Morphological-molecular incongruence in Sphagnum majus ssp. majus and ssp. norvegicum

Authors: Nieto-Lugilde, Marta, Robinson, Sean, Aguero, Blanka, Duffy, Aaron, Imwattana, Karn, et al.

Source: The Bryologist, 125(2) : 294-310

Published By: The American Bryological and Lichenological Society

URL: https://doi.org/10.1639/0007-2745-125.2.294

Morphological-molecular incongruence in *Sphagnum majus* ssp. *majus* and ssp. *norvegicum*

Marta Nieto-Lugilde[1,7], Sean Robinson[2], Blanka Aguero[1], Aaron Duffy[1], Karn Imwattana[1], Kristian Hassel[3], Kjell Ivar Flatberg[3], Hans K. Stenøien[3], Anna V. Shkurko[4], Vladimir E. Fedosov[5,6] and A. Jonathan Shaw[1]

[1] *Department of Biology & L. E. Anderson Bryophyte Herbarium, Duke University, Durham, NC 27708, U.S.A.;* [2] *Department of Biology, SUNY Oneonta, Oneonta, NY 13820, U.S.A.;* [3] *Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway;* [4] *Tsitsin Main Botanical Garden, Moscow, Russia;* [5] *Moscow State University, Biological Faculty, Leninskiye Gory 1-12, Moscow, 119234, Russia;* [6] *Botanical Garden-Institute, FEB RAS, Makovskogo Street, 142, Vladivostok, 690024, Russia*

Abstract. Species delimitation is problematic in many plant groups and among the mosses, *Sphagnum* is one of the more contentious genera because of high levels of morphological variation. The allopolyploid species, *Sphagnum majus*, comprises one such problematic complex. Two morphologically differentiated but overlapping subspecies have been described. We conducted morphometric and molecular analyses with samples from around the Northern Hemisphere to test for phenotypic and phylogenetic differentiation between the subspecies. Although field collections of the two species can be statistically differentiated morphologically, there is substantial overlap. Genome-scale molecular data do not suggest any differentiation between *S. majus* ssp. *majus* and ssp. *norvegicum*, including samples assigned to the two taxa from sympatric sites. Sequence data from the plastid genome were employed to infer parentage of allopolyploid *S. majus*. Our results support the hypothesis that *S. annulatum* is the paternal parent and *S. cuspidatum* is the maternal parent. We conclude that the morphological differences between them are either plastic responses to habitat heterogeneity or segregating genetic variation within a single taxon. Formal taxonomic recognition of two taxa is not supported by our molecular data.

Keywords. Allopolyploidy, peatmosses, phylogenetic species, RADseq, subspecies.

◆ ◆ ◆

The biologically meaningful delimitation of species can be important to conservation biology, including policy decisions (Agapow et al. 2004), and is critical to any ecological or evolutionary based inquiry where species identification is required. The identification of units that are appropriate for species status has been the subject of numerous discussions ranging from philosophical and theoretical to methodological (De Queiroz 2007; Zachos 2018). The so-called Biological Species Concept (BSC) defines species as groups of organisms that can successfully interbreed and produce fertile offspring but are reproductively isolated from other species (Mayr 1942, 1963). Most botanists try to employ a phylogenetic approach to species delimitation although in many cases appropriate molecular data are lacking. In practice many or most species are recognized nomenclaturally because they "look different" and have distinct ecological and/or geographic distributions. Even with the best of data, determining the hierarchical *level* in a phylogenetic tree that suitably corresponds to the species category can be arbitrary (Mishler 2009; Wilson 1992). Objective methods have been developed for species delimitation based on patterns of molecular variation (e.g., Zhang et al. 2013) but these too are not without caveats and the application of names always involves some level of subjective judgement.

[7] Corresponding author's e-mail: marta.nietolugilde@duke.edu

DOI: 10.1639/0007-2745-125.2.294

0007-2745/22/$1.85/0

Some evolutionarily significant units appropriate for conservation consideration are below the species level because of genetic and morphological/ecological differentiation can be resolved within taxonomic species (Fraser & Bernatchez 2001; Moritz 1994). Disagreement about taxonomic ranks below species, including varieties or subspecies, has generated heated discussions among taxonomists (Mishler 2009; Wilson 1992). Species with a wide morphological diversity are at the center of these discussions because of difficulties establishing (or agreeing on) limits in a morphological (and ecological) continuum even when well-differentiated extremes can be identified (Barrett & Freudenstein 2011; Burley & Pritchard 1990; Nieto-Lugilde et al. 2018).

Species in the moss genus *Sphagnum* L. have been especially controversial with regard to species delimitation. For example, Crum (1984) recognized 51 species in North America whereas McQueen & Andrus (2007) recognized 89 species in their *Flora of North America* treatment for the genus. Recent molecular systematic research has led to conclusions that some morphologically defined *Sphagnum* taxa do not correspond to genetic/phylogenetic units worthy of species status (e.g., Duffy et al. 2020), and in other cases what was thought to be a single taxon is better understood as a complex of closely related species (Kyrkjeeide et al. 2016; Yousefi et al. 2017). *Sphagnum*-dominated peatlands have global impacts on ecology, hydrology, and biogeochemistry, and the genus has long served as a model for research in community ecology, so what we do or do not consider species has significant impacts for downstream research outside the field of systematics per se. Moreover, *Sphagnum* has recently become an important model for functional and ecological genomics because of a genus-wide sequencing effort being conducted by the Joint Genome Institute (JGI) of the U.S. Department of Energy (Weston et al. 2018).

Sphagnum subgenus *Cuspidata* contains a number of morphologically polymorphic complexes of closely related morphotypes that may or may not comprise multiple phylogenetic species. One such group centers around *Sphagnum majus* (Russow) C.E.O.Jensen. Morphological variation in *S. majus* was well studied by K. I. Flatberg (1987). He described variation relative to environmental gradients in Norwegian peatlands—"dry-wet" and "poor-rich"—through a study of some 300 herbarium specimens. He concluded with the recognition of two subspecies: *S. majus* ssp. *majus* and *S. majus* ssp. *norvegicum* Flatberg.

The subspecies of *Sphagnum majus* differ in color, branch and stem leaf shape and size, hyalocyst pore size and number on the outer surface of the divergent branch leaves, and spore size and granulation, but they show considerable morphological overlap (Flatberg 1987). The most consistent differences are that *S. majus* ssp. *norvegicum* has yellowish green to only somewhat brownish capitula, with broad and acute-obtuse stem leaves, broad branch leaves that are moderately involute, with pores on the outer surface rather few, fairly large (>1/3 cell width), in one irregular row. *Sphagnum majus* ssp. *majus* typically has brown capitula, with narrow and acute stem leaves, narrower branch leaves that are pronouncedly involute, with smaller pores (<1/3 cell width) in one or two irregular rows on the outer surface. The two subspecies are also said to have slightly different but overlapping habitat preferences and geography (Flatberg 1987). *Sphagnum majus* ssp. *majus* is almost exclusively found in minerotrophic mires, whereas *S. majus* ssp. *norvegicum* is more frequent in ombrotrophic mires. The broader circumboreal distribution of the *S. majus* subspecies is uncertain because of taxonomic uncertainties.

Crum's (1984) revision of North American *Sphagnum* was published before Flatberg described *S. majus* ssp. *norvegicum* and Crum attributed *S. majus* s.l. to both eastern and western North America, restricted to northern regions in Canada and Alaska. McQueen & Andrus (2007) report *S. majus* ssp. *majus* in Alaska and British Columbia, in eastern Canada, and in the United States south to New Jersey and Pennsylvania, in poor fens to ombrotrophic bogs. They document *S. majus* ssp. *norvegicum* from a broader range in western North America that includes the Northwest Territories, the Yukon Territory and eastward to Saskatchewan, and it is also widespread in eastern North America south to New Jersey and Pennsylvania. It is said to occur in "weakly minerotrophic mires" (McQueen & Andrus 2007).

