in the Northern Hemisphere and is an emblematic plant of Arctic, sub-Arctic and alpine environments. This moss reproduces mainly by clonal growth forming extensive carpets and contributes to the nitrogen budget of the tundra due to its epiphytic N<sub>2</sub>-fixing bacteria (Tallis, 1958; Baddeley, Thompson & Lee, 1994; Klarenberg *et al.*, 2021). Early phytogeographic studies on *R. lanuginosum* demonstrated slight morphological variation among populations in the Northern Hemisphere (Tallis, 1958; Vitt & Marsh, 1988; Ellis & Tallis, 2003). Recent molecular studies coupled with morphological examination in Scandinavian populations suggested the presence of three well-differentiated molecular lineages based on the nuclear ribosomal ITS and the plastid genes *rpl16* and *trnG* (Hedenäs, 2019, 2020a). These lineages are genetically differentiated but lack distinctive morphological traits supporting molecular results. On a global scale, it is unclear whether those cryptic taxa also occur across the entire distribution of *R. lanuginosum* or if this is a phenomenon specific to the Scandinavian peninsula.

Spatial genetic structure in bryophytes has been commonly investigated using isozymes (e.g. Boisselier-Dubayle *et al.*, 1995; Cronberg, Molau & Sonesson, 1997; Werner & Guerra, 2004), DNA fingerprinting (e.g. Spagnuolo, Terracciano & Giordano, 2009; Hutsemékers *et al.*, 2010; Hassel, Gunnarsson & Gunnarsson, 2011; Korpelainen *et al.*, 2013; Hutsemékers, Hardy & Vanderpoorten, 2013; Shaw *et al.*, 2014) and a few Sanger-generated markers (e.g. Grundmann *et al.*, 2008; Hedenäs, 2010, 2020a; Stech & Quandt, 2014; Biersma *et al.*, 2020). Nonetheless, those approaches may fail to resolve species complexes, recent expansions or shallow relationships due to the limited data (Piñeiro *et al.*, 2012; Biersma *et al.*, 2018). High-throughput sequencing methods, such as genotyping-by-sequencing (GBS), outperform Sanger-generated sequences by producing thousands of loci in non-model species (e.g. Westergaard *et al.*, 2019; Pérez-Escobar *et al.*, 2020; Alonso-García *et al.*, 2021), rendering GBS and similar techniques more suitable for elucidating evolutionary histories. GBS is an emergent method applied in a few phylogeographical studies of bryophytes (Baughman *et al.*, 2017; Lewis *et al.*, 2017; Alonso-García *et al.*, 2020; Ledent *et al.*, 2020). In addition, molecular approaches coupled with past species distribution models could help to shed light on the historical events that influenced the geographical distribution of cryptic taxa (Gavin *et al.*, 2014).

This research aims to determine whether *R. lanuginosum* comprises differentiated molecular taxa across its distribution range in the Northern Hemisphere. We tested whether populations inhabiting previously glaciated and ice-free areas during the LGM differed in patterns of genetic diversity. We used two molecular approaches: a conserved nuclear marker (ITS) to reconstruct the evolutionary timeframe and ancestral areas of the species; and the GBS method to infer the genetic structure and estimate the demographic history. Additionally, we complemented the demographic inference by estimating changes in the species distribution range through three periods: present, LGM, and Last Interglacial (LIG; c. 110 kya).

## MATERIAL AND METHODS

### SAMPLING AND DNA EXTRACTION

*Racomitrium lanuginosum* specimens were obtained from herbaria ALA, CANM, FLAS, L, QFA, TRH, UBC and VT. In addition, fresh material was collected from Kuujuarapik and Umiujaq in Hudson Bay, eastern Canada, in 2017 and 2019. Specimens were deposited in the QFA herbarium following the first author's collection numbers and identified by the authors using regional keys. Samples from Central and Southern Europe were not available for the study to assess the genetic diversity and presence of glacial refugia in those regions. We included some samples of the congeners *R. pruinosum* (Wilson) Müll.Hal. and *R. geronticum* Müll.Hal. A single shoot (gametophyte) per specimen was ground with liquid nitrogen resulting in a single genet per sample. DNA extraction was performed using the CTAB method (Murray & Thompson, 1980). DNA was used for both PCR and GBS but used different samples for each approach (see each type of analysis).

### SANGER-BASED SEQUENCING AND PHYLOGEOGRAPHIC ANALYSES

The ITS region was amplified using primers 18SF and 26SR (Rydin, Pedersen & Friis, 2004). The primer sequences are 5'-GCT TGT CTC AAA GAT TAA GCC-3' for 18SF and 5'-ACT TCC ATG ACC ACC GTC CT-3' for 26SR. PCRs were performed using a total volume of 25 µL: 0.1 µL TopTaq polymerase (Qiagen), 2.5 µL TopTaq buffer (Qiagen), 0.5 µL dNTP mix (10 mM, GeneDireX, Inc.), 1 µL of bovine serum albumin (New England BioLabs), 1 µL of each primer (2.5 mM), 17.9 µL of ultra-pure distilled water (Invitrogen) and 1 µL DNA.

ITS sequences were manually edited using Geneious v.2020.1.2 (Kearse *et al.*, 2012). We produced 61 sequences mainly from North American specimens. Additionally, 64 sequences from Scandinavian

populations of *R. lanuginosum* (Hedenäs, 2020a) and 20 sequences from different locations in Europe (Stech & Larraín, unpubl. data) were added to the dataset (see also [Supporting Information, Table S1](#)). Two sequences of the closely related species *R. pruinosum* were included as an outgroup following Larraín *et al.* (2013). One hundred and forty-seven sequences were aligned using Clustal Omega v.1.2.2. (Sievers & Higgins, 2014). We evaluated the best nucleotide substitution model for the alignment using the phangorn v.2.7.1 library (Schliep, 2011) in R v.3.6.3 (R Core Team, 2017). We selected the GTRGAMMA+I model with the lowest value of the Akaike information criterion corresponding to 4137.920.

To estimate divergence times of *R. lanuginosum*, we conducted Bayesian analyses using BEAST v.1.10.4 and its complementary software (BEAUTi, Tracer & TreeAnnotator) (Drummond & Rambaut, 2007) on the CIPRES Science Gateway 3.3 (Miller, Pfeiffe & Schwartz, 2010). We tested both the strict and the uncorrelated log-normal relaxed molecular clocks. Due to the absence of known *R. lanuginosum* fossils, we applied a general nucleotide substitution rate estimated for mosses (Laenen *et al.*, 2014) with a mean of  $4.45 \times 10^{-4}$  substitutions/site/million years and a standard deviation of  $1.77 \times 10^{-6}$ . We assessed convergence in Tracer and according to the ESS values. Convergence was achieved after  $5 \times 10^8$  generations. Thus, we performed analyses with  $5 \times 10^8$  generations, sampling every 25 000 generations and discarding 10% in each run. We followed Zaccara *et al.* (2020) to evaluate the best tree prior. We performed analyses under the strict and uncorrelated log-normal molecular clocks using Yule, birth-death, coalescent constant size and Bayesian Skyline tree priors. To evaluate the best model, we estimated the marginal likelihood estimate of each model through stepping-stone sampling and path sampling with 150 path steps and one million iterations. According to Bayes factors, the model that better explained the data was the strict clock with a coalescent constant size prior tree (see also [Supporting Information, Table S2](#)). We performed two independent runs using the best model. We corroborated the BEAST phylogenetic tree using maximum likelihood. We estimated phylogenetic relationships using RAxML v.8.2.9 (Stamatakis, 2014) under the GTRGAMMA + I model with 100 bootstrap replicates. All phylogenetic trees were generated in FigTree v.1.4.4 (Rambaut, 2018).

Ancestral area reconstruction was conducted in RASP v1.1.7 (Yu *et al.*, 2015) using the trees produced in BEAST. We classified the samples into Alaska, eastern North America, Europe and Asia to infer the ancestral areas of each clade (see [Supporting Information, Table S1](#)). Then, we removed the outgroup and compared the fit of six different models to our

data using BioGeoBEARS: DEC, DEC+J, DIVALIKE, DIVALIKE+J, BAYEAREALIKE, BAYEAREALIKE+J. We selected the DIVALIKE+J model based on the Akaike information criterion corrected for small samples.

Geographical groups were defined according to the spatial distribution of the samples to compute population genetic statistics. We randomly selected eight samples per geographical group to avoid bias due to different sample sizes. For each group, we calculated the number of polymorphic sites and haplotypes, haplotype diversity, nucleotide diversity and demographic statistics, including Fu & Li's  $D$  and Tajima's  $D$  in DnaSP software v.6.12.03 (Rozas *et al.*, 2017). Finally, we estimated the same summary statistics for each group, but using all samples (different sample sizes per group).

#### GBS DATA PROCESSING

##### *Library preparation and sequencing*

Library preparation was carried out using a GBS double-digested approach. We used the cutting enzymes *Pst*I and *Msp*I to improve the number of restriction sites and fragment size to identify single nucleotide polymorphisms (SNPs) (Sonah *et al.*, 2013). We used a volume of 10  $\mu$ L with DNA concentrations of *c.* 20 ng/ $\mu$ L per sample. Libraries were sequenced in an Ion Torrent system producing single-end sequences at the Genomic Analysis Platform of the Institute of Integrative Biology and Systems of Laval University (Quebec, Canada). We sequenced 168 samples in two plates. The first plate produced 67 512 653 reads and the second one yielded 151 484 471 reads. Additionally, we sequenced 16 samples of *R. lanuginosum* from northern Europe (Svalbard and Jan Mayen) and the congeneric *R. geronticum*. We used the same cutting enzymes, but sequencing was performed on an Illumina HiSeq 2000 system producing single-end reads at the Leibniz Institute of Plant Genetics and Crop Plant Research IPK, Gatersleben (Gatersleben, Germany). This run produced 16 210 221 reads. One hundred and eighty-four samples were processed in this study.

##### *Raw read processing*

Per base sequence quality of raw reads was checked using FastQC v.0.11.8 (Andrews, 2010). We standardized reads to a length of 90 bp based on read length distribution. Next, sequences were cleaned, trimmed and demultiplexed using sample-specific barcodes in the *process\_radtags* program of Stacks v.2.3 (Catchen *et al.*, 2013). A second quality check was performed on demultiplexed sequences to validate the processed data. We applied an alternative approach

using Trimmomatic v.0.36 (Bolger, Lohse & Usadel, 2014) to cut and clean reads before *process\_radtags*. However, the number of processed reads was lower than the standard processed reads; hence, we did not proceed with further analyses of this dataset. We used three different approaches to construct loci: *de novo*; reference-based using BWA; and Bowtie2 aligners.

##### *De novo assembly*

To assemble loci *de novo*, we performed tests in *denovo\_map.pl* program of Stacks to determine the optimized parameter combination. We selected 12 samples that best represented read coverage and phylogenetic groups based on preliminary ITS analyses (see also Supporting Information, Table S3).

We applied the *-r80* loci method proposed by Paris, Stevens & Catchen (2017) to select the parameter combination that maximizes the number of SNPs shared by at least 80% of individuals in a population (*-r* 80). This method consists of varying values for the Stacks parameters *-m*, *-n* and *-M*. First, we varied *-m*, a parameter that constructs stacks on the basis of identical matching reads. We then tested the values *-m* ranged from  $m = 1$  to  $m = 6$  with the rest of the main parameters held constant ( $M = 3$  and  $n = 1$ ). We proceeded to test *-M*, which represents the minimum number of different nucleotides allowed to combine stacks into a locus. We iterated *-M* from  $M = 1$  to  $M = 4$  (constant values  $m = 2$  and  $n = 1$ ). The last parameter was *-n*, which defines the number of mismatches between each sample and the locus catalogue to merge them into a single locus (i.e. fixed alleles in a population representing an alternative allele compared to other populations). This parameter was tested from  $n = 1$  to  $n = 6$  (constant values  $m = 2$ ,  $M = 2$ ). We examined the number of loci and SNPs recovered in each iteration for each parameter. To define a threshold for parameter selection, we only considered those values in which the number of loci and SNPs reach a relatively constant increase at *-R* 80. We tested two candidate parameter combinations  $m = 3$ ,  $M = 1$  and  $n = 3$  and  $m = 2$ ,  $M = 1$  and  $n = 3$ . The final parameter candidates consisted of a trade-off between sequence coverage and number of loci and SNPs. We finally selected  $m = 3$ ,  $M = 1$  and  $n = 3$  as the optimized parameter combination to build loci *de novo*.

*De novo* loci were built using Stacks. For the first step, *ustacks*, we input *-m* 3 and *-M* 1. For *cstacks* we used *-n* 3 and for *gstacks* we created a catalogue using 117 samples out of 184 to discard uninformative loci (loci present in a few samples). The rest of the pipeline components were run with default settings using a population map where each sample represented a population (sample-population map).

### Mapping to a reference

In the absence of an available genome of *R. lanuginosum*, we selected the transcriptomes of *R. elongatum* Ehrh. ex Frisvoll (ID: ABCD) and *R. varium* (Mitt.) A.Jaeger (ID: RDOO) from the 1000 plants initiative (Carpenter *et al.*, 2019; Leebens-Mack *et al.*, 2019) to align the reads. We merged the two transcriptomes into one single reference file to improve the number of aligned reads. Individual output files produced by `process_radtags` were aligned to the reference file using two aligners, BWA v.0.7.17 (Li & Durbin, 2009) and Bowtie2 v.2.3.4.1 (Langmead & Salzberg, 2012). The `mem` command and default options were employed for BWA, which conducts a local alignment starting with exact local matches. The second alignment was performed on Bowtie2 with the `--very-sensitive` option, and one mismatched allowed in seed alignment ( $-N 1$ ). To have compatible input files for Stacks, we transformed the alignment from SAM to BAM format using Samtools v.1.8 (Li *et al.*, 2009). A sample-population map was used as input for the subsequent pipeline steps. We ran the `gstacks` program with default settings to build loci and identify SNPs for both approaches.

### Genotype calling for de novo and reference-based data

For each approach (de novo, reference-based BWA and Bowtie2), we ran populations program using a sample-population map. A first genotype calling was done without a filter in the number of loci shared among samples (using `-R` instead of `-r` because each sample represented a population). Then, the number of generated loci and the percentage of missing data for each sample were recorded. Missing data were evaluated in `adegenet` v.2.0.2 package (Jombart, 2008) in R, transforming `vcf` files into `genind` and `genlight` formats. We filtered samples with high levels of missing data (> 95%) to create a preliminary good-quality dataset consisting of 130 samples. Next, the Stacks populations program was run for each approach, keeping loci shared by at least 80% of samples (`-R 80`) with no heterozygotes due to the haploid condition of moss gametophytes (`--max-obs-het 0`) and a minimum allele frequency of 0.05 (`--min-maf`); however, we did not find SNPs for the three approaches using this stringent filter. Then, we changed the filter to `-R 40`, resulting in different amounts of loci, SNPs and missing data per sample. This information was used to compare the three approaches. The *de novo* approach yielded a lower number of SNPs than the reference-based approaches, and the BWA approach recovered the highest number of SNPs (Table 1). Therefore, BWA approach was selected for further analysis.

We assessed the effect of `-R 40` and `-R 20` filters for the BWA approach. Filtered datasets represented a trade-off between the number of sites (loci and SNPs) and the quality of the data (missing data per sample). The `-R 40` dataset contained fewer SNPs (411) of high quality (< 60% missing data), and `-R 20` produced more SNPs (844) but more missing data (> 60%) (for detailed BWA results see also Supporting Information, Table S4). It has been shown that SNPs with high-missing data (> 80%) yield better-supported phylogenetic trees than few good-quality SNPs (e.g. Tripp *et al.*, 2017). Hence, we selected the `-R 20` dataset to conduct the analyses. Finally, we constrained the dataset to a group of 127 samples, including the congeneric *R. pruinosum* and, for some analyses, *R. geronticum*. None of the 16 Illumina HiSeq sequenced samples passed the filters and were not included in the analyses (see also Supporting Information, Table S4). Different output files were produced in populations program for subsequent analysis. The `-R 20` dataset was mainly used to investigate the phylogeography of *R. lanuginosum*.

## GBS ANALYSES

### Phylogenetic analyses

To estimate maximum-likelihood trees, we used `-R 40` and `-R 20` datasets employing unlinked SNPs (`-write-single-snp`) in populations program of Stacks (i.e. one SNP per locus). Unlinked SNPs have been used in phylogenetic inference because they reduce linked loci accounting for independent evolutionary histories (Grewe *et al.*, 2017, 2018; Alonso-García *et al.*, 2020, 2021). The unlinked-SNP datasets resulted in 324 and 608 SNPs for `-R 40` and `-R 20`, respectively. Maximum-likelihood trees were inferred using RAxML. Runs were performed using the GTR +  $\Gamma$  model with 100 bootstrap replicates and the ascertainment bias option (`--asc-corr = lewis`) that accounts for the variability of all positions. Phylogenetic trees were rooted with *R. pruinosum*.

**Table 1.** Summary of the number of SNPs recovered in *Racomitrium lanuginosum* using two population filters (`-R`) for three different approaches: *de novo*, reference-based using BWA; and Bowtie2 aligners. The BWA approach produced more SNPs using both population filters

Approach	Number of SNPs	
	<code>-R 40</code>	<code>-R 20</code>
<i>de novo</i>	23	757
BWA	465	844
Bowtie2	333	786

We only focused on the *-R 20* dataset for the following analyses due to there being more recovered SNPs. First, a phylogenetic network was constructed to assess the relationships of individuals under a framework that consider events that may lead to reticulate relationships (e.g. admixture and hybridization). We used a matrix of 4620 loci shared by at least 20% of the 126 samples (*-R 20*). We built the network in SplitsTree v.4.16.1 (Huson & Bryant, 2006) using the NeighborNet method with uncorrected *P* distances.

To estimate a coalescent-based tree that accounts for loci incongruence, we used SVDquartets v.1.0 (Chifman & Kubatko, 2014). This software handles multilocus data to infer species and lineage level phylogenetic trees under a coalescent model and performs well when using loci composed of a small number of sites (Chou et al., 2015). We used the same matrix recovered from the *-R 20* dataset to conduct this analysis. We inferred a lineage tree using SVDquartets in PAUP v.4.0a (Swofford, 2003) under the multispecies coalescent model without a partition of samples and evaluating all possible quartets with 100 bootstraps replicates.

### Genetic structure analyses

To detect genetic structure in *R. lanuginosum* we used principal component analyses (PCA) and a haplotype co-ancestry matrix. We included the congeneric *R. geronticum* only for these analyses. Dimensionality-reduction methods have been applied to investigate population structure and have proved to be adequate for datasets with high levels of missing data as a complementary approach (Grewé et al., 2017; Alonso-García et al., 2021). We conducted preliminary PCA using both *-R 20* and *-R 40* datasets, and the results were similar; thus, we only kept the *-R 20* dataset due to the better resolution of genetic groups. We extracted SNPs from loci shared by at least 20% of 127 samples (*-R 20*), resulting in a vcf file with 844 SNPs. The vcf file was transformed into genind and genlight formats to perform PCA using adegenet.

A haplotype co-ancestry matrix of the nearest neighbour was constructed with the software fineRADstructure v.1.0 (Malinsky et al., 2018). Suitable haplotype input data (file.fineRADpainter) were generated in populations program of Stacks from the same *-R 20* dataset as in the PCA. We reordered loci according to linkage disequilibrium using the *sampleLD.R* script. The co-ancestry matrix was estimated and visualized in the finestructureGUI, and samples were grouped according to co-ancestry levels. Each sample was labelled according to the genetic group membership resulting from phylogenetic analyses to compare clustering methods.

Genetic diversity based on recovered loci was estimated for the four genetic groups and ten geographic groups. Geographical groups were delimited according to the co-ancestry levels and sample geographical distribution. Summary statistics such as nucleotide diversity, haplotype diversity, private alleles and pairwise fixation index ( $F_{ST}$ ) among groups were calculated in populations program from Stacks (*-r 60, --max-obs-het 0, --min-maf 0.05*) using population maps based on genetic and geographical groups (see also Supporting Information, Table S4). We applied a sample-size correction using eight samples per group to perform these estimations.

We inferred the relationships of geographical groups and evaluated the presence of migration among populations using allele frequencies in TreeMix v.1.13 (Pickrell & Pritchard, 2012). This analysis estimates population relationships based on allele frequencies (SNPs). The software assumes that SNPs are unlinked or account for linkage disequilibrium by defining blocks of SNPs (*-k*) for resampling. The results are displayed as a maximum-likelihood graph of population splits where branch length represents the amount of genetic drift among populations. Then, the covariance matrix is analysed to select pairs of populations that do not fit well with the model, and the software tries to include migration edges that improve the fitting of the model (Pickrell & Pritchard, 2012). TreeMix performs better with low missing data per population; thus, a more stringent population filter (*-r 60*) and a popmap based on geographic groups were used to generate the input file in Stacks. The *-r 60* and unlinked SNPs (*-write-single-snps*) filters applied on 124 samples of *R. lanuginosum* resulted in 406 SNPs. We selected the Alaska Peninsula population to root the graph (*-root*) based on co-ancestry levels and  $F_{ST}$  group pairwise comparisons. We iterated *-m* from 0 to 10 with ten bootstrap replicates, each with a constant window size of *-k 10* and the *-noss* option to turn off sampling correction. We estimated the optimized number of migrations among populations by analysing the results of the iterations. We used OptM v.0.1.5, which infers the most likely value for *-m* using the TreeMix output (Fitak, 2021). The estimation is based on the second-order rate of change ( $\Delta m$ ) of the composite likelihood  $L(m)$ , which is similar to the  $\Delta K$  method used for detecting the number of genetic clusters in Structure software (Evanno, Regnaut & Goudet, 2005). The results of OptM suggested that one migration edge optimize the fit of the graph model; however, this value did not reach the 99.8% variance explained by the model considered to confidently infer migration edges (Pickrell & Pritchard, 2012; Fitak, 2021) (see also Supporting Information, Fig. S1).

## SPECIES DISTRIBUTION MODELLING

*Data compilation*

To assess range distribution shifts of *R. lanuginosum* during the Quaternary glaciations, we modelled its current, LGM and LIG distribution ranges. Occurrence records of *R. lanuginosum* were downloaded from the Global Biodiversity Information Facility (GBIF) database and complemented with samples used in this study. A total of 28 205 records were obtained from the GBIF portal. Raw data were manually filtered in the following order: (1) removal of occurrence records from the Southern Hemisphere currently recognized as *R. geronticum* and *R. pruinosum* (26 359); (2) occurrence records acquired from human observations, fossils, unknown sources, material samples and literature were excluded (7855); (3) occurrence records identified as species synonyms were kept (7817) and (4) occurrence records without coordinates were excluded (7797).

Manual filtered data consisting of 7797 records were cleaned using CoordinateCleaner v.2.0 library (Zizka *et al.*, 2019) in R. We first removed records with no coordinates (including 0,0). Then, default tests were implemented to clean the data ('capitals', 'centroids', 'equal', 'gbif', 'institutions', 'zeros', 'countries', 'seas'). A second cleaning was conducted to discard records labelled as temporal outliers, with low coordinate precision, outside the distribution range (> 30 decimal latitude), older than 1945 and duplicates. A cleaned data set consisting of 2421 occurrences was generated (see also [Supporting Information, Fig. S2](#)).

Climatic variables were obtained from the PaleoClim web-site (Brown *et al.*, 2018). We downloaded 19 standard bioclimatic variables in a spatial resolution of 10 arc-minutes (c. 20 km) for all periods. Present bioclimatic data was based on Karger *et al.* (2017). Data for the LGM followed Karger *et al.* (2021) from the CHELSA-TraCE21k algorithm in 100-year time steps from the last 21 kya. Finally, we used the LIG bioclimatic variables according to Otto-Bliesner *et al.* (2006) and based on climate simulations starting from c. 130 kya.