Sphagnum majus is an allopolyploid species (i.e., diploid gametophytes, tetraploid sporophytes: Maass & Harvey 1973; Såstad et al. 2000; Temsch et al. 1998). Morphological similarity suggests a

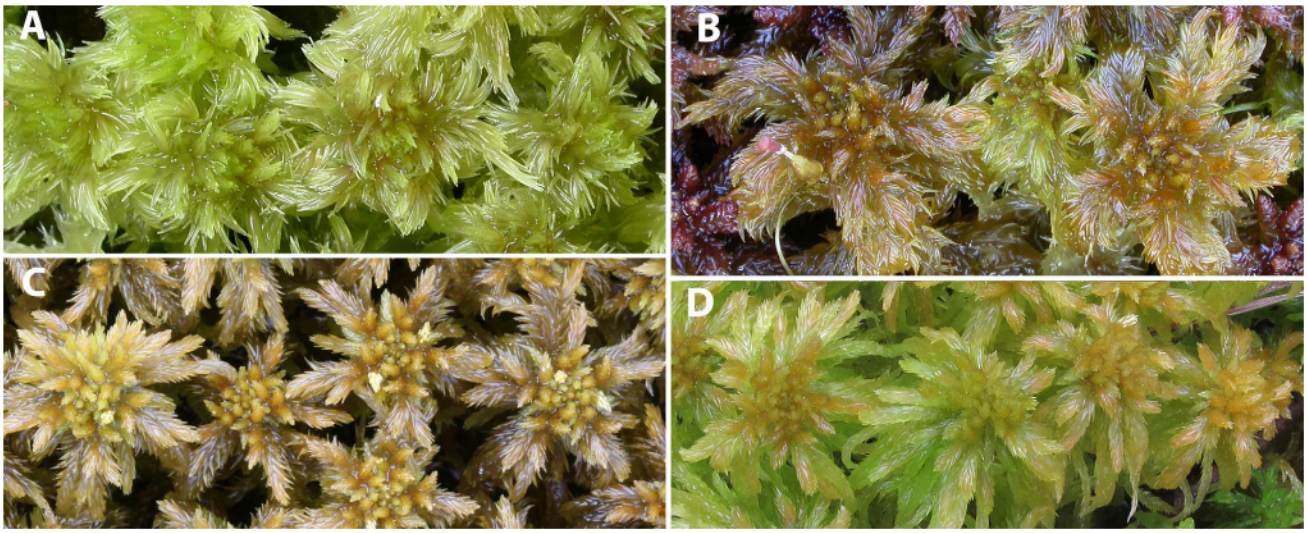

Figure 1. A. *Sphagnum cuspidatum*: CANADA. Newfoundland: On Hwy. 360, 78 km S of Hwy. 1, E side of road, *B. Shaw 6796*, 17 Aug. 2005. **B.** *S. majus* ssp. *majus*: CANADA. Newfoundland: On Hwy. 1, 3 km SW of junction with Hwy. 402, E side of road, *B. Shaw 6844*, 20 Aug. 2005. **C.** *S. annulatum*: U.S.A. Maine: Penobscot Co., Marble Fen, *B. Shaw 9681*, 17 Jun. 2009. **D.** *S. majus* ssp. *norvegicum*: U.S.A. New York: Warren Co., Along Carl Turner Rd. just E of I-87, S of Schroon Lake, 0.4 mi NE of jct. with Schroon River Rd. (private property on W side of rd.), *B. Shaw 18991*, 24 Jun. 2019.

relationship of *S. majus* to *S. cuspidatum* Ehrh. ex Hoffm. and *S. annulatum* H.Lindb. ex Warnst. (**Fig. 1**). Intraspecific genetic variation and interspecific relationships of *S. majus* were investigated by Såstad and co-workers (2000) studying 13 populations from Norway, Canada (Newfoundland), and the United States (New York) with nine isozyme loci. They found that the subspecific classification of *S. majus* was not supported by the observed patterns of genetic variation. Isozymes supported *S. cuspidatum* as a progenitor, but *S. annulatum* was fixed for enzyme bands not found in *S. majus*. The possibility that *S. balticum* (Russow) C.E.O.Jensen or another undetected or extinct taxon is a progenitor for *S. majus* could not be rejected. A recent study using RADseq data confirmed a close genetic/phylogenetic relationship of *S. majus* to *S. annulatum* but did not support a relationship to *S. balticum* (Duffy et al. 2020). Nevertheless, the placement of an allopolyploid such a *S. majus* in a phylogenetic reconstruction can be unpredictable and the derivation of *S. majus* from hybridization between *S. cuspidatum* and *S. annulatum* seems likely but not confirmed.

In this study we aimed to resolve parentage of allopolyploid *Sphagnum majus* and test the hypothesis that the two subspecies of *S. majus*, namely ssp. *majus* and ssp. *norvegicum*, represent distinct clades based on molecular data that correlate with morphological variation. We include samples from a broad range around the North Hemisphere, including localities where the two morphotypes both occur at the same site, and use high-resolution population genomic data (RADseq) to assess relationships. We measured morphological traits thought to distinguish the two taxa so we could relate molecular and morphological patterns of variation. Note that we use "*S. majus*" to refer to both *S. majus* subspecies, and we specify *S. m. majus* or *S. m. norvegicum* when we are specifically referring to only one of the two.

Materials and Methods

Plant material. A total of 110 samples were included in this study; voucher information is provided in **Supplementary Table S1**. These included 44 samples determined morphologically as *Sphagnum majus majus* and 47 as *S. m. norvegicum*, distributed across the Northern Hemisphere (**Fig. 2**). Fifty-three collections were from Europe (of which 25 were from Norway, 13 from European Russia, four from the United Kingdom, four from Belgium, three from Finland, two from Estonia, one from Austria, and one from France), 31 from eastern North America, four from western North America, and three from central Russia. Twenty-eight samples (31%) came from 11 sites with both subspecies. Five samples were included from each possible parental species, *S. annulatum*, *S.*

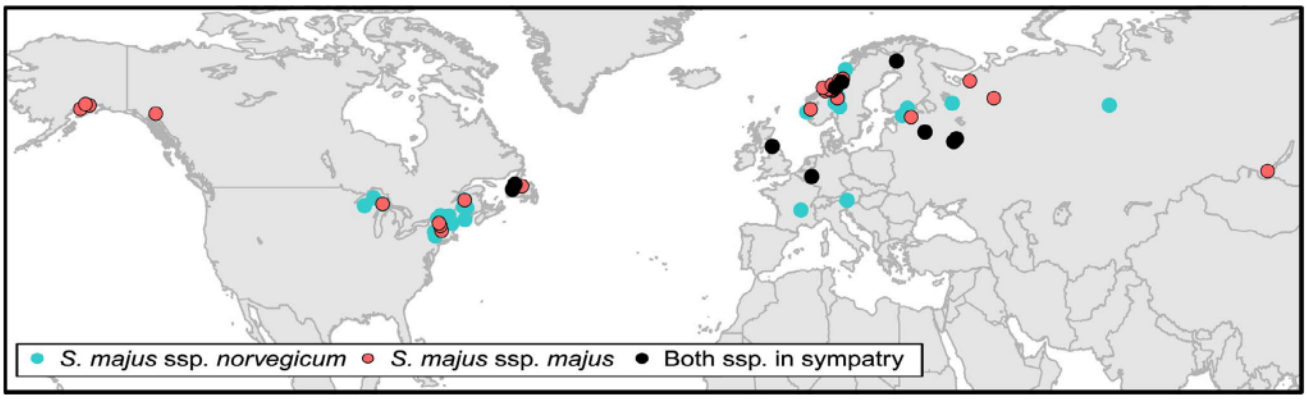

Figure 2. Geographic locations of *Sphagnum majus* samples included in this study.

balticum and *S. cuspidatum*. Two samples of *S. pulchrum* (Lindb.) Warnst. and two of *S. riparium* Ångstr. were used as outgroups. A single capitulum was sampled from each plant and the remaining tissue from sampled plants are archived in the Duke University herbarium (DUKE).

Biometric study. We measured 16 morphological characters (**Table 1**), selected following the results of Flatberg (1987) and our own observations. Of these, three were qualitative, plus 13 quantitative characters. From each collection, one gametophyte shoot was sampled. Three leaves from the middle of one divergent branch from the upper part of the shoot just below the capitulum from each gametophyte were dissected and the following characters were studied: lamina length and width, leaf habit and lamina length/width. Three basal cells and three apical cells from each leaf were selected and the following traits were measured: cell length and width, number of pores per fibril interval, total number of pores per cell, pore diameter and pore diameter/cell width ratio.

Statistical analyses. We measured morphological traits using a light microscope (Olympus BX41) and data were imported into R (R Core Team, 2020). For the evaluation of morphological characters, the specimens were grouped by subspecies based on preliminary assessments of morphological patterns. We used nonparametric Wilcoxon tests to compare differences between subspecies with signif-

Table 1. Morphological characters included in the biometric study, with indication of the number of items studied, type (QL = qualitative, QT = quantitative) and status characters.

	Character	Type and status character
Leaf	Leaf length	QT (mm)
(3 leaves per plant)	Leaf width	QT (mm)
	Leaf habit	QL: Straight to usually strongly subsecund (1); Straight to slightly subsecund (2)
	Leaf width/Leaf length ratio	QT
Basal cell	Cell length	QT (µm)
(9 basal cells: 3 from each leaf coming from 1 plant)	Cell width	QT (µm)
	Predominant number of pores per fibril interval	QL: Zero (0), One (1); Two (2)
	Total number of pores per cell	QT (integer)
	Pore diameter	QT (µm)
	Pore diameter/Cell width	QT
Apical cell	Cell length	QT (µm)
(9 mid-media cells: 3 from each leaf coming from 1 plant)	Cell width	QT (µm)
	Predominant number of pores per fibril interval	QL: One (1); Two (2)
	Total number of pores per cell	QT (integer)
	Pore diameter	QT (µm)
	Pore diameter/Cell width	QT

icance levels subjected to Benjamini-Hochberg corrections. The set of R functions for morphometric analysis was based on Koutecký (2015). The total number of pores per basal cell were log transformed before analysis. The "number of pores per fibril interval" was split as two binary characters: presence/absence and if present, one/two. To avoid highly correlated characters ($r > 0.95$), we calculated a Pearson matrix of the correlation coefficients of the characters. Multivariate analysis of variance (MANOVA) was used to assess the statistical significance of group differences by considering all of the variables simultaneously. Principal Component Analyses (PCA) were conducted on all traits (not shown), and also on just those traits for which the MANOVA indicated statistical significance of differences between the *S. majus* subspecies.

DNA extraction, library preparation and sequencing. Because methods employed here were modified relative to those described in our previous studies (Duffy et al. 2020), we describe our protocol in some depth. Approximately 100mg of dried tissue was employed for genomic DNA extractions following a modification of Doyle & Doyle's (1987) CTAB protocol described here. Two stainless steel ball bearings were added to each tube with the plant material, frozen with liquid nitrogen for 5–10 seconds. The frozen tissue was ground in a Genogrinder at the 1× setting, speed 500 rps for 1.5 minutes. 500µL CTAB isolation buffer (500 µL of 2× stock CTAB buffer, 0.02 g PVP-40 and 2 µL β-mercaptoethanol) was added to each sample, heated to 60°C, and incubated for 60 minutes. During incubation tubes were gently mixed by inverting every 20–30 minutes. An equal volume of chloroform-isoamyl (24:1) was added and was mixed gently by inversion (30–50×) to produce an emulsion. The emulsified solution was centrifuged for 5 minutes at 4500 rpm. The aqueous phase was transferred to a new tube with wide-cut tips to reduce DNA shearing, to which an equal volume of ice-cold isopropanol was added and mixed by inversion. DNA was then precipitated at –20°C for 30 minutes. Tubes were centrifuged for 20 minutes at 4500 rpm. The pellet was washed with 500 µL of cold 70% ethanol, and the tubes were then centrifuged for 5 minutes at 4500 rpm. This step was repeated 3 times. The pellet was air-dried overnight. Finally, the pellet was resuspended in 25 µL DEP water. DNA concentrations were measured using a Qubit 2.0 Fluorometer (Life Technologies) and standardized to 20ng/µL.

Genomic libraries were made following the double digestion restriction site-associated DNA sequencing (ddRADseq) protocol of Parchman et al. (2012), with modifications described by Duffy et al. (2020). The libraries were cleaned and size-selected for fragments of approximately 350bp using AMPur XPbeads (Beckman Coulter), checked for quality on a BioAnalyzer (Agilent), and sequenced on a single lane of Illumina NovaSeq 6000 with 150bp single-end reads at the Genome Sequencing Shared resource operated by the Duke Center for Genomic and Computational Biology (https://oit.duke.edu/comp-print/research/).

RADseq data pipeline. SNP discovery was performed with ipyrad v.0.7.29 (Eaton 2014) using default parameters except as noted here. Reads were processed as datatype "ddrad" to match the library preparation method. A maximum of two mismatched bases were allowed in the barcode during demultiplexing, Illumina adapter sequences and low-quality bases were trimmed from the reads and trimmed reads less than 92 bases long or with more than five low quality bases were discarded. Multiple ipyrad runs were performed using a range of read clustering thresholds to identify the clustering threshold (0.90) that maximized the number of variable and parsimony informative loci and to verify that the results of downstream analyses are not sensitive to clustering threshold. Only loci present in at least 80% of the samples were kept for final analyses. Separate assemblies for *S. majus*, coded as diploids, and the remaining species (*S. annulatum*, *S. balticum*, *S. cuspidatum* and *S. pulchrum*), coded as haploids, were constructed for estimates of heterozygosity, error rates, and consensus base calling. The parameter that differed for haploids versus diploids was the maximum number of alleles allowed in individual consensus reads after accounting for sequencing errors. Only one allele was permitted for haploids and two for the diploids. The resulting files from these steps where then merged to apply the final clustering, filtering and file-formatting steps of the pipeline on a single dataset. We also generated multiple data sets in the final step of the ipyrad pipeline based on different subsets of the data, to assess if SNP calling based on these various subsets impacted downstream inferences. In the first data set we dropped two

outgroups species (*S. pulchrum* and *S. riparium*). For the second data set we kept just *S. majus*. These data sets yielded consistent results. The sequences, phylogenetic alignment, and STRUCTURE files analyzed in this study are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.r7sqv9sdv (Nieto-Lugilde et al. 2022).

Phylogenetic analyses. IQ-TREE 2.0.3 (Minh et al. 2020) was used to infer phylogenetic relationships among sequences by maximum likelihood (ML) using concatenated loci identified by ipyrad for the complete data set. The ML tree was estimated using random starting trees and the ultrafast bootstrap approximation (UFBoot: Hoang et al. 2018; Minh et al. 2013). The best nucleotide substitution model was estimated using the ModelFinder algorithm as implemented in IQ-TREE v 2.0.3 (Minh et al. 2020) and chosen according to the Bayesian information criterion (BIC). The rapid hill-climbing search algorithm was used to estimate the best ML tree using 1000 bootstrap replicates to determine support for branches. To visualize potential admixture between samples in the *Sphagnum majus* complex, a network was constructed using the NeighborNet algorithm implemented in SplitsTree v. 4.17 (Huson & Bryant 2006), and a bootstrap analysis was performed with 1,000 replicates.

To investigate maternal parentage of allopolyploid *Sphagnum majus* we assumed that the plastid genome is inherited maternally (Natcheva & Cronberg 2007). The reference genome for *S. angustifolium* (Russow) C.E.O.Jensen (Phytozome v12: https://phytozome.jgi.doe.gov/) was used to identify plastid sequences in our RADseq data. We mapped each RADseq locus to that reference genome, allowing up to nine mismatches, with the clustering threshold for identifying loci set to 90% in ipyrad. We concatenated the plastid loci and phylogenetic relationships were inferred with IQ-TREE 2.0.3 (Minh et al. 2020) as used above.

Population structure. STRUCTURE 2.3.4 (Pritchard et al. 2000) was used to assess any detectable genetic structure within *Sphagnum majus*. Two sample partitions were used, (I) samples of *S. majus* and the three putative parental species, *S. annulatum, S. balticum,* and *S. cuspidatum*, and (II) a subset of samples that included only *S. majus*. One SNP per locus was randomly selected to avoid including tightly linked SNPs. STRUCTURE does not permit mixed haploid-diploid models in a single analysis, so all samples were coded diploid, with the haploids coded as homozygous. The analyses were also conducted with all samples coded as haploid (results not shown). The best number of clusters (K) was evaluated using the ΔK according to method of Evanno et al. (2005) implemented in STRUCTURE HARVESTER v0.6.94 (http://taylor0.biology.ucla.edu/structureHarvester/, Earl & vonHoldt 2012) based on 10 independent runs using an admixture model with correlated allele frequencies for each K from 1 to 10 with 50,000 steps of burn-in and 500,000 steps per run. Regardless of how this method evaluated the "optimal" K, we explored different levels of K to assess the possibility of additional structure in the data. In addition to the entire *S. majus* complex, STRUCTURE was also used on a subset of the samples to explore finer-scale genetic structure considering the subspecies of *S. majus* separately. The web application STRUCTURE PLOT v2.0 (http://omicsspeaks.com/strplot2/; Ramasamy et al. 2014) was used to visualize the STRUCTURE results.

Summary statistics for genetic diversity were calculated in VCFtools v0.1.16 (Danecek et al. 2011) using the same data set as analyzed using STRUCTURE. We estimated the following parameters: mean nucleotide diversity per site (Pi) and Tajima's D. Fst (Weir & Cockerham 1984) was estimated among species (and putative subspecies).

Results

Molecular data characterization. After demultiplexing, trimming, removing barcodes and adapter sequences, filtering for quality, and removing samples with low read counts, 303 million reads of 35 to 92 bp were retained across 88 individual plant samples (69 samples of *Sphagnum majus*, 5 each of *S. annulatum*, *S. balticum*, and *S. cuspidatum,* and 2 each of *S. pulchrum* and *S. riparium*), with the number of reads per individual ranging from 247,591 to 6,993,163 (median ± SD = 3,716,434 ± 1,623,604). In a preliminary analysis, the 90% clustering similarity threshold generated the highest number of loci and parsimony informative sites and was therefore used to construct the final data sets for further analysis.

This assembly pipeline produced 6692 loci distributed across the genome (**Supplementary Fig. S1**); the final matrix included 442,082 bp, and 6240 of those loci contained one or more SNPs. 5915 contained one or more phylogenetically informative SNPs (i.e., those shared by two or more individuals). The mean locus coverage per individual was 86.45%.