*Distribution modelling*

SDMtoolbox v.2.4 (Brown, 2014) was used to prepare the input data for distribution modelling. First, we created a layer containing only the known *R. lanuginosum* distribution range in the Northern Hemisphere to delimitate the area for model prediction. According to Collart *et al.* (2021), the distribution model for *R. lanuginosum* as one species in Sweden performed better than the separate cryptic species models. Therefore, we did not produce individual genetic group distribution models due to the small dataset and the

uncertain niche differentiation among genetic groups. Second, we performed pairwise correlation analyses of environmental variables to discard autocorrelated ones. As a result, five independent environmental variables were recovered from the analyses and used in species distribution models: *bio\_1*: annual mean temperature in °C; *bio\_2*: mean diurnal range in °C; *bio\_4*: temperature seasonality, *bio\_8*: mean temperature of the wettest quarter in °C and *bio\_15*: precipitation seasonality as a coefficient of variation. Then, we reduced spatial autocorrelation of occurrences by filtering them into two classes: a maximum area of 50 km<sup>2</sup> and a minimum of 20 km<sup>2</sup> for areas with environmental variation. Finally, we created a bias file to optimize the selection of occurrences and environmental layer points using a buffered minimum convex polygon based on observed localities.

The potential distribution ranges of *R. lanuginosum* in the current time and projections into the LGM and LIG were modelled using MaxEnt v.3.4 (Phillips, Anderson & Schapire, 2006) and SDMtoolbox. We performed geographically structured *k*-fold cross-validation for the models (*k* = 3). Automatic selection of the best model consisted of running models using eight regularization multipliers (0.5, 1, 1.5, 2, 2.5, 3, 4, 5) and five model classes to evaluate the performance in terms of omission rate, the area under the curve and model complexity. The best model was the linear and quadratic with a regularization parameter of five. Finally, we used the output files to generate potential distribution maps representing expected suitable areas for *R. lanuginosum* using the 10-percentile training presence as a threshold for all projections: current, LGM and LIG.

## RESULTS

## PHYLOGEOGRAPHY AND MOLECULAR DATING BASED ON ITS

The alignment of the 147 ITS sequences included 715 bp with 110 indels (93 excluding the outgroup) and 41 (39) variable sites, of which 37 (35) were potentially parsimony-informative sites. Bayesian phylogenetic analyses of ITS sequences recovered three clades coded as blue, orange and green (Fig. 1), previously identified as cryptic taxa. The tree topology based on maximum likelihood is consistent with the Bayesian inference (see [Supporting Information, Fig. S3](#)). There is little spatial genetic structure, with some samples from North America being closely related to European specimens. The blue clade is formed by *R. lanuginosum* samples inhabiting Arctic environments such as north-eastern Canada, northern Alaska and the Scandinavian Mountains (46 samples). The green



**Table 2.** Genetic diversity and demographic statistics of *Racomitrium lanuginosum* geographical groups with corrected sample sizes based on ITS marker. Eight randomly selected individuals per group were used to perform the analysis. Populations with fewer than eight individuals were discarded. For each group, the number of samples, polymorphic sites (S), number of haplotypes (h), haplotype (Hd) and nucleotide ( $\pi$ ) diversity with their respective standard deviations (SD) and Fu & Li's  $D$  and Tajima's  $D$  statistics are presented. The Swedish northern Mountain and northern Norway populations have the highest haplotype and nucleotide diversity

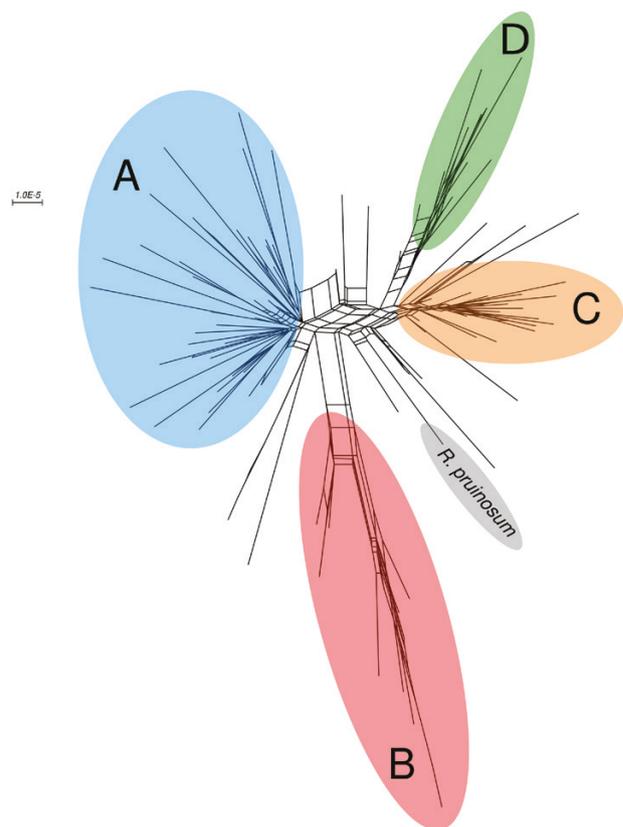
Geographic group	S	h	Hd	SD Hd	$\pi$	SD $\pi$	Fu & Li's $D$	Tajima's $D$
Alaska Peninsula	18	2	0.571	0.094	0.0162	0.0027	1.5771 *	2.4929 **
Eastern Quebec	28	3	0.464	0.2	0.0181	0.0072	1.4833 *	0.5006
Western Quebec	36	4	0.75	0.139	0.0215	0.0054	1.5269 *	1.0185
Middle Lowland-Sweden	18	3	0.75	0.096	0.0141	0.0031	1.5771 *	1.9337
Middle Mountain-Sweden	19	4	0.75	0.139	0.0142	0.0032	1.2045	1.6149
Northern Mountain-Sweden	39	4	0.821	0.101	0.0282	0.0039	0.8214	1.0066
Northern Norway	36	6	0.929	0.084	0.0271	0.0059	1.2426	1.3102
Northern Quebec	36	5	0.857	0.108	0.0231	0.0048	0.2680	0.2167
Southern Sweden	19	4	0.75	0.139	0.0142	0.0032	1.2045	1.6149
Western Europe	19	4	0.786	0.113	0.0142	0.0030	1.2045	1.6149

\*  $p < 0.05$ ; \*\*  $p < 0.01$

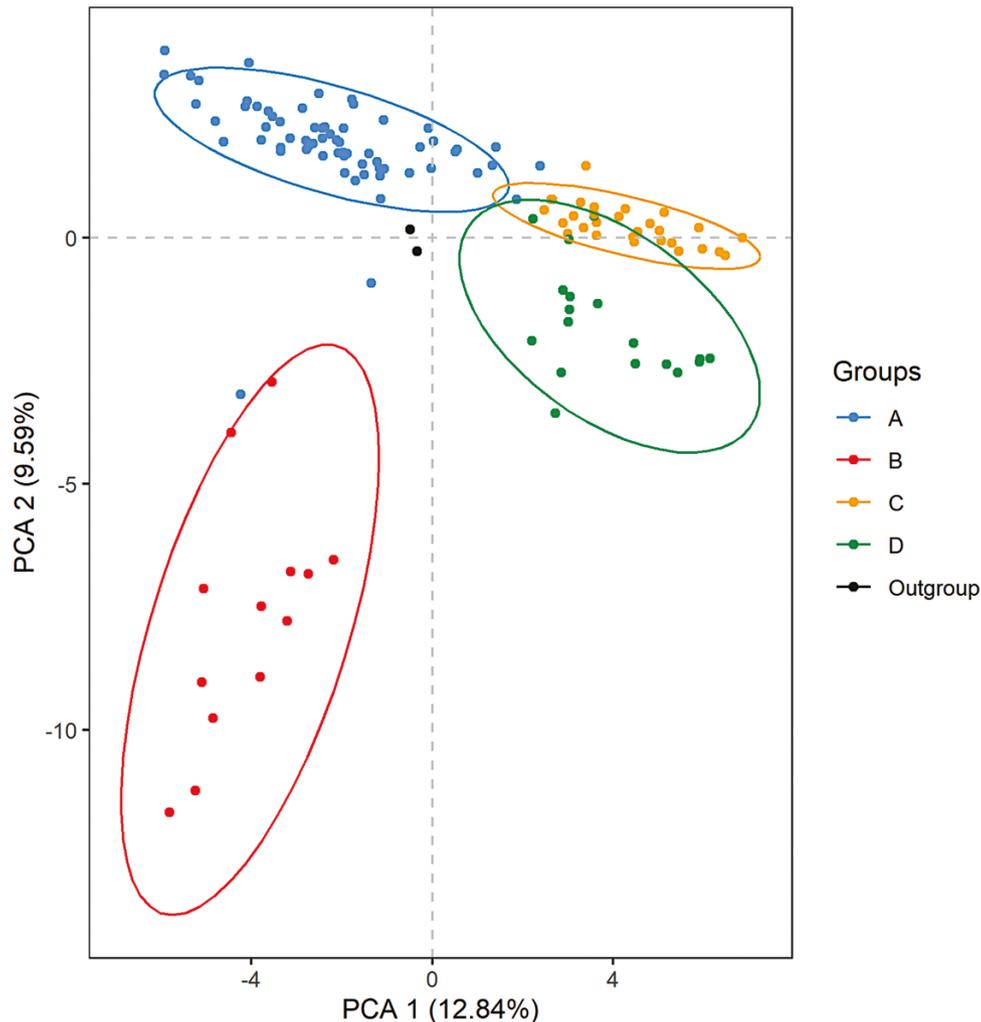
The coalescent-based phylogenetic analysis (SVDquartets) of concatenated loci resulted in 10 009 125 quartets, of which 68.61% were compatible, 31.38% were incompatible and five were discarded. The lineage tree topology was similar to the maximum-likelihood phylogenetic tree but with weaker bootstrap support (see also [Supporting Information, Fig. S7](#)). Some group A members were recovered as monophyletic subgroups intermixed with group B. Group C had low support ( $< 70$ ), and D was again recovered as monophyletic.

A PCA based on SNP data ( $-R 20$ ) supported the four genetic groups identified in the phylogenetic network, with C and D being closely related and B representing the most divergent group in *R. lanuginosum* ([Fig. 3](#)). Some samples across the phylogenetic network and PCA plot had ambiguous membership. Nonetheless, the haplotype co-ancestry matrix ( $-R 20$ ) identified the genetic groups of these samples and further highlighted the genetic structure of *R. lanuginosum* ([Fig. 4](#)). The equivalence of ITS clades and fineRAD groups for each sample is presented in the co-ancestry matrix. The distribution of genetic groups is sympatric, with groups A and D mainly distributed in North America and Greenland, whereas B and C occur in North America and Europe ([Fig. 5](#)). Some samples of group A have an estimated co-ancestry level of five with groups C and D. Group B is well-differentiated but had a co-ancestry level of five with some Scandinavian samples of group D. These results are supported by the  $F_{ST}$  comparisons among genetic groups, with group B having the highest  $F_{ST}$  values indicative of a divergent population (see also [Supporting Information, Table S7](#)).

Summary statistics of the ten geographical groups reflect the spatial genetic diversity ([Table 3](#); see



**Figure 2.** Phylogenetic network based on uncorrected  $P$  distances of 4620 concatenated loci shared by at least 20% of 126 samples ( $-R 20$ ) of *Racomitrium lanuginosum* and one of *R. pruinosum* as the outgroup. Colour codes are analogous to ITS phylogenetic analyses, but with genetic group B in red.



**Figure 3.** Principal component analysis based on 844 SNPs shared by at least 20% of 127 samples ( $-R 20$ ) of *Racomitrium lanuginosum* and the outgroup (*R. geronticum* and *R. pruinosum*). Ellipses represent the 95% confidence intervals, and the group colour code is the same as in the phylogenetic network. Group B is the most genetically differentiated.