Phylogenetic reconstructions. Iqtree estimated TVM+F+R10 as the best-fit substitution model according to the BIC (Bayesian Information Criterion). Samples identified morphologically as *Sphagnum magus majus* and *S. m. norvegicum* are not resolved by the molecular data, nor do their geographic provenance correspond to phylogenetic relationships. Samples of the two subspecies from the same locality, and in some cases growing a few centimeters from each other, show little or no differentiation (**Fig. 3**). Rooted with *S. riparium*, the *S. majus* clade, without distinguishing subspecies, is resolved as sister to *S. cuspidatum*, this clade is sister to *S. annulatum,* and this in turn to *S. pulchrum*. *Sphagnum majus* does not appear to be closely related to *S. balticum* (**Fig. 3**).

SplitsTree network analyses also recover each of the species as monophyletic, but does not resolve the two subspecies of *Sphagnum majus* (**Fig. 4**). As with the ML tree, *S. majus* shares more recent common ancestors with *S. cuspidatum* and *S. annulatum* than it does with *S. pulchrum* or *S. balticum*. SplitsTree indicates substantial conflict in the data, especially between *S. majus*, *S. cuspidatum*, and *S. annulatum*, consistent with the latter being ancestral to allopolyploid *S. majus* (**Fig. 4**).

Parentage of allopolyploid S. majus. Mapping RADseq data to the genome of *Sphagnum angustifolium* yielded eight loci that mapped to the plastid genome, for a total of 584bp (of 37 to 86 bp). In total, only eight variable and 5 parsimony informative sites were identified. BIC identified the best-fit substitution model as TPM2u+F for these data. The ML tree indicates that *S. majus* shares identical plastid sequences with *S. cuspidatum* (**Fig. 5**). Support values of the clades of *S. annulatum* and *S. pulchrum* based on plastid data alone are, not surprisingly, quite low (52–56% respectively).

Population structure and genetic diversity. When all species are analyzed as diploids, STRUCTURE analyses suggest that K=2 genetic groups is optimal. All *Sphagnum majus* samples belong to the orange group in **Fig. 6A**, *S. balticum* belongs to the yellow group, and both *S. cuspidatum* and *S. annulatum* are both more or less equally admixed for those two genetic groups. At K=4, all four species are resolved as separate genetic groups but the two subspecies of *S. majus* are still not distinguished. In analyses of *S. majus* alone, K=3 is considered optimal and all three groups are present in plants identified as both *S. m. majus* and *S. m. norvegicum* (**Fig. 6B**). The yellow genetic group is predominant in samples of *S. majus* (both subspecies) from eastern North America and that group is absent from other samples or represented by a small minority component in admixed individuals. At K=2 (not considered optimal), the eastern North American samples are comprised completely or predominantly by the orange genetic group. Some samples from outside eastern North America show minority admixture with that group; these are the same samples that appear to show a genetic relationship with eastern North American *S. majus* plants in the analyses of all four species (**Fig. 6A**).

Genetic diversity as estimated by π is highest in *Sphagnum cuspidatum* despite the small sample size for this species, and lowest in *S. annulatum* (**Table 2**). The two subspecies of *S. majus*, and *S. balticum*, have intermediate levels of genetic diversity. Tajima's D is strongly positive in the haploid species but is negative in *S. majus* (**Table 2**). Combining *S. m. majus* and *S. m. norvegicum* as one taxon does not increase the level of genetic diversity above the levels characterizing each subspecies individually, confirming that their genetic makeup is very similar. The four species, *S. majus, S. annulatum*, *S. cuspidatum* and *S. balticum*, are well-differentiated from one another, with Fst values of 0.7–0.9 (**Table 3**). In contrast, Fst between the two *S. majus* subspecies is essentially zero.

Biometric analyses. Results of the morphological analyses are summarized in **Fig. 7**, and numerical values (means ± SD) are provided in **Supplementary Table S2**. Thirty-seven out of 136 pairs of traits (27%) are significantly correlated (**Supplementary Table S3**). Among the correlated traits are leaf length and width, suggesting size variation. In fact, leaf size appears to be correlated with cell length, and with pore size. *Sphagnum m. majus* and *S. m. norvegicum*, preliminarily identified mainly by color, are discriminated by the MANOVA

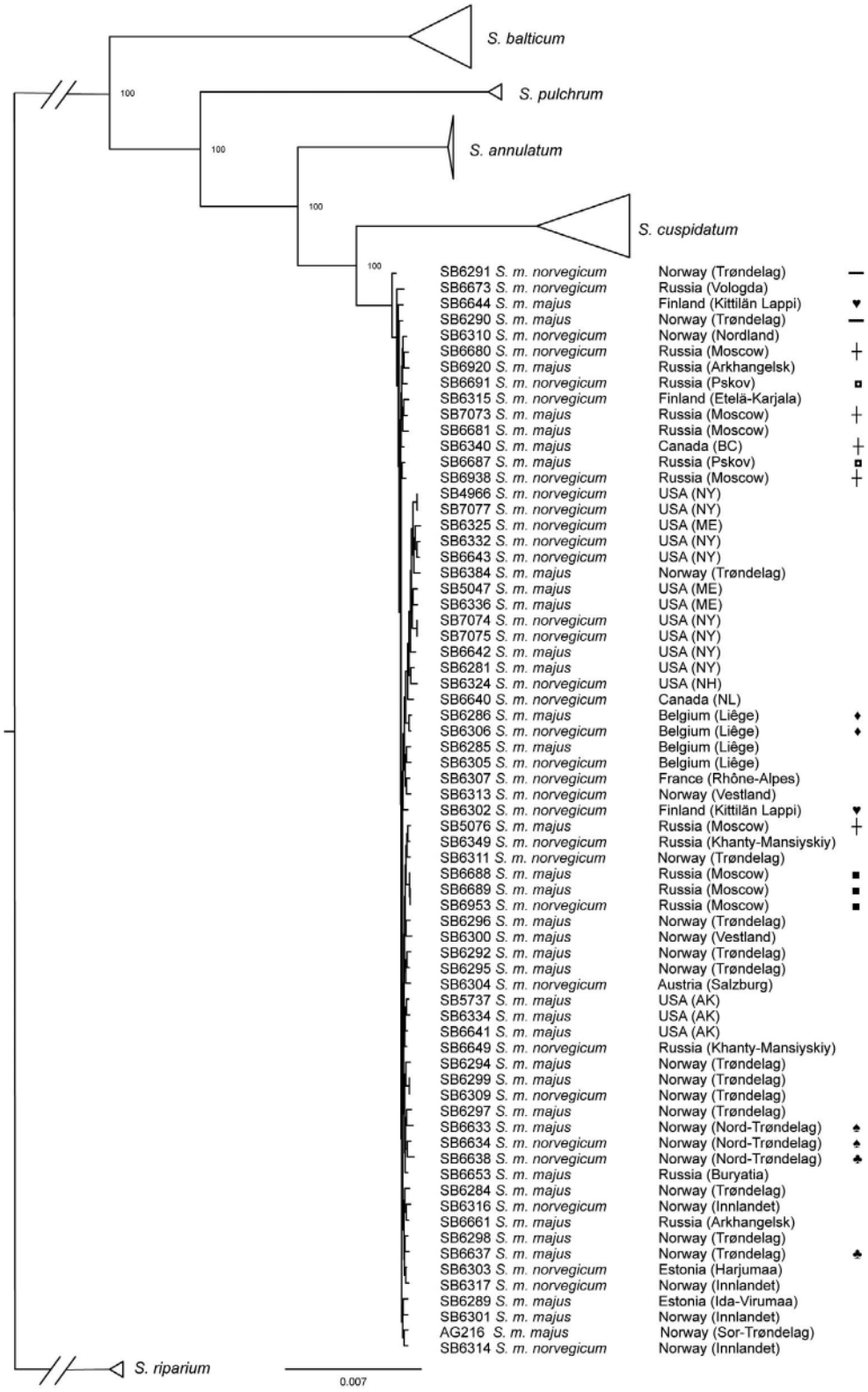

Figure 3. Phylogenetic relationships among *Sphagnum majus* and related species based on RADseq loci. Nodes supported at >95% are labeled with support values. Symbols on the right mark samples from sites where both subspecies were collected; samples with the same symbol were collected from the same locality (for locality details see Supplementary Table S1).

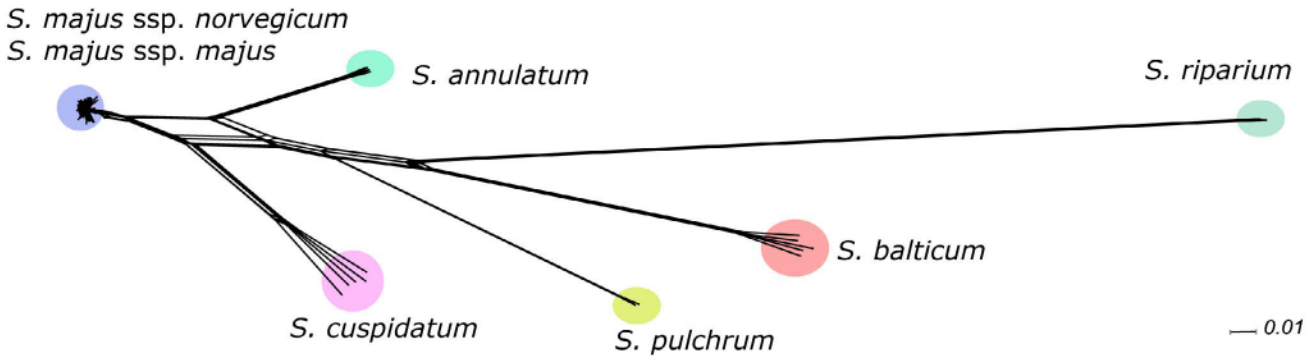

Figure 4. Results of SplitsTree network analyses (Neighbor-Net algorithm) for *Sphagnum majus* and related species.

($p = 0.05$, **Table 4**). The two subspecies differ ($p < 0.05$) in six individual branch leaf traits: leaf length, leaf width/length ratio, apical cell length, total number of pores per cell, pore diameter, and pore diameter/cell width ratio (**Supplementary Table S4**). Wilcoxon's tests also confirm clear differences between *S. majus* subspecies (**Supplementary Table S2**).

In the PCA, 59.4% of the total variance is explained by the first two PCs (34.8% PC1, 24.6% PC2). PC3 explained 16.6% and the remaining below 6.5%. PC1 separates *Sphagnum majus majus* from *S. m. norvegicum*, but with substantial overlap (**Fig. 8A**), while PC2 do not clearly separate any taxonomic group. However, there seems to be a minor tendency (with exceptions) for samples from North America to present positive values relative to PC2 while the European samples tend to have negative values (data not shown). Considering the sites where the subspecies co-occur, we detected a high degree of variation in morphological patterns. All UK samples (indicated with a "slash" symbol in **Fig. 8A**), for example, are located in the morphological space of overlap between the subspecies. In contrast, mixed stand samples of the two subspecies from localities in Norway (Lurudalen: "spades"), Finland (Tuulijoki mire: "heart"), and Canada (Gros Morne National Park: "number") differ morphologically. Co-occurring samples from localities in Belgium (Wames: "diamond") and Canada (Western Region: "asterisk") identified as the two subspecies fall within the range of variation characterizing the *S. m. norvegicum* cluster, whereas plants from two populations in Norway (Flamyra: "dash"; and Forra: "club") fall within the *S. m. majus* cluster. The relationships of individual morphological traits to the PCA axes are summarized in **Fig. 8B**.

Discussion

This research demonstrated a set of branch leaf morphological characters that support the recognition of *Sphagnum majus majus* and *S. m. norvegicum*. Some intermediate plants exist but the two are statistically differentiated in morphology. Flatberg (1987) suggested that the subspecies differ in stem leaf traits as well, but we did not include these in our morphometric analyses. Our results nevertheless indicate that the taxa are statistically distinguishable with the traits we measured. In contrast to the evidence suggesting that the subspecies differ in morphology, molecular data indicate that these morphotypes do not form monophyletic groups. It therefore appears that the subspecific morphological variation has segregated repeatedly within a broader monophyletic species. Both *S. m. majus* and *S. m. norvegicum*, defined morphologically, occur in the plastid DNA clade that includes all samples of both *S. cuspidatum* and *S. majus*. These results demonstrate unambiguously that *S. cuspidatum* is the maternal parent of allopolyploid *S. majus*.

Although common garden and/or reciprocal transplant experiments would be necessary to show that the differentiating morphological traits are genetically based, at least in part, the fairly clear resolution of two groups rather than a continuum suggests that there is some degree of genetic determination. There is also very likely a genotype × environment interaction such that expression of the traits, even for a given genotype, is variable across environments. Flatberg (1987) already identified wet-dry and nutrient richness gradients as potentially important. The observation that these

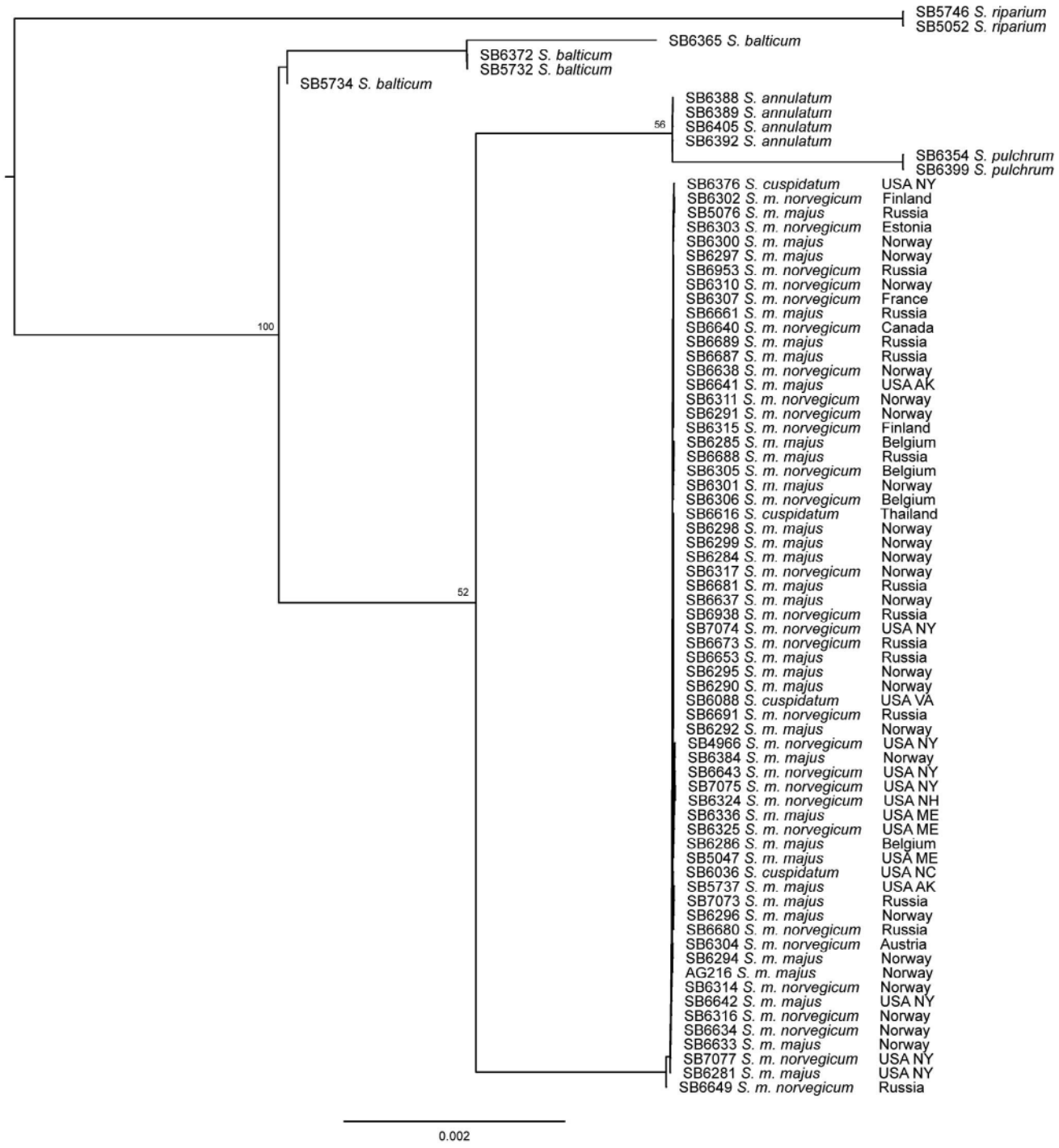

Figure 5. Phylogenetic relationships among plastid DNA sequences from samples of *Sphagnum majus* and related species. See Methods for details of how these data were obtained from RADseq results by mapping reads to the *S. angustifolium* reference genome.

morphotypes often differ in habitat may support the view that there is an underlying environmental component to the morphological differences.