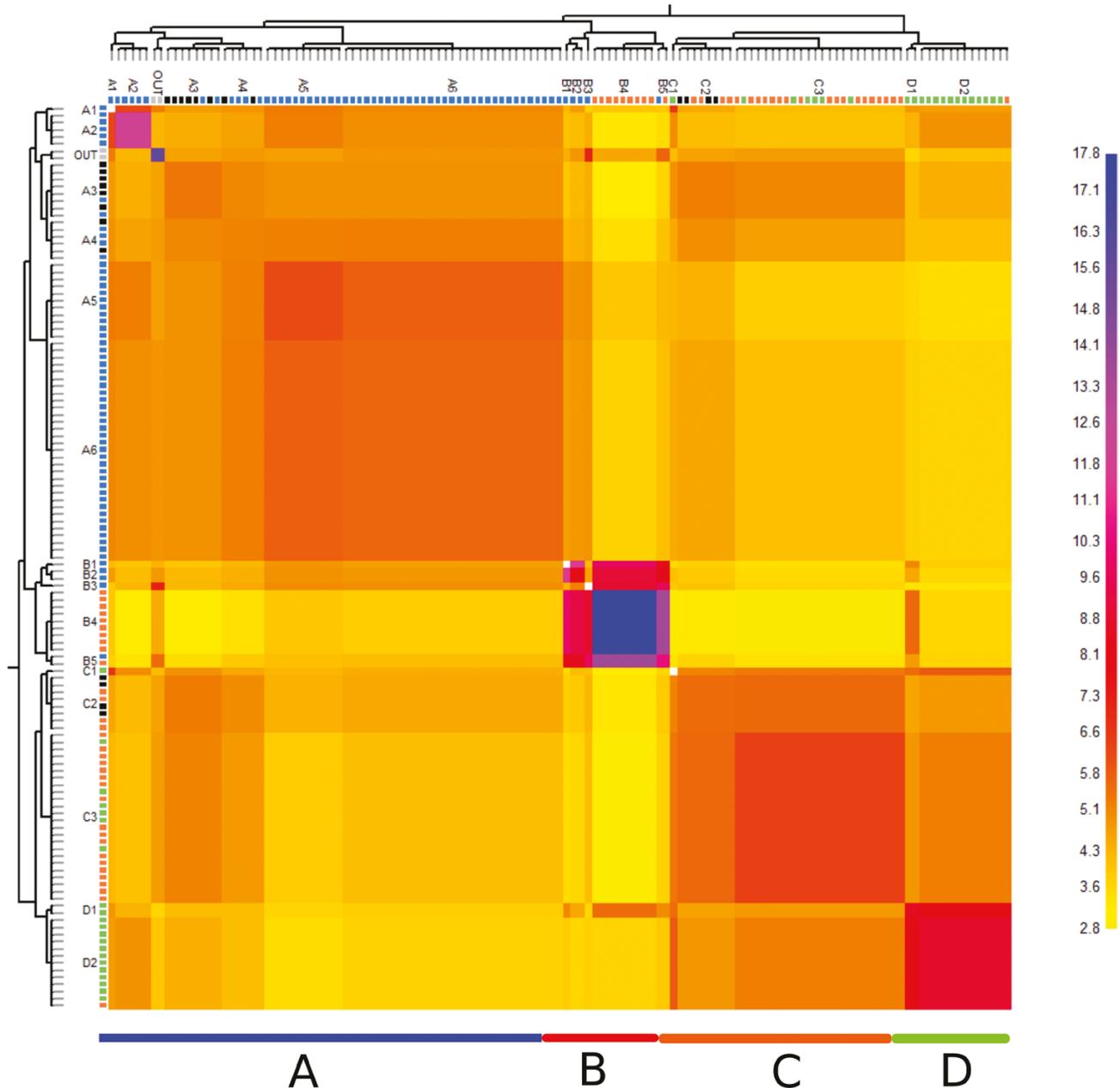
also [Supporting Information Table S8](#)). Haplotype and nucleotide diversities are higher in the Alaska Peninsula, followed by the Scandinavian Mountains and Western Quebec (see also [Supporting Information Fig. S8](#)). Polymorphic sites were higher among than within groups. The  $F_{ST}$  pairwise comparisons suggest that the Scandinavian Lowlands and Ontario are more differentiated from the rest of *R. lanuginosum* populations (see also [Supporting Information Table S9](#)).

Relationships among geographic groups inferred by TreeMix resulted in an inconsistent graph topology across replicates. However, the Scandinavian Mountains population always appeared at the base of the graph. The rest of populations were clustered, with Ontario being the most differentiated. Quebec populations (northern, western and eastern) were more related to the Scandinavian Lowlands and

Greenland populations. The evaluation of the number of migration edges indicated that one migration edge better explains the population graph with two recurrent scenarios (5/10 replicates): migration from the Alaska Peninsula or an ancient hypothetical North American population to Western Quebec. Two main topologies resulted from the ten TreeMix replicates and represented the scenarios mentioned above (see also [Supporting Information, Fig. S9](#)).

#### PRESENT AND PAST DISTRIBUTION RANGES

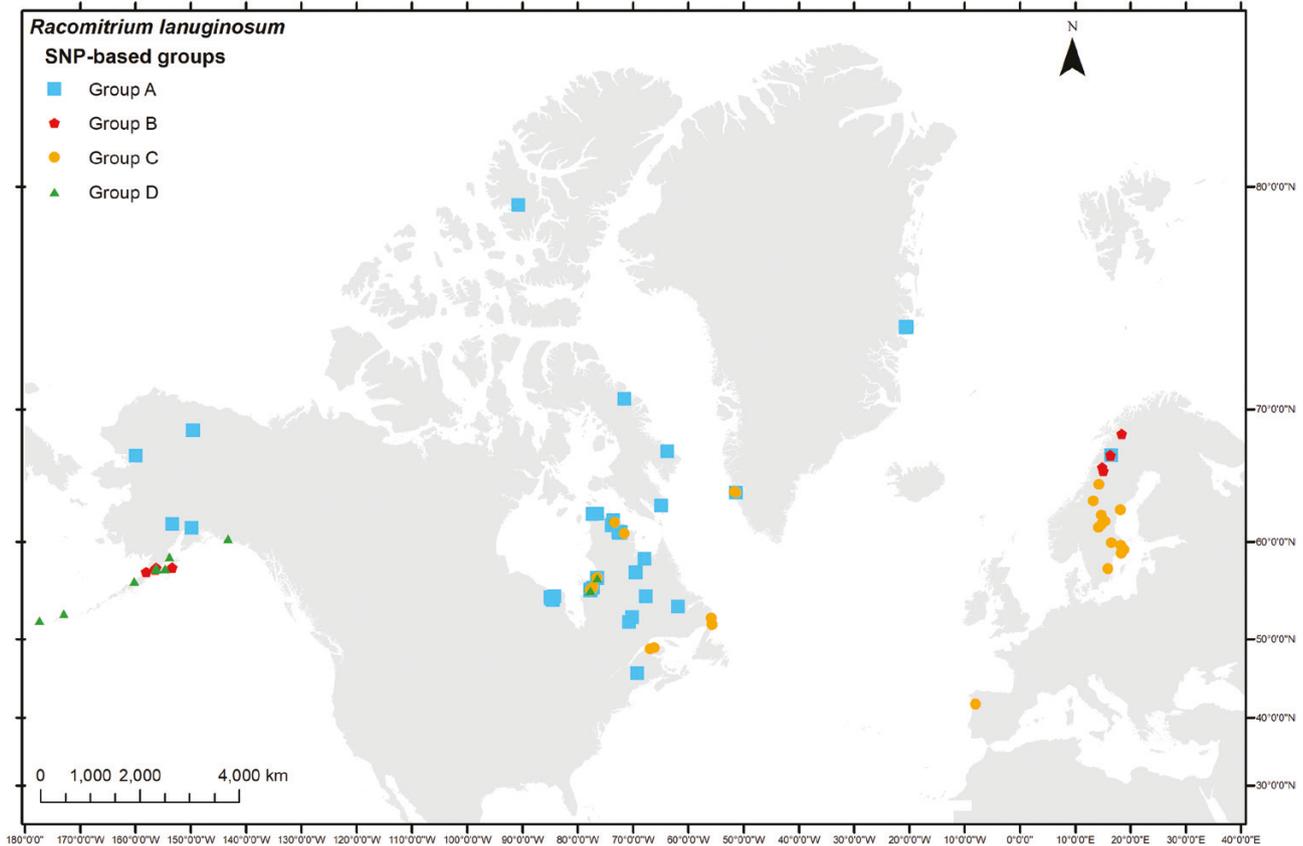
The present species distribution agrees with the known sub-Arctic and Arctic range of the species (see also [Supporting Information, Fig. S10](#)). The model accuracy result using the area under the receiver operating characteristic curve was  $AUC = 0.78$ . Past



**Figure 4.** Co-ancestry matrix of *Racomitrium lanuginosum* based on 844 SNPs at  $-R$  20 of 127 samples. On the right, the colour scale ranges from yellow, which indicates a low level of shared co-ancestry between samples, to blue, indicating a high level of co-ancestry. The left and top axes include trees showing the relationships among groups and small squares corresponding to each sample with colours representing the ITS clades (blue, orange and green) as a comparison. The four genetic groups recovered from the co-ancestry matrix are highlighted at the bottom (A, B, C, D). Samples of the same genetic group shared more co-ancestry than samples of different groups. Group B has the highest level of co-ancestry, indicating strong genetic differentiation.

distribution models of *R. lanuginosum* suggest a loss of suitable habitats during the LGM compared with the present and the LIG. The LGM model indicates that the most suitable areas for *R. lanuginosum* for that period were the coastline of western North America and Western Europe (see also [Supporting Information](#),

[Fig. S11](#)). The LIG model showed fewer habitats than the present model but a wider range than the LGM (see also [Supporting Information](#), [Fig. S12](#)). The current potential spatial distribution is broader than at the LGM, indicating a species range expansion after the retreat of glaciers.



**Figure 5.** Geographical distribution map of *Racomitrium lanuginosum* genetic groups recovered from SNP-based analyses. Group distribution patterns are similar to ITS clades, but group B is represented in southern Alaska (the Alaska Peninsula) and the Scandinavian Mountains. Genetic groups have a sympatric distribution in the Northern Hemisphere.

**Table 3.** Genetic summary statistics of ten *Racomitrium lanuginosum* geographic groups with corrected sample size based on GBS data. Five randomly selected individuals per group were used to perform the analysis. These results are based on 50 samples and loci shared by at least 60% of individuals per geographical area ( $r = 60$ ). Filters recovered 4707 loci and 1196 SNPs. For each geographical group, the mean number of samples recorded per loci, the number of total, private and variant sites, single nucleotide polymorphisms (SNPs) and the percentage of polymorphic loci (Percentage PL) are presented. In addition, the mean haplotype (Hd) and nucleotide ( $\pi$ ) diversity and their standard errors (SE) were estimated for each group using SNPs. The Scandinavian Mountains and the Alaska Peninsula have the highest genetic diversity

Geographic group	Samples per loci	Total sites	Private sites	Variant sites	SNPs	Percentage PL	Hd	SE Hd	$\pi$	SE $\pi$
Alaska Peninsula	4.40	301004	78	739	413	0.14	0.2526	0.0084	0.2878	0.0096
Alaska Mainland	4.51	297942	24	684	167	0.06	0.0980	0.0067	0.1119	0.0077
Ontario	4.57	324082	26	675	184	0.06	0.1001	0.0064	0.1150	0.0074
Northern Quebec	4.51	273251	18	585	198	0.07	0.1369	0.0080	0.1581	0.0093
Nunavut	4.57	316122	54	794	265	0.08	0.1351	0.0069	0.1542	0.0079
Western Quebec	4.62	363098	73	852	415	0.11	0.1974	0.0071	0.2263	0.0082
Eastern Quebec	4.41	281586	4	532	216	0.08	0.1689	0.0090	0.1936	0.0104
Greenland	4.39	275072	14	601	266	0.10	0.1844	0.0086	0.2121	0.0100
Scandinavian Lowlands	4.62	300052	24	681	135	0.04	0.0824	0.0065	0.0938	0.0074
Scandinavian Mountains	4.39	301237	25	606	360	0.12	0.2573	0.0088	0.2959	0.0102

## DISCUSSION

The results support the hypothesis that *R. lanuginosum* comprised differentiated molecular lineages distributed sympatrically in the Northern Hemisphere. The populations inhabiting ice-free areas during the LGM had higher genetic diversity than those previously located under the ice sheet. Species distribution models suggested a range contraction during the LGM. In addition, we found evidence of long-distance dispersal and post-glacial expansion. Regarding the GBS data, the analyses performed well using loci with high levels of missing data to infer phylogenetic relationships, as previously highlighted in other studies (Lewis *et al.*, 2017; Tripp *et al.*, 2017; Alonso-García *et al.*, 2020).

## CRYPTIC TAXA AND THEIR SPATIAL DISTRIBUTION

The ITS marker resolved the deep phylogenetic relationships of groups, and SNP data helped define the fine genetic structure. The molecular dating based on ITS estimated that the ancestor of *R. lanuginosum* originated during the Late Miocene and Early Pliocene, the onset of changes from a warm to a cool climate that reached its coldest point during the Pleistocene (De Schepper *et al.*, 2014; Willeit *et al.*, 2019). These analyses remain conservative due to the lack of precisely dated fossils and a single molecular marker, but they provide a suitable proxy for lineage divergence as exemplified in the globally distributed moss *Bryum argenteum* Hedw. (Zaccara *et al.*, 2020). The divergence between clade A and the other clades is similar to the congeneric *R. pruinosum*. This result is supported by phylogenetic trees and the NeighborNet network, which points to a well-differentiated and relatively ancient molecular lineage. Hedenäs (2020a) also reported three lineages in Scandinavian populations of *R. lanuginosum* as differentiated from each other as from the outgroups *R. pruinosum* and *R. geronticum*. They were thus interpreted as cryptic taxa. In bryophytes, recognition of cryptic species in recent research has been conducted using multilocus molecular phylogenies (e.g. McDaniel & Shaw, 2003; Hedenäs, 2018) and complete organellar genomes (Myszczynski *et al.*, 2017; Šlipiko *et al.*, 2020).

The genetic structure analyses of *R. lanuginosum* resolved three clades based on ITS and four genetic groups using SNP data. These results confirm that *R. lanuginosum* comprises distinct molecular lineages across the Northern Hemisphere. The three ITS clades found in this study are similar to the cryptic species proposed by Hedenäs (2020a) in *R. lanuginosum* of Scandinavia, with one restricted to the mountain range (L3 = blue) and the other two widely distributed in lowlands (L1 and L2 = orange and green; see also

Supporting Information, Fig. S4). The main difference between ITS and SNP-based phylogenetic trees was the recovery of group B, only distributed in the Alaskan Peninsula and northern Scandinavia. The haplotype 10 of the ITS orange clade corresponds to the SNP-based group B (see also Supporting Information, Table S1). As suggested by the SVDquartets results, the incongruent position of group B may be caused by the retention of ancient polymorphisms.

The four *R. lanuginosum* genetic groups are also supported by the phylogenetic network, PCA, coalescent-based analysis, co-ancestry matrix and  $F_{ST}$  indexes. Particularly, genetic groups were not recovered as monophyletic in the phylogenetic network suggesting cryptic molecular taxa. Most genetic groups have sympatric distribution in eastern North America and Europe. The occurrence of sympatric molecular lineages in the Northern Hemisphere, particularly the North Atlantic, is a widespread phenomenon among mosses (Kyrkjeeide *et al.*, 2016a; Lewis *et al.*, 2017; Hedenäs, Kuznetsova & Ignatov, 2020). Long-distance dispersal in *R. lanuginosum* could be an infrequent event due to the rare production of sporophytes. However, phylogenetic analyses, ancestral area reconstructions and TreeMix graphs revealed potential transatlantic dispersal from North America to Europe. Groups A and B are mainly distributed in Arctic regions, whereas C and D are widespread in the sub-Arctic and some Arctic localities. Accordingly, previous distribution models performed for the three *R. lanuginosum* cryptic taxa in Scandinavia differentiated one lineage with suitable mountain habitats and the other two in lower elevations (Collart *et al.*, 2021). The geographical distribution of group A in Arctic environments and groups C and D in the sub-Arctic may support the hypothesis of similar morphology but differential physiological responses of cryptic taxa to the environment. In addition, a thorough morphological examination of the recovered molecular taxa could provide critical characters to differentiate them.

## POST-GLACIAL DEMOGRAPHIC INFERENCE

The climatic events of the Quaternary affected the demographic history of *R. lanuginosum*, but did not seem to be directly related to the divergence of molecular lineages. Nucleotide and haplotype diversity within groups, based on both ITS and SNPs, agreed with estimates of divergence times. The ITS blue clade has higher haplotype and nucleotide diversity (Table 2) and diverged 5.1 Mya (95% HPD 6.5–3.97). The ITS marker failed to discriminate group B as an independent clade, and it was clustered in the orange clade as an Alaskan sub-clade. In addition, the northern Scandinavian populations have higher

genetic diversity and more haplotypes. According to demographic statistics, the Alaska Peninsula has experienced population size contraction, probably due to glaciations (Table 2).

SNP summary statistics of geographical groups revealed the Alaska Peninsula, the Scandinavian Mountains and Western Quebec as the most genetically diverse areas (Table 3; see also Supporting Information, Fig. S8, Tables S5 and S8). For example, the Alaska Peninsula and the Scandinavian Mountains populations were more differentiated based on the TreeMix graph (see also Supporting Information, Fig. S9). The genetic signatures in these geographical areas may be related to historical events and the preservation of different haplotypes through time. The LGM distribution model indicated that the Alaska Peninsula was a suitable habitat for *R. lanuginosum*, and the ice sheet did not entirely cover it (see also Supporting Information, Fig. S11). In other bryophytes and vascular plants, refugial populations seemed to be genetically diverse and different from other populations (Petit *et al.*, 2003; Hedenäs, 2008, 2012; Kyrkjeeide *et al.*, 2016a; Westergaard *et al.*, 2019; Pérez-Escobar *et al.*, 2020). Thus, some *R. lanuginosum* populations could have survived glaciations in Beringia, a well-known Quaternary refugium for other plants (Abbott *et al.*, 2000; Abbott & Brochmann, 2003). Our current data indicate that the Alaska Peninsula populations may represent glacial refugia due to the high genetic diversity, private alleles and their location in ice-free areas during the LGM.

Also, the Scandinavian Mountains and Western Quebec had high genetic diversity. The Scandinavian Mountains may be represented by populations that survived under the ice sheet or *in situ*, as exemplified in the hypothesized refugia located on the coast of Norway (Kyrkjeeide *et al.*, 2014; Westergaard *et al.*, 2019). Gene flow from the Alaska Peninsula or the Scandinavian Mountains may account for the higher diversity in Western Quebec than in nearby areas (see also Supporting Information, Figs S8 and S9). Nevertheless, the retention of ancient polymorphisms due to *in situ* survival in microrefugia or under the ice sheet could also explain the high genetic diversity in this population. For example, at the retreating edge of the Twin Glacier in Ellesmere Island, Canada (1410 to 1690 years BP), *R. lanuginosum* was discovered with a high content of chlorophyll comparable to extant populations even though the pigment degraded after some days of light exposure (La Farge, Williams & England, 2013). Additionally, bryophyte populations have been found in an ice-retreated margin of Teardrop Glacier in Nunavut, Canada, dating from 404 to 614 years BP; after *in vitro* culture, four species exhibited regrowth from ice entombment (Bergsma, Svoboda & Freedman, 1984). Even more prolonged

survival under permafrost has been reported for *Chorisodontium aciphyllum* (Hook.f. & Wilson) Broth. in the Antarctic area (c. 1.5 kya) by Roads, Longton & Convey (2014). Nunataks might also function as refugial areas (e.g. Pan *et al.*, 2020; Paus, 2021), especially for *R. lanuginosum* that can survive high Arctic conditions. These examples illustrate the potential ability of bryophytes to cope with extreme cold periods, at least for hundreds of years.

The LGM might have influenced the demographic dynamics of *R. lanuginosum* contracting its distribution range as recovered by the distribution models. Haplotype and nucleotide diversities are low in recent North American deglaciated areas and the Scandinavian Lowlands (Table 3; see also Supporting Information, Fig. S8). These results agree with population bottlenecks during the LGM followed by post-glacial colonization for other Amphi-Atlantic bryophytes (Désamoré *et al.*, 2016). The comparison of the present distribution model with the LGM and LIG shows a range expansion to northern habitats. Refugial populations of *R. lanuginosum*, such as in the Alaska Peninsula, could have dispersed to ice-free areas in the Northern Hemisphere as glaciers retreated (Abbott & Brochmann, 2003). These dispersion routes are evident in the TreeMix population graph. The Alaska Peninsula and the Scandinavian Mountains are at the base, followed by northern North America, eastern North America and the Scandinavian Lowlands (see also Supporting Information, Fig. S9). For Europe, the species could survive cold conditions between the European Alps and the northern ice sheet in the tundra and then migrate to Scandinavia. This scenario is exemplified by the mosses *Rhytidium rugosum* (Hedw.) Kindb. and *Drepanocladus turgescens* (T. Jensen) Loeske, which both have genetically distinct populations in Scandinavian Mountains and Lowlands (Hedenäs, 2014, 2015; Hedenäs & Bisang, 2019). Moreover, the addition of southern European populations could shed light on post-glacial dynamics in *R. lanuginosum* at a continental scale, specifically to test whether this species also follows the European southern refugium hypothesis (see Kyrkjeeide *et al.*, 2014). These results show how the LGM affected the spatial genetic structure of the *R. lanuginosum*.

Finally, our data indicate that geographical groups have experienced migration events. For example, the Western Quebec population could have experienced gene flow with the Alaska Peninsula population (see also Supporting Information, Fig. S9). TreeMix analyses must be considered exploratory because they were conducted with a sliding window of ten SNPs; however, these results agreed with the levels of co-ancestry observed among groups in fineRADstructure and  $F_{ST}$  results (Fig. 4; see also Supporting Information, Table S9).

## CONCLUSIONS

*Racomitrium lanuginosum* comprises distinct molecular lineages, most probably corresponding to cryptic species with sympatric distribution across the Northern Hemisphere. We recovered three clades with the ITS marker (blue, green and orange) and four SNP-based genetic groups (A–D) with different distributions in Arctic and sub-Arctic environments diverging around 5 Mya. Long-distance dispersal has occurred from eastern North America to Scandinavia. Populations from the Alaska Peninsula could be considered glacial refugia according to their nucleotide and haplotype diversity, private alleles,  $F_{ST}$  estimates and the geographical distribution in ice-free areas during the LGM. There is evidence of *in situ* survival in the Scandinavian Mountains. Furthermore, the species probably experienced a range contraction during the LGM followed by post-glacial dispersal from refugial populations northwards in the Northern Hemisphere. Finally, we conclude that complex events such as cryptic speciation, long-distance dispersal and post-glacial expansion are shaping the evolutionary history of *R. lanuginosum*. Applying an integrative approach with a Sanger-sequenced marker, GBS and distribution models allowed elucidation of factors that shaped the observed genetic signatures. The phylogeography of this northern moss exemplifies the complex demographic history in cold environments and contributes to recognizing evolutionary patterns in the Northern Hemisphere.

## ACKNOWLEDGEMENTS

We would like to thank all the herbaria that kindly loaned us specimens for this study. We thank the Genomic Analysis Platform of Laval University for the advice on sequencing and bioinformatic processing. We also thank the Centre for Northern Studies (CEN) for the facilities provided in Kuujuarapik and Umiujaq. We thank Herbière Louis-Marie (QFA) for providing plant material and facilities, especially to Kim Damboise and Catherine Boudreault, for their help during fieldwork. DAEO thanks Marta Alonso García for help with bioinformatic pipelines. Finally, we thank the anonymous referees for their comments to improve the manuscript. Specimens from Jan Mayen and Svalbard were collected by M.S. and colleague H. Kruijer with support from the Dutch Research Council (NWO) grant ALW-NAP/08-01, European Commission (EU) grant ARCFAC-026129-2008-31, Royal Netherlands Navy (Jan Mayen expedition 2014) and NWO-funded Netherlands Scientific Expedition Edgeøya Spitsbergen (SEES) 2015. This research received financing from the Mexican National

Council for Science and Technology (CONACYT) through a doctoral fellowship awarded to DAEO, the Discovery Grant (NSERC): RGPIN-2016-05967 and the Canadian Foundation for Innovation (CFI): 39135. This article is part of the PhD thesis of DAEO.