At sites where the two putative subspecies have been identified by collectors as occurring at the same sites, morphological differentiation between them is most often weak, even though on a global scale the morphotypes are statistically distinguishable. Among collectors, there seems to be agreement that as field-growing plants, the subspecies differ in color

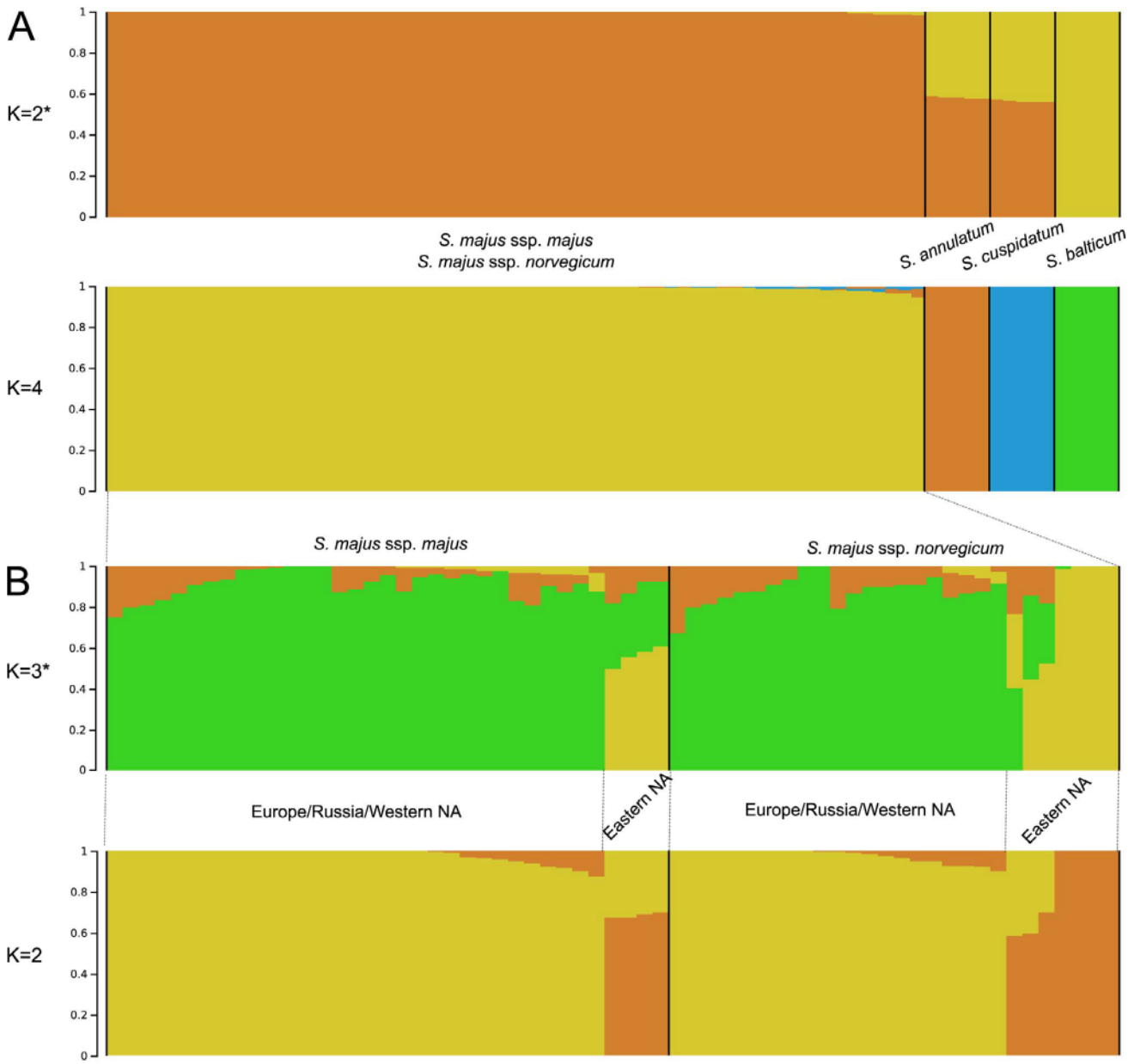

Figure 6. Results of STRUCTURE analyses of RADseq loci. **A.** *Sphagnum majus* and putative parental species. **B.** *S. majus* subspecies. For each analysis, the optimal K-value (indicated with *) is presented and another K-value when it provides additional clustering information. *S. annulatum*, *S. balticum* and *S. cuspidatum* were not analyzed separately due to small sample sizes.

and shape of the capitula, but microscopic differentiation between them is less clear (Flatberg 1987; Graulich 2021). Moreover, within-plant morphological variation has been observed in traits that, on average, distinguish the subspecies (Flatberg 1987; this study). Flatberg suggested that seasonal fluctuation in water levels may induce variation in branch leaves, but we observed this variation even in leaves from a single branch. If changes in water level impact within-branch leaf morphology the temporal scale of such genotype × environment interaction would need to be very fine. *Sphagnum majus* is a dioicous allopolyploid, and sexual leaf dimorphism seems to occur, as in most *Sphagnum* species (Flatberg 1987). Male plants tend to have relatively broader stem leaves and divergent branch leaves than the female plants, while the stem leaves are slightly longer and the divergent branch leaves

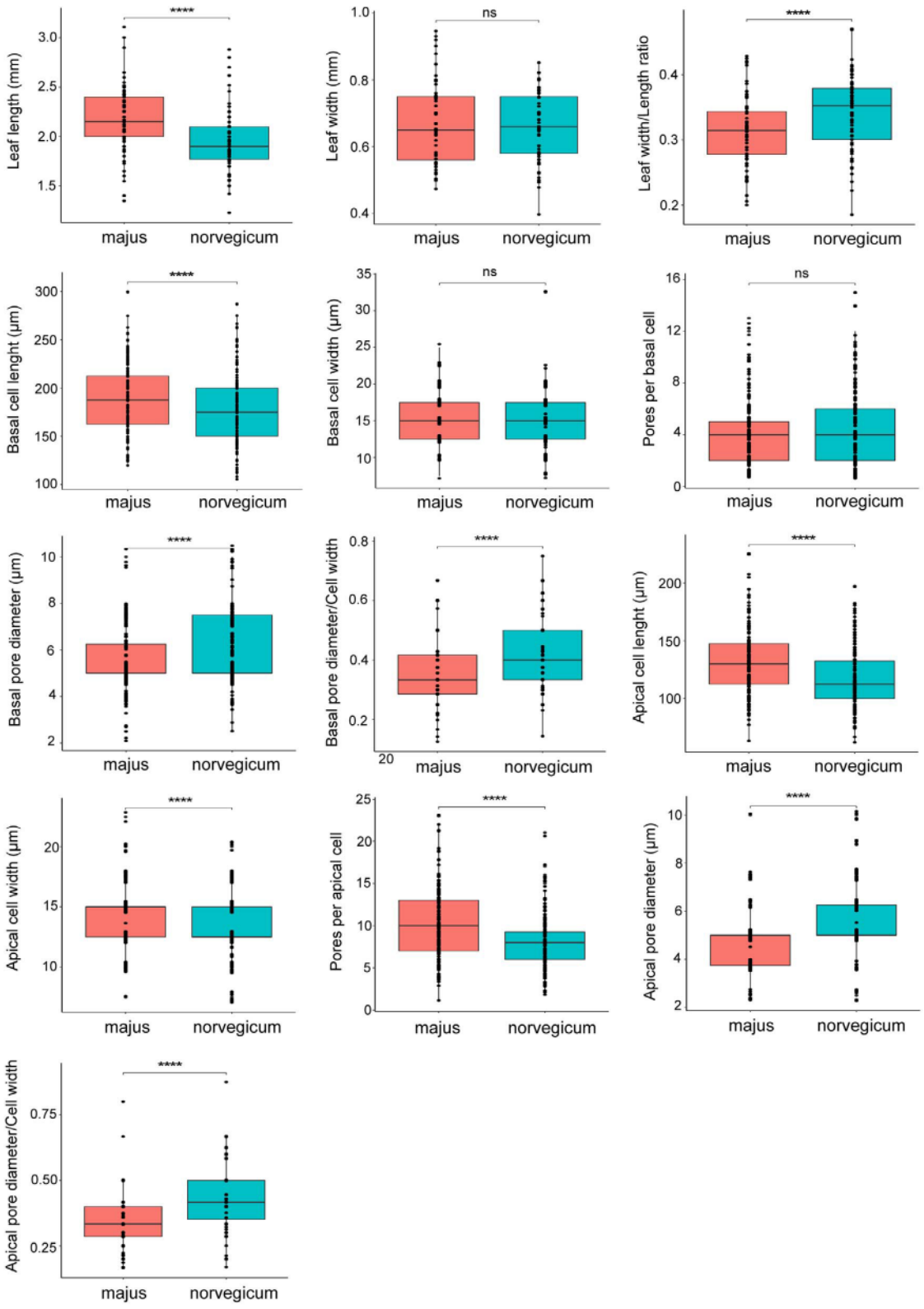

Figure 7. Box-plots of quantitative morphological characters studied for *Sphagnum majus* ssp. *majus* and *S. majus* ssp. *norvegicum*. Not significant *p*-value and significant *p*-value <0.05 after being corrected with Benjamini-Hochberg test are indicated with ns and ****, respectively.

Table 2. Number of individuals and genetic diversity statistics (mean nucleotide divergency (π), and Tajima's D statistic) for each species.

Species	N	π	Tajima's D
S. majus	63	0.056	-0.20
S. majus ssp. *majus*	35	0.055	-0.11
S. majus ssp. *norvegicum*	28	0.056	-0.13
S. cuspidatum	5	0.104	0.56
S. annulatum	5	0.005	0.64
S. balticum	5	0.064	0.64

Table 4. MANOVA of morphological characters included in the biometric study. Values with statistically significant difference ($\alpha<0.05$) are indicated with *.

	Df	Pillai	Approx. F	num Df	den Df	Pr (>F)
(Intercept)	1	0.99984	10478.9	17	29	<2e-16 *
Subspecies	1	0.58623	2.4	17	29	0.01767 *
Residuals	45					

slightly shorter; The stem leaves of male plants are more obtuse at the apex than of the female plants, and the hyalocyst pores on convex surfaces of the divergent branch leaves are smaller in male than in female plants (Flatberg 1987). The extent to which differences between male and female gametophytes could contribute to the variation patterns we (and Flatberg 1987) observed is unknown as we were unable to assess the sex of most samples in our data set.

Samples of *Sphagnum majus* from eastern North America are genetically differentiated, as indicated by our STRUCTURE results. Eastern American plants are characterized by a predominant genotype group that is rare or absent in plants sampled from other areas. No other geographic patterns were detected, and phylogenetic relationships among samples at that level near the tips of the tree were generally unsupported.

Divergent inferences from morphological vs. molecular data have been described in other groups within *Sphagnum* subgenus *Cuspidata*. Species recognized because of morphological variation where taxonomic separation was not supported by molecular data include *S. atlanticum* R.E.Andrus vs. *S. torreyanum* Sull. (Shaw et al. 2009), the separation of *S. isoviitae* Flatberg and *S. brevifolium* (Lindb. ex Braithw.) Röll from *S. fallax* (H.Klinngr.) H.Klinngr. (Duffy et al. 2020; Såstad 1999), and *S. cuspidatum* vs. *S.* viride Flatberg (Hanssen et al. 2000). Color morphs have also been studied in *S. palustre* L. (Stenøien et al. 2014), but as in *S. majus*, these morphs did not correspond to molecular variation. The lack of molecular differentiation between putative taxa is of course negative evidence and could reflect insufficient data rather than a real absence of differentiation. Molecular evidence based on sequences for a single gene, or even a few genes, are suspect in this regard. Conservative plastid (or mitochondrial) genes may not show divergence when faster evolving nuclear genes do or would. Older isozyme data are very conservative, notwithstanding that many species can be differentiated at isozyme loci. In the present case, we note that our analyses were based on over 400,000 nucleotides from over 6000 short DNA fragments distributed across all chromosomes in the genome, and that these data were sufficient to resolve subtle differentiation of eastern North American *S. majus* plants. We thus argue that the data are amply sufficient to observe differentiation between *S. m. majus* and *S. m. norvegicum* if it existed.

Såstad et al. (2000) hypothesized that the two most likely parents for allopolyploid *Sphagnum majus* are *S. cuspidatum* and *S. annulatum*, although they suggested that the second parent could be *S. balticum* rather than *S. annulatum*. *Sphagnum balticum* is thought to be one parent of a somewhat similar allopolyploid, *S. jensenii* H.Lindb., along with *S. annulatum* (Såstad et al.

Table 3. Weir and Cockerham mean (lower triangle) and weighted (upper triangle) Fst estimates for *Sphagnum majus* and related species.

Species	*S. majus* ssp. *majus*	*S. majus* ssp. *norvegicum*	*S. cuspidatum*	*S. annulatum*	*S. balticum*
S. majus* ssp. *majus	—	0.0037	0.7580	0.8186	0.8941
S. majus* ssp. *norvegicum	0.0007	—	0.7426	0.8117	0.8885
S. cuspidatum	0.4904	0.4721	—	0.8532	0.8207
S. annulatum	0.4943	0.4900	0.6546	—	0.9152
S. balticum	0.6610	0.6512	0.6166	0.8005	—

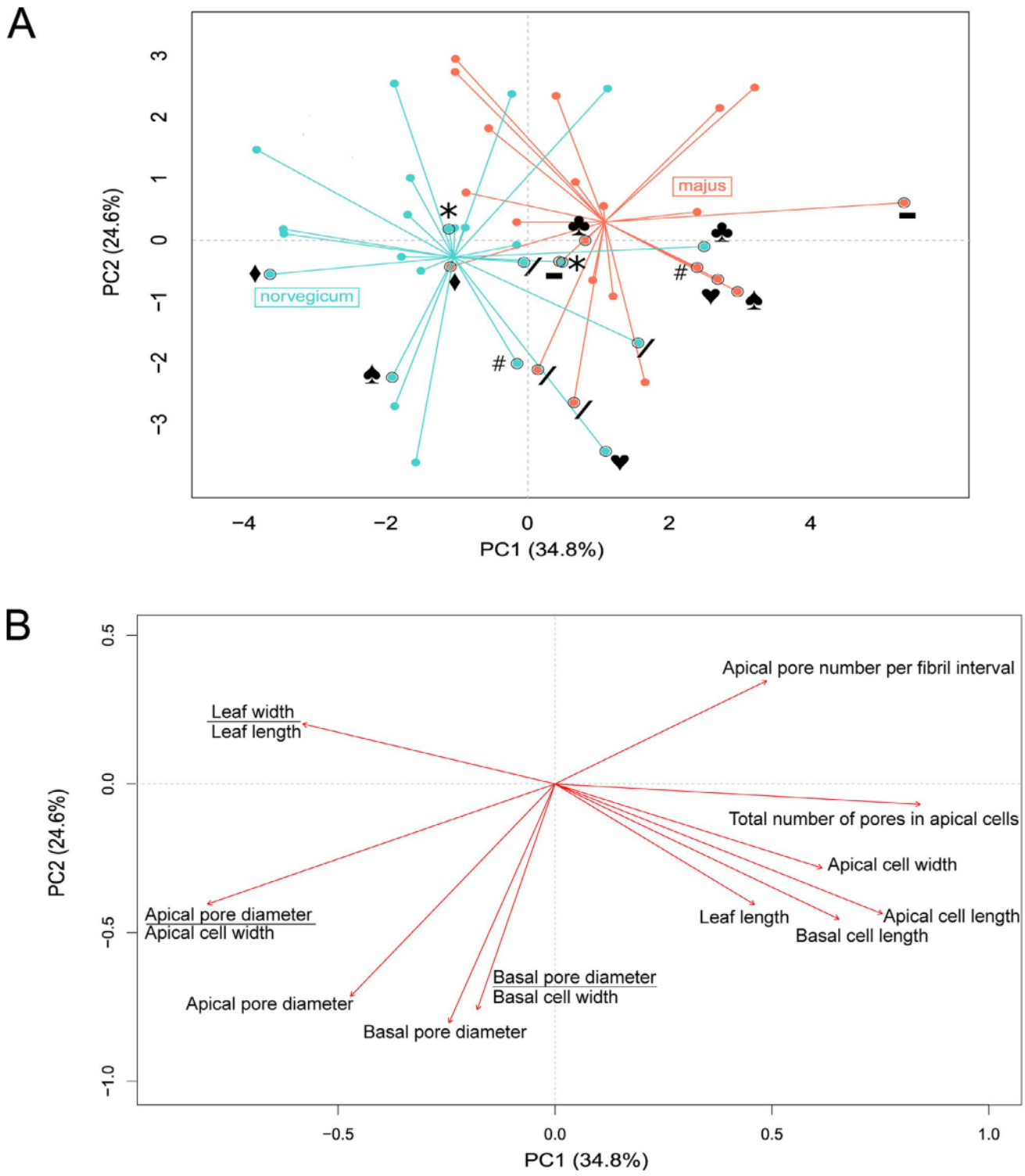

Figure 8. Results of Principal Components Analysis of morphological variation in *Sphagnum majus*. **A.** Plot of samples in relation to the first two principal components. **B.** Relationships of individual morphological traits to the first two principal components.

1999). It is also implicated in the parentage of the Norwegian endemic *S. troendelagicum* Flatberg. along with *S. tenellum* (Brid.) Bory (Stenøien et al. 2011). Our data do not indicate a close relationship between *S. balticum* and *S. majus*, although the former does seem to be repeatedly involved in different allopolyploidization events. Plastid sequences provide clear evidence that *S. cuspidatum* is the maternal parent. Morphologically, *S. majus* shares clear features with both *S. cuspidatum* and *S. annulatum* (Såstad et al. 2000). Molecular data strongly corroborate the hypothesis that these two haploid species are the parents of allopolyploid *S. majus*.

Although we infer a close phylogenetic relationship between *Sphagnum majus* and its two parents, *S. annulatum* and *S. cuspidatum*, it is noteworthy that our STRUCTURE analyses did not show evidence of fixed heterozygosity for genetic groups present in the parental species. From this surprising result we infer that the origin of *S. majus* was sufficiently long ago that it has diverged to the point of comprising its own genetic lineage, and its allopolyploid legacy is now genetically unclear. Its allopolyploid status was nevertheless clearly shown by fixed heterozygosity at isozyme loci (Såstad et al. 2000) and microsatellites (Shaw et al. 2018). While the absence of any allopolyploid signal in our RADseq data is puzzling, we explored various approaches to the analyses of our data and consistently recovered no evidence of fixed heterozygosity.