## AUTHOR CONTRIBUTIONS

DAEO and JCVA conceived the study. DAEO collected the Canadian samples, extracted DNA from voucher specimens, produced the ITS sequences and GBS data and conducted the analyses. LH provided DNA and ITS sequences of Scandinavian samples. MS and JL produced ITS sequences from European populations. MS, JL, DQ and DH collected, sequenced and processed samples from Svalbard and sister species using GBS. DH helped with the GBS data processing; DAEO wrote the manuscript with the comments of the co-authors.

## DATA AVAILABILITY

Sequences for ITS correspond to GenBank submission SUB9624643: accessions MZ227145 to MZ227205. Genotyping-by-sequencing data was deposited under the BioProject ID PRJNA735773 containing the BioSample accessions: SAMN19596317 to SAMN19596443. The raw occurrence records downloaded from GBIF containing 28 205 entries are available at <https://doi.org/10.15468/dl.ktasda>. Scripts for all the analyses are available from the corresponding author on request.

## REFERENCES

- Abbott RJ, Brochmann C. 2003.** History and evolution of the Arctic flora: in the footsteps of Eric Hultén. *Molecular Ecology* **12**: 299–313.
- Abbott RJ, Smith LC, Milne RI, Crawford RMM, Wolff K, Balfour J. 2000.** Molecular analysis of plant migration and refugia in the Arctic. *Science* **289**: 1343–1346.
- Alonso-García M, Grewe F, Payette S, Villarreal AJC. 2021.** Population genomics of a reindeer lichen species from North American lichen woodlands. *American Journal of Botany* **108**: 159–171.
- Alonso-García M, Villarreal A JC, McFarland K, Goffinet B. 2020.** Population genomics and phylogeography of a clonal bryophyte with spatially separated sexes and extreme sex ratios. *Frontiers in Plant Science* **11**: 1–15.
- Alors D, Grande FD, Cubas P, Crespo A, Schmitt I, Molina MC, Divakar PK. 2017.** Panmixia and dispersal from the Mediterranean Basin to Macaronesian islands of a macrolichen species. *Scientific Reports* **7**: 40879.

- Andrews S.** 2010. *FASTQC. A quality control tool for high throughput sequence data.* Available at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Baddeley JA, Thompson DBA, Lee JA.** 1994. Regional and historical variation in the nitrogen content of *Racomitrium lanuginosum* in Britain in relation to atmospheric nitrogen deposition. *Environmental Pollution* **84**: 189–196.
- Baughman JT, Payton AC, Paasch AE, Fisher KM, McDaniel SF.** 2017. Multiple factors influence population sex ratios in the Mojave Desert moss *Syntrichia caninervis*. *American Journal of Botany* **104**: 733–742.
- Bergsma BM, Svoboda J, Freedman B.** 1984. Entombed plant communities released by a retreating glacier at central Ellesmere Island, Canada. *Arctic* **37**: 49–52.
- Biersma EM, Convey P, Wyber R, Robinson SA, Downton M, van de Vijver B, Linse K, Griffiths H, Jackson JA.** 2020. Latitudinal biogeographic structuring in the globally distributed moss *Ceratodon purpureus*. *Frontiers in Plant Science* **11**: 1–14.
- Biersma EM, Jackson JA, Bracegirdle TJ, Griffiths H, Linse K, Convey P.** 2018. Low genetic variation between South American and Antarctic populations of the bank-forming moss *Chorisodontium aciphyllum* (Dicranaceae). *Polar Biology* **41**: 1–12.
- Biersma EM, Jackson JA, Hyvönen J, Koskinen S, Linse K, Griffiths H, Convey P.** 2017. Global biogeographic patterns in bipolar moss species. *Royal Society Open Science* **4**: 170147.
- Boisselier-Dubayle MC, Jubier MF, Lejeune B, Bischler H.** 1995. Genetic variability in the three subspecies of *Marchantia polymorpha* (Hepatitaceae): isozymes, RFLP and RAPD markers. *Taxon* **44**: 363–376.
- Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Brown JL.** 2014. SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution* **5**: 694–700.
- Brown JL, Hill DJ, Dolan AM, Carnaval AC, Haywood AM.** 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data* **5**: 180254.
- Carpenter EJ, Matasci N, Ayyampalayam S, Wu S, Sun J, Yu J, Jimenez Vieira FR, Bowler C, Dorrell RG, Gitzendanner MA, Li L, Du W, Ullrich KK, Wickett NJ, Barkmann TJ, Barker MS, Leebens-Mack JH, Wong GKS.** 2019. Access to RNA-sequencing data from 1,173 plant species: the 1000 Plant Transcriptomes Initiative (1KP). *GigaScience* **8**: giz126.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA.** 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**: 3124–40.
- Chifman J, Kubatko L.** 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* **30**: 3317–3324.
- Chou J, Gupta A, Yaduvanshi S, Davidson R, Nute M, Mirarab S, Warnow T.** 2015. A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics* **16**: S2.
- Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM.** 2009. The Last Glacial Maximum. *Science* **325**: 710–714.
- Collart F, Hedenäs L, Broennimann O, Guisan A, Vanderpoorten A.** 2021. Intraspecific differentiation: implications for niche and distribution modelling. *Journal of Biogeography* **48**: 415–426.
- Cronberg N, Molau U, Sonesson M.** 1997. Genetic variation in the clonal bryophyte *Hylocomium splendens* at hierarchical geographical scales in Scandinavia. *Heredity* **78**: 293–301.
- De Schepper S, Gibbard PL, Salzmann U, Ehlers J.** 2014. A global synthesis of the marine and terrestrial evidence for glaciation during the Pliocene Epoch. *Earth-Science Reviews* **135**: 83–102.
- Désamoré A, Patiño J, Mardulyn P, McDaniel SF, Zanatta F, Laenen B, Vanderpoorten A.** 2016. High migration rates shape the post-glacial history of amphiatlantic bryophytes. *Molecular Ecology* **25**: 5568–5584.
- Drummond AJ, Rambaut A.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**: 214.
- Ellis CJ, Tallis JH.** 2003. Ecology of *Racomitrium lanuginosum* in British blanket mire - evidence from the palaeoecological record. *Journal of Bryology* **25**: 7–15.
- Evanno G, Regnaut S, Goudet J.** 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.
- Fitak RR.** 2021. OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols* **6**: bpab017.
- Fuselier L, Davison PG, Clements M, Shaw B, Devos N, Heinrichs J, Hentschel J, Sabovljevic M, Szövényi P, Schuette S, Hofbauer W, Shaw AJ.** 2009. Phylogeographic analyses reveal distinct lineages of the liverworts *Metzgeria furcata* (L.) Dumort. and *Metzgeria conjugata* Lindb. (Metzgeriaceae) in Europe and North America. *Biological Journal of the Linnean Society* **98**: 745–756.
- Gavin DG, Gavin DG, Fitzpatrick MC, Gugger PF, Heath KD, Dobrowski SZ, Hampe A, Hu FS, Ashcroft MB, Bartlein PJ, Blois JL, Carstens BC, Davis EB, Lafontaine G De, Edwards ME, Fernandez M, Henne PD, Herring EM, Tsai Y hsin E, Williams JW.** 2014. Climate refugia joint inference from fossil records, SDMs and phylogeography. *New Phytologist* **204**: 37–54.
- Grewe F, Huang JP, Leavitt SD, Lumbsch HT.** 2017. Reference-based RADseq resolves robust relationships among closely related species of lichen-forming fungi using metagenomic DNA. *Scientific Reports* **7**: 1–11.
- Grewe F, Lagostina E, Wu H, Printzen C, Lumbsch HT.** 2018. Population genomic analyses of RAD sequences resolves the phylogenetic relationship of the lichen-forming fungal species *Usnea antarctica* and *Usnea aurantiacoatra*. *MycKeys* **113**: 91–113.
- Grundmann M, Ansell SW, Russell SJ, Koch MA, Vogel JC.** 2008. Hotspots of diversity in a clonal world - the Mediterranean moss *Pleurochaete squarrosa* in Central Europe. *Molecular Ecology* **17**: 825–838.

- Hassel K, Gunnarsson U, Gunnarsson U. 2011. The use of inter simple sequence repeats (ISSR) in bryophyte population studies. *Oikos* **28**: 152–157.
- Hedenäs L. 2008. Molecular variation and speciation in *Antitrichia curtispindula* s.l. (Leucodontaceae, Bryophyta). *Botanical Journal of the Linnean Society* **156**: 341–354.
- Hedenäs L. 2010. Phylogeography and origin of European *Sanionia uncinata* (Amblystegiaceae, Bryophyta). *Systematics and Biodiversity* **8**: 177–191.
- Hedenäs L. 2012. Global phylogeography in *Sanionia uncinata* (Amblystegiaceae: Bryophyta). *Botanical Journal of the Linnean Society* **168**: 19–42.
- Hedenäs L. 2014. Intraspecific genetic variation in selected mosses of Scandinavian interglacial refugia suggests contrasting distribution history patterns. *Botanical Journal of the Linnean Society* **176**: 295–310.
- Hedenäs L. 2015. *Rhytidium rugosum* (Bryophyta) colonized Scandinavia from at least two glacial refugial source populations. *Botanical Journal of the Linnean Society* **179**: 635–657.
- Hedenäs L. 2018. Conservation status of the two cryptic species of *Hamatocaulis vernicosus* (Bryophyta) in Sweden. *Journal of Bryology* **40**: 307–315.
- Hedenäs L. 2019. On the frequency of northern and mountain genetic variants of widespread species: essential biodiversity information in a warmer world. *Botanical Journal of the Linnean Society* **191**: 440–474.
- Hedenäs L. 2020a. Cryptic speciation revealed in Scandinavian *Racomitrium lanuginosum* (Hedw.) Brid. (Grimmiaceae). *Journal of Bryology* **42**: 117–127.
- Hedenäs L. 2020b. Cryptic and morphologically recognizable species diversity within Scandinavian *Plagiopus oederianus* (Bryophyta: Bartramiaceae). *Lindbergia* **43**: linbg.01130.
- Hedenäs L, Bisang I. 2019. Are the remains of the Central European population of *Drepanocladus turgescens* genetically distinct from Scandinavian populations? *Herzogia* **32**: 209–218.
- Hedenäs L, Kuznetsova OI, Ignatov MS. 2020. A revision of the genus *Tomentypnum* (Amblystegiaceae) in northern Eurasia. *Bryologist* **123**: 377–395.
- Hernández-Rojas AC, Kluge J, Krömer T, Carvajal-Hernández C, Silva-Mijangos L, Miede G, Lehnert M, Weigand A, Kessler M. 2020. Latitudinal patterns of species richness and range size of ferns along elevational gradients at the transition from tropics to subtropics. *Journal of Biogeography* **47**: 1383–1397.
- Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**: 183–195.
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254–267.
- Hutsemekers V, Hardy OJ, Mardulyn P, Shaw AJ, Vanderpoorten A. 2010. Macroecological patterns of genetic structure and diversity in the aquatic moss *Platyhypnidium riparioides*. *New Phytologist* **185**: 852–864.
- Hutsemekers V, Hardy OJ, Vanderpoorten A. 2013. Does water facilitate gene flow in spore-producing plants? Insights from the fine-scale genetic structure of the aquatic moss *Rhynchostegium riparioides* (Brachytheciaceae). *Aquatic Botany* **108**: 1–6.
- Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405.
- Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M. 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* **4**: 170122.
- Karger DN, Nobis MP, Normand S, Graham CH, Niklaus E. 2021. CHELSA-TraCE21k v1.0. Downscaled transient temperature and precipitation data since the Last Glacial Maximum. *Climate of the Past*. <https://doi.org/10.5194/cp-2021-30>.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.
- Klarenberg IJ, Keuschnig C, Russi Colmenares AJ, Warshan D, Jungblut AD, Jónsdóttir IS, Vilhelmsson O. 2021. Long-term warming effects on the microbiome and *nifH* gene abundance of a common moss species in sub-Arctic tundra. *New Phytologist* doi: [10.1111/nph.17837](https://doi.org/10.1111/nph.17837)
- Köckinger H, Hedenäs L. 2017. A farewell to *Tortella bambergeri* (Pottiaceae) as understood over the last decades. *Journal of Bryology* **39**: 213–225.
- Korpelainen H, von Cräutlein M, Kostamo K, Virtanen V. 2013. Spatial genetic structure of aquatic bryophytes in a connected lake system. *Plant Biology* **15**: 514–521.
- Kyrkjeeide MO, Hassel K, Flatberg KI, Shaw AJ, Brochmann C, Stenøien HK. 2016a. Long-distance dispersal and barriers shape genetic structure of peatmosses (*Sphagnum*) across the Northern Hemisphere. *Journal of Biogeography* **43**: 1215–1226.
- Kyrkjeeide MO, Hassel K, Flatberg KI, Shaw AJ, Yousefi N, Stenøien HK. 2016b. Spatial genetic structure of the abundant and widespread peatmoss *Sphagnum magellanicum* Brid. *PLoS One* **11**: 1–19.
- Kyrkjeeide MO, Stenøien HK, Flatberg KI, Hassel K. 2014. Glacial refugia and post-glacial colonization patterns in European bryophytes. *Lindbergia* **2**: 47–59.
- La Farge C, Williams KH, England JH. 2013. Regeneration of Little Ice Age bryophytes emerging from a polar glacier with implications of totipotency in extreme environments. *Proceedings of the National Academy of Sciences, USA* **110**: 9839–9844.
- Laenen B, Shaw B, Schneider H, Goffinet B, Paradis E, Désamoré A, Heinrichs J, Villarreal JC, Gradstein SR, McDaniel SF, Long DG, Forrest LL, Hollingsworth ML, Crandall-Stotler B, Davis EC, Engel J, Von Konrat M, Cooper ED, Patiño J, Cox CJ, Vanderpoorten A, Shaw AJ. 2014. Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts. *Nature Communications* **5**: 5134.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