Multiple lines of molecular evidence indicate that the two subspecies of *Sphagnum majus*, ssp. *majus* and ssp. *norvegicum*, do not represent evolutionary/phylogenetic lineages despite some morphological differentiation. These include the lack of resolution from either phylogenetic or phenetic (STRUCTURE) analyses of genomic scale sequence data. This in turn supports the view that the differences in morphology reflect either phenotypic plasticity or segregating variation within a single gene pool, or some combination of these (i.e., a genotype × environment interaction). Our research therefore does not support the recognition of the subspecies in an evolutionary context, although one could argue that such morphotypes could be useful to distinguish for some ecological applications.

Acknowledgments

This research was supported by NSF grants DEB-1737899 and DEB-1928514. We thank two anonymous reviewers and Associate Editor Rafael Medina for their valuable comments, which helped to improve the first draft of the manuscript.

Literature Cited

Agapow, P. M., O. R. P. Bininda-Emonds, K. A. Crandall, J. L. Gittleman, G. M. Mace, J. C. Marshall & A. Purvis. 2004. The impact of species concept on biodiversity studies. The Quarterly Review of Biology 79: 161–179.

Barrett, C. F & J. V. Freudenstein. 2011. An integrative approach to delimiting species in a rare but widespread mycoheterotrophic orchid. Molecular Ecology 20: 2771–2786.

Burley, J. S & N. M. Pritchard. 1990. Revision of the genus *Ceratodon* (Bryophyta). Harvard Papers in Botany 1(2): 17–76.

Crum, H. A. 1984. Sphagnopsida, Sphagnaceae. North American Flora, ser. 2, part 11: 1–180.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean & R. Durbin. 2011. The variant call format and VCFtools. Bioinformatics 27(15): 2156–2158.

De Queiroz, K. 2007. Species concepts and species delimitation. Systematic Biology 56(6): 879–886.

Doyle, J. J. & J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11–15.

Duffy, A. M., B. Aguero, H. K. Stenøien, K. I. Flatberg, M. S. Ignatov, K. Hassel & A. J. Shaw. 2020. Phylogenetic structure in the *Sphagnum recurvum* complex (Bryophyta) in relation to taxonomy and geography. American Journal of Botany 107(9): 1283–1295.

Earl, D. A. & B. M. vonHoldt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4(2): 359–361.

Eaton, D. A. R. 2014. PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30(13): 1844–1849.

Evanno, G., S. Regnaut & J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology 14(8): 2611–2620.

Flatberg, K. I. 1987. Taxonomy of *Sphagnum majus* (Russ.) C. Jens. Kongelige Norske Videnskabers Selskab Skrifter 2: 1–42.

Fraser, D. J. & L. Bernatchez. 2001. Adaptive evolutionary conservation: towards a unified concept for defining conservation units. Molecular Ecology 10: 2741–2752

Graulich, A. 2021. Distribution, ecology, morphology and reproductive biology of *Sphagnum majus* in the south of its range (Hautes-Fagnes, Belgium). Lindbergia 2021(1): (18 February 2021)

Hanssen, L., S. M. Såstad & K. I. Flatberg. 2000. Population structure and taxonomy of *Sphagnum cuspidatum* and *S. viride*. The Bryologist 103(1): 93–103.

Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh & L. S. Vinh. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Molecular Biology and Evolution 35(2): 518–522.

Huson, D. H. & D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23(2): 254–267.

Koutecký, P. 2015. MorphoTools: A set of R functions for morphometric analysis. Plant Systematics and Evolution 301(4): 1115–21.

Kyrkjeeide, M.O., K. Hassel, K. I. Flatberg, A. J. Shaw, N. Yousefi & H. K. Stenøien. 2016. Spatial genetic structure of the abundant and widespread peatmoss *Sphagnum magellanicum* Brid. PloS one 11(2): e0148447.

Maass, W. S. G. & M. J. Harvey. 1973. Studies on the taxonomy and distribution of *Sphagnum* VII. Chromosome numbers in *Sphagnum*. Nova Hedwigia 24: 193–205.

Mayr, E. 1942. Systematics and the origin of species from the viewpoint of a zoologist. Columbia University Press, New York.

Mayr, E. 1963. Animal species and evolution. Harvard University Press, Cambridge.

McQueen, C. B. & R. E. Andrus. 2007. Sphagnaceae. Pages 45–101. In: Bryophytes: Mosses, part1. Flora of North America. Vol. 27 (ed. Committee FoNAE). Oxford University Press, New York.

Minh, B. Q., M. A. T. Nguyen & A. von Haeseler. 2013. Ultrafast approximation for phylogenetic bootstrap. Molecular Biology and Evolution 30(5): 1188–1195.

Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler & R. Lanfear. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37(5): 1530–1534.

Mishler, B. D. 2009. Three centuries of paradigm changes in biological classification: Is the end in sight? Taxon 58(1): 61–67.

Moritz, C. 1994. Defining "Evolutionarily Significant Units" for conservation. Trends in Ecology and Evolution 9: 373–375.

Natcheva, R. & N. Cronberg. 2007. Maternal transmission of cytoplasmic DNA in interspecific hybrids of peat mosses, *Sphagnum* (Bryophyta). Journal of Evolutionary Biology 20: 1613–1616.

Nieto-Lugilde, M., O. Werner, S. F. McDaniel & R. Ros-Espín. 2018. Environmental variation obscures species diversity in southern European populations of the moss genus *Ceratodon*. Taxon 67: 673–692.

Parchman, T. L., Z. Gompert, J. Mudge, F. D. Schilkey, C. W. Benkman & C. A. Buerkle. 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. Molecular Ecology 21(12): 2991–3005.

Pritchard, J. K., M. Stephens & P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

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

Ramasamy, R. K., S. Ramasamy, B. B. Bindroo & V. G. Naik. 2014. STRUCTURE PLOT: A program for drawing elegant STRUCTURE bar plots in user friendly interface. Springer Plus 3(1): 431.

Såstad, S. M. 1999. Genetic and environmental sources of variation in leaf morphology of *Sphagnum fallax* and *Sphagnum isoviitae* (Bryopsida): Comparison of experiments conducted in the field and laboratory. Canadian Journal of Botany 77(1): 1–10.

Såstad, S. M., K. I. Flatberg & N. Cronberg. 1999. Electrophoretic evidence supporting a theory of allopolyploid origin of the peatmoss *Sphagnum jensenii*. Nordic Journal of Botany 19(3): 355–362.

Såstad, S. M., K. I. Flatberg & L. Hanssen. 2000. Origin, taxonomy and population structure of the allopolyploid peat moss *Sphagnum majus*. Plant Systematics and Evolution 225(1): 73–84.

Shaw, A. J., B. E. Carter, B. Aguero, D. P. da Costa & A. Crowl. 2018. Range change evolution of peat mosses (*Sphagnum*) within and between climate zones. Global Change Biology 2018: 1–13.

Shaw, B., S. Terracciano & A. J. Shaw. 2009. A genetic analysis of two recently described peat moss species, *Sphagnum atlanticum* and *S. bergianum* (Sphagnaceae). Systematic Botany 34(1): 6–12.

Stenøien, H. K., A. J. Shaw, K. Stengrundet & K. I. Flatberg. 2011. The narrow endemic Norwegian peat moss *Sphagnum troendelagicum* originated before the last glacial maximum. Heredity 106(2): 370–382.

Stenøien, H. K., K. Hassel, R. Segreto, R. Gabriel, E. F. Karlin, A. J. Shaw & K. I. Flatberg. 2014. High morphological diversity in remote island populations of the peat moss *Sphagnum palustre*: glacial refugium, adaptive radiation or just plasticity? The Bryologist 117: 95–109.

Temsch, E. M., J. Greilhuber & R. Krisai. 1998. Genome size in *Sphagnum* (peat moss). Botanica Acta 111(4): 325–330.

Weir, B. S. & C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38(6): 1358–1370.

Weston, D. J., M. R. Turetsky, M. G. Johnson, G. Granath, Z. Lindo, L. R. Belyea, S. K. Rice, D. T. Hanson, K. A. M. Engelhardt, J. Schmutz, E. Dorrepaal, E. S. Euskirchen, H. K. Stenøien, P. Szövényi, M. Jackson, B. T. Piatkowski, W. Muchero, R. J. Norby, J. E. Kostka, J. B. Glass, H. Rydin, J. Limpens, E. Tuittila, K. K. Ullrich, A. Carrell, B. W. Benscoter, J. G. Chen, T. A. Oke, M. B. Nilsson, P. Ranjan, D. Jacobson, E. A. Lilleskov, R. S. Clymo & A. J. Shaw. 2018. The Sphagnome Project: enabling ecological and evolutionary insights through a genus-level sequencing project. New Phytologist 217: 16–25.

Wilson, E. O. 1992. The diversity of life. Harvard University Press, Cambridge.

Yousefi, N., K. Hassel, K. I. Flatberg, P. Kemppainen, E. Trucchi, A. J. Shaw, M. O. Kyrkjeeide, P. Szövényi & H. K. Stenøien. 2017. Divergent evolution and niche differentiation within the common peatmoss *Sphagnum magellanicum*. American Journal of Botany 104: 1060–1072.

Zachos, F. E. 2018. (New) Species concepts, species