- Larraín J, Quandt D, Stech M, Muñoz J. 2013. Lumping or splitting? The case of *Racomitrium* (Bryophytina: Grimmiaceae). *Taxon* **62**: 1117–1132.
- Ledent A, Désamoré A, Laenen B, Mardulyn P, McDaniel SF, Zanatta F, Patiño J, Vanderpoorten A. 2019. No borders during the post-glacial assembly of European bryophytes. *Ecology Letters* **22**: 973–986.
- Ledent A, Gauthier J, Pereira M, Overson R, Laenen B, Mardulyn P, Gradstein SR, de Haan M, Ballings P, Van der Beeten I, Zartman CE, Vanderpoorten A. 2020. What do tropical cryptogams reveal? Strong genetic structure in Amazonian bryophytes. *New Phytologist* **228**: 640–650.
- Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S, Porsch M, Quint M, Rensing SA, Soltis DE, Soltis PS, Stevenson DW, Ullrich KK, Wickett NJ, DeGironimo L, Edger PP, Jordon-Thaden IE, Joya S, Liu T, Melkonian B, Miles NW, Pokorny L, Quigley C, Thomas P, Villarreal JC, Augustin MM, Barrett MD, Baucom RS, Beerling DJ, Benstein RM, Biffin E, Brockington SF, Burge DO, Burris JN, Burris KP, Burtet-Sarramegna V, Caicedo AL, Cannon SB, Çebi Z, Chang Y, Chater C, Cheeseman JM, Chen T, Clarke ND, Clayton H, Covshoff S, Crandall-Stotler BJ, Cross H, DePamphilis CW, Der JP, Determann R, Dickson RC, Di Stilio VS, Ellis S, Fast E, Feja N, Field KJ, Filatov DA, Finnegan PM, Floyd SK, Fogliani B, García N, Gâteblé G, Godden GT, Goh F (Qi Y, Greiner S, Harkess A, Heaney JM, Helliwell KE, Heyduk K, Hibberd JM, Hodel RGJ, Hollingsworth PM, Johnson MTJ, Jost R, Joyce B, Kapralov MV, Kazamia E, Kellogg EA, Koch MA, Von Konrat M, Könyves K, Kutchan TM, Lam V, Larsson A, Leitch AR, Lentz R, Li FW, Lowe AJ, Ludwig M, Manos PS, Mavrodiev E, McCormick MK, McKain M, McLellan T, McNeal JR, Miller RE, Nelson MN, Peng Y, Ralph P, Real D, Riggins CW, Ruhsam M, Sage RF, Sakai AK, Scascitella M, Schilling EE, Schlösser EM, Sederoff H, Servick S, Sessa EB, Shaw AJ, Shaw SW, Sigel EM, Skema C, Smith AG, Smithson A, Stewart CN, Stinchcombe JR, Szövényi P, Tate JA, Tiebel H, Trapnell D, Villegente M, Wang CN, Weller SG, Wenzel M, Weststrand S, Westwood JH, Whigham DF, Wu S, Wulff AS, Yang Y, Zhu D, Zhuang C, Zuidof J, Chase MW, Pires JC, Rothfels CJ, Yu J, Chen C, Chen L, Cheng S, Li J, Li R, Li X, Lu H, Ou Y, Sun X, Tan X, Tang J, Tian Z, Wang F, Wang J, Wei X, Xu X, Yan Z, Yang F, Zhong X, Zhou F, Zhu Y, Zhang Y, Ayyampalayam S, Barkman TJ, Nguyen NP, Matasci N, Nelson DR, Sayyari E, Wafula EK, Walls RL, Warnow T, An H, Arrigo N, Baniaga AE, Galuska S, Jorgensen SA, Kidder TI, Kong H, Lu-Irving P, Marx HE, Qi X, Reardon CR, Sutherland BL, Tiley GP, Welles SR, Yu R, Zhan S, Gramzow L, Theißen G, Wong GKS. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**: 679–685.
- Lewis LR, Biersma EM, Carey SB, Holsinger K, McDaniel SF, Rozzi R, Goffinet B. 2017. Resolving the Northern Hemisphere source region for the long-distance dispersal event that gave rise to the South American endemic dung moss *Tetraplodon fuegianus*. *American Journal of Botany* **104**: 1651–1659.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- Malinsky M, Trucchi E, Lawson DJ, Falush D. 2018. RADpainter and fineRADstructure: population inference from RADseq data. *Molecular Biology and Evolution* **35**: 1284–1290.
- McDaniel SF, Shaw AJ. 2003. Phylogeographic structure and cryptic speciation in the trans-Antarctic Moss *Pyrrhobryum mnioides*. *International Journal of Organic Evolution* **57**: 205–215.
- McDaniel SF, Shaw AJ. 2005. Selective sweeps and intercontinental migration in the cosmopolitan moss *Ceratodon purpureus* (Hedw.) Brid. *Molecular Ecology* **14**: 1121–1132.
- Medina R, Lara F, Goffinet B, Garilleti R, Mazimpaka V. 2013. Unnoticed diversity within the disjunct moss *Orthotrichum tenellum* s.l. validated by morphological and molecular approaches. *Taxon* **62**: 1133–1152.
- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *2010 Gateway Computing Environments Workshop (GCE)*. IEEE, 1–8.
- Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids* **8**: 4321–4325.
- Myszczyński K, Bączkiewicz A, Buczkowska K, Ślipiko M, Szczecińska M, Sawicki J. 2017. The extraordinary variation of the organellar genomes of the *Aneura pinguis* revealed advanced cryptic speciation of the early land plants. *Scientific Reports* **7**: 1–12.
- Ochyra R, Bednarek-Ochyra H. 2007. *Racomitrium*. In: Flora of North America Editorial Committee, eds. *Flora of North America North of Mexico, Provisional Publication*. New York: Oxford.
- Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A. 2006. Simulating Arctic climate warmth and icefield retreat in the Last Interglaciation. *Science* **311**: 1751–1753.
- Pan D, Hülber K, Willner W, Schneeweiss GM. 2020. An explicit test of Pleistocene survival in peripheral versus nunatak refugia in two high mountain plant species. *Molecular Ecology* **29**: 172–183.
- Panier JR, Stevens JR, Catchen JM. 2017. Lost in parameter space: a road map for Stacks. *Methods in Ecology and Evolution* **8**: 1360–1373.
- Patiño J, Hedenäs L, Dirkse G, Ignatov M, Papp B, Müller F, González-Mancebo J, Vanderpoorten A. 2017. Species delimitation in the recalcitrant moss genus *Rhynchostegiella* (Brachytheciaceae). *Taxon* **66**: 293–308.

- Patiño J, Vanderpoorten A. 2018.** Bryophyte biogeography. *Critical Reviews in Plant Sciences* **2689**: 175–209.
- Paus A. 2021.** Lake Heimtjønn at Dovre, Mid-Norway, reveals remarkable late-glacial and Holocene sedimentary environments and the early establishment of spruce (*Picea abies*), alder (*Alnus cf. incana*), and alpine plants with present centric distributions. *Quaternary International* **580**: 38–52.
- Pérez-Escobar OA, Bogarín D, Schley R, Bateman RM, Gerlach G, Harpke D, Brassac J, Fernández-Mazuecos M, Dodsworth S, Hagsater E, Blanco MA, Gottschling M, Blattner FR. 2020.** Resolving relationships in an exceedingly young Neotropical orchid lineage using genotyping-by-sequencing data. *Molecular Phylogenetics and Evolution* **144**: 106672.
- Petit RJ, Aguinagalde I, De Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG. 2003.** Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**: 1563–1565.
- Phillips SJ, Anderson RP, Schapire RE. 2006.** Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231–259.
- Pickrell JK, Pritchard JK. 2012.** Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* **8**: e1002967.
- Piñeiro R, Popp M, Hassel K, Listl D, Westergaard KB, Flatberg KI, Stenøien HK, Brochmann C. 2012.** Circumarctic dispersal and long-distance colonization of South America: the moss genus *Cinclidium*. *Journal of Biogeography* **39**: 2041–2051.
- Pisa S, Biersma EM, Convey P, Patiño J, Vanderpoorten A, Werner O, Ros RM. 2014.** The cosmopolitan moss *Bryum argenteum* in Antarctica: recent colonisation or *in situ* survival? *Polar Biology* **37**: 1469–1477.
- R Core Team. 2017.** R development core team. *R: A Language and Environment for Statistical Computing* **55**: 275–286.
- Rambaut A. 2018.** *FigTree v.1.4.4*. Available at: <http://tree.bio.ed.ac.uk/software/figtree/>
- Renner MAM. 2020.** Opportunities and challenges presented by cryptic bryophyte species. *Telopea* **23**: 41–60.
- Renner MAM, Heslewood MM, Heinrichs J. 2018.** Geometric morphometric methods achieve type specimen assignment in the cryptic *Plagiochila arbuscula* complex (Plagiochilaceae: Jungermanniopsida) with the minimum of morphological evidence. *Botanical Journal of the Linnean Society* **186**: 108–128.
- Roads E, Longton RE, Convey P. 2014.** Millennial timescale regeneration in a moss from Antarctica. *Current Biology* **24**: R222–R223.
- Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017.** DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* **34**: 3299–3302.
- Rydin C, Pedersen KR, Friis EM. 2004.** On the evolutionary history of *Ephedra*: Cretaceous fossils and extant molecules. *Proceedings of the National Academy of Sciences, USA* **101**: 16571–16576.
- Schliep KP. 2011.** phangorn: phylogenetic analysis in R. *Bioinformatics* **27**: 592–593.
- Shaw J. 2001.** Biogeographic patterns and cryptic speciation in bryophytes. *Journal of Biogeography* **28**: 253–261.
- Shaw AJ, Golinski GK, Clark EG, Shaw B, Stenøien HK, Flatberg KI. 2014.** Intercontinental genetic structure in the amphipacific peatmoss *Sphagnum miyabeum* (Bryophyta: Sphagnaceae). *Biological Journal of the Linnean Society* **111**: 17–37.
- Sievers F, Higgins DG. 2014.** Clustal Omega. *Current Protocols in Bioinformatics* **48**: 3.13.1–3.13.16.
- Sim-Sim M, Afonina OM, Almeida T, Désamoré A, Laenen B, Garcia CA, González-Mancebo JM, Stech M. 2017.** Integrative taxonomy reveals too extensive lumping and a new species in the moss genus *Amphidium* (Bryophyta). *Systematics and Biodiversity* **15**: 451–463.
- Ślipiko M, Myszczyński K, Buczkowska K, Bączkiewicz A, Szczecińska M, Sawicki J. 2020.** Molecular delimitation of European leafy liverworts of the genus *Calypogeia* based on plastid super-barcodes. *BMC Plant Biology* **20**: 1–15.
- Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, Normandeau É, Laroche J, Larose S, Jean M, Belzile F. 2013.** An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS One* **8**: 1–9.
- Spagnuolo V, Terracciano S, Giordano S. 2009.** Clonal diversity and geographic structure in *Pleurochaete squarrosa* (Pottiaceae): different sampling scale approach. *Journal of Plant Research* **122**: 161–170.
- Stamatakis A. 2014.** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.
- Stech M, Quandt D. 2014.** 20,000 species and five key markers: the status of molecular bryophyte phylogenetics. *Phytotaxa* **9**: 196.
- Struck TH, Feder JL, Bendiksby M, Birkeland S, Cerca J, Gusarov VI, Kistenich S, Larsson KH, Liow LH, Nowak MD, Stedje B, Bachmann L, Dimitrov D. 2018.** Finding evolutionary processes hidden in cryptic species. *Trends in Ecology and Evolution* **33**: 153–163.
- Stubbs RL, Folk RA, Soltis DE, Cellinese N. 2020.** Diversification in the Arctic: biogeography and systematics of the North American *Micranthes* (Saxifragaceae). *Systematic Botany* **45**: 802–811.
- Swofford DL. 2003.** *PAUP\*. Phylogenetic analysis using parsimony (\*and other methods)*. Sunderland: Sinauer.
- Szövényi P, Terracciano S, Ricca M, Giordano S, Shaw AJ. 2008.** Recent divergence, intercontinental dispersal and shared polymorphism are shaping the genetic structure of amphipacific peatmoss populations. *Molecular Ecology* **17**: 5364–5377.
- Tallis JH. 1958.** Studies in the biology and ecology of *Rhacomitrium Lanuginosum* Brid.: I. Distribution and ecology. *Journal of Ecology* **46**: 271–288.
- Tripp EA, Tsai YHE, Zhuang Y, Dexter KG. 2017.** RADseq dataset with 90% missing data fully resolves recent radiation

- of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. *Ecology and Evolution* **7**: 7920–7936.
- Vanderpoorten A, Patiño J, Désamoré A, Laenen B, Górski P, Papp B, Holá E, Korpelainen H, Hardy O. 2019.** To what extent are bryophytes efficient dispersers? *Journal of Ecology* **107**: 2149–2154.
- Vigalondo B, Garilleti R, Vanderpoorten A, Patiño J, Draper I, Calleja JA, 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). *Molecular Phylogenetics and Evolution* **140**: 106598.
- Vitt DH, Marsh C. 1988.** Population variation and phytogeography of *Racomitrium lanuginosum* and *R. pruinosum*. *Beiheft zur Nova Hedwigia* **90**: 235–260.
- Werner O, Guerra J. 2004.** Molecular phylogeography of the moss *Tortula muralis* Hedw. (Pottiaceae) based on chloroplast *rps4* gene sequence data. *Plant Biology* **6**: 147–157.
- Westergaard KB, Zemp N, Bruederle LP, Stenøien HK, Widmer A, Fior S. 2019.** Population genomic evidence for plant glacial survival in Scandinavia. *Molecular Ecology* **28**: 818–832.
- Willeit M, Ganopolski A, Calov R, Brovkin V. 2019.** Mid-Pleistocene transition in glacial cycles explained by declining CO<sub>2</sub> and regolith removal. *Science Advances* **5**: 1–9.
- Yu Y, Harris AJ, Blair C, He X. 2015.** RASP (reconstruct ancestral state in phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution* **87**: 46–49.
- Zaccara S, Patiño J, Convey P, Vanetti I, Cannone N. 2020.** Multiple colonization and dispersal events hide the early origin and induce a lack of genetic structure of the moss *Bryum argenteum* in Antarctica. *Ecology and Evolution* **10**: 8959–8975.
- Zizka A, Silvestro D, Andermann T, Azevedo J, Duarte Ritter C, Edler D, Farooq H, Herdean A, Ariza M, Scharn R, Svantesson S, Wengström N, Zizka V, Antonelli A. 2019.** CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* **10**: 744–751.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Figure S1.** The composite likelihood and optimized number of migration edges ( $-m$ ) inferred by OptM from  $-m$  0 to 10. A, Mean composite likelihood  $L(m)$  with its SD (left axis) and the proportion of variance explained (right axis) of ten iterations of  $-m$  values. The iterations did not reach the 0.998 proportion threshold of the explained variance recommended to estimate migration edges in TreeMix analyses (see Materials and methods, GBS analyses, genetic structure analyses). B, Second-order rate of change for each  $-m$  value ( $\Delta m$  values). One migration edge was selected according to the highest  $\Delta m$  for this dataset.

**Figure S2.** Map of *Racomitrium lanuginosum* vetted records from GBIF portal consisting of 2421 occurrences.

**Figure S3.** Maximum-likelihood tree of 147 samples of *Racomitrium lanuginosum* based on ITS rooted on *R. pruinosum*. The three recovered clades are colour-coded. Only bootstrap values higher than 90 are presented. Several polytomies resulted within clades, and groups were collapsed instead.

**Figure S4.** Distribution map of *Racomitrium lanuginosum* clades recovered from ITS phylogenetic analyses. There is a low spatial genetic structure. Clade blue occurs more North in the Arctic and clades orange and green are more common in the sub-Arctic zones.

**Figure S5.** Maximum-likelihood tree of *Racomitrium lanuginosum* based on 605 SNPs found in at least 20% of 126 samples. Only groups with bootstrap values > 70 are represented in colours according to ITS-based clades as a topology comparison. Two clades were recovered but with a new group clustered within clade blue. The clades represented by orange and green ITS samples are still sister groups in this SNP-based phylogenetic tree.

**Figure S6.** Maximum-likelihood tree of *Racomitrium lanuginosum* based on 324 SNPs found in at least 40% of 126 samples ( $-R$  40). Only groups with bootstrap values > 70 are represented in colours according to ITS-based clades as a topology comparison. A group mainly composed of Alaskan samples in orange and some samples from the green ITS clade were well supported. The  $-R$  20 tree had better branch support than this phylogenetic tree.

**Figure S7.** 50% majority consensus tree of *Racomitrium lanuginosum* resulted from an SVDquartets coalescent-based analysis of 4620 loci shared by at least 20% of 126 samples ( $-R$  20). Only clades with bootstrap values > 70 are represented by the same colour code that genetic groups in previous analyses. Group C had low support and was not shown on the tree.

**Figure S8.** Map of ten geographic groups of *Racomitrium lanuginosum* and their nucleotide diversity. These results are based on 124 samples and 4181 loci shared by at least 60% of individuals per geographic area ( $-r$  60). The filters recovered 552 SNPs. Nucleotide diversity is presented according to a colour scale (top) for each geographic group based on SNP data. The Alaska Peninsula, the Scandinavian Mountains and Western Quebec have the highest nucleotide diversity. For detailed information on this map, refer to Supporting Information, Table S8.

**Figure S9.** TreeMix population graphs of *Racomitrium lanuginosum* geographic groups based on 406 unlinked SNPs of 124 samples using a  $-r$  50 per group filter. The window size is  $k$ -10, and one migration edge was estimated for the graph model. Two recurrent scenarios (5/10 replicates) are shown. A, Migration from the Alaska Peninsula to Western Quebec. B, Migration from an ancient hypothetical North American population to Western Quebec. In both cases, the Alaska Peninsula and the Scandinavian Mountains are more differentiated at the base of the graph.

**Figure S10.** The present distribution model of *Racomitrium lanuginosum* shows predicted suitable areas in green. The distribution threshold is based on a 10% training presence (AUC = 0.78). The currently suitable area is wider than areas predicted in the past, indicating a possible range expansion.

**Figure S11.** Last Glacial Maximum (c. 21 kya) distribution model of *Racomitrium lanuginosum* showing predicted suitable areas in green. Ice sheets are presented in blue shapes. The distribution threshold is based on a 10% training presence (AUC = 0.78). Ice-free areas such as Southern Alaska and Western Europe were predicted as suitable habitats.

**Figure S12.** Last Interglacial (c. 130 kya) distribution model of *Racomitrium lanuginosum* showing predicted suitable areas in green. The distribution threshold is based on a 10% training presence (AUC = 0.78).

**Table S1.** Internal transcribed spacer sequences of *Racomitrium lanuginosum* used for phylogenetic analyses indicating the dataset source, sequence name and code, herbarium code, voucher specimen, collection date, geographic area, locality, coordinates in latitude and longitude, haplotype, geographic group, samples randomly selected for sample-size corrected summary statistics clade represented by colour, ancestral area reconstruction analysis and accession number. Information that was not applicable, not found or missing is indicated with 'n.a.'.

**Table S2.** Model log-likelihoods of four tree priors under a strict and uncorrelated log-normal clock for BEAST analyses of *Racomitrium lanuginosum*. A total of eight models were evaluated using path sampling (PS) and stepping-stone (SS) approaches. The strict clock with a coalescent constant size prior tree was the model that better explained the data.

**Table S3.** Samples of *Racomitrium lanuginosum* used for de novo parameter selection tests in Stacks. The 12 selected samples are representatives of the complete dataset in terms of the number of reads, ITS-based clades and geographic areas.

**Table S4.** Data of 184 genotyping-by-sequencing samples of *Racomitrium lanuginosum* and congenics. Information includes the sequence name and code, herbarium code, voucher specimen, collection date, geographic area, locality, coordinates in latitude and longitude, genetic group based on fineRADstructure, geographic group, samples randomly selected for sample-size corrected summary statistics, accession number and sequencing technology. Additionally, we provide the alignment results with BWA using two available transcriptomes (*R. varium* and *R. elongatum*). Results include the number reads, mapped reads, proportion of mapped reads, recovered loci and mean coverage from gstacks and missing data for each sample using three different filters:  $-R$  00 (9815 SNPs) 184 samples,  $-R$  40 (411 SNPs) 127 samples and filter  $-R$  20 (844 SNPs) 127 samples. Samples with accession numbers correspond to those used for phylogeographic analyses.

**Table S5.** Genetic diversity and demographic statistics of *Racomitrium lanuginosum* geographic groups based on ITS marker. For each group, the number of samples, polymorphic sites (S), number of haplotypes (h), haplotype (Hd) and nucleotide ( $\pi$ ) diversity is presented along with their respective standard deviations (SD) and Fu & Li's  $D$  and Tajima's  $D$  statistics.

**Table S6.** Evolutionary timeframe for *Racomitrium lanuginosum* using ITS sequences under a strict clock and a coalescent constant size tree prior. The mean age in a million years and the 95% highest posterior density (HPD) are indicated.

**Table S7.** Mean  $F_{ST}$  pairwise comparison between genetic groups of *Racomitrium lanuginosum* based on fineRADstructure using loci shared by at least 60% of individuals per group ( $-r$  60). Group B has high values suggesting higher differentiation.

**Table S8.** Genetic summary statistics of ten *Racomitrium lanuginosum* geographic groups. These results are based on 124 samples and loci shared by at least 60% of individuals per geographic area ( $-r$  60). Filters recovered 4181 loci and 552 SNPs. For each geographic group is presented the number of samples, samples per loci, total, private and variant sites, single nucleotide polymorphisms (SNPs), the percentage of polymorphic loci (% PL). In addition, the mean haplotype (Hd) and nucleotide ( $\pi$ ) diversity and their standard errors (SE) were estimated for each group using SNPs. The Alaska Peninsula, the Scandinavian Mountains and Western Quebec have the highest genetic diversity.

**Table S9.** Mean  $F_{ST}$  pairwise comparison among geographic groups of *Racomitrium lanuginosum* with corrected sample sizes (eight samples) using loci shared by at least 60% of individuals per population ( $-r$  60). The Alaska Peninsula and Ontario have high values suggesting higher differentiation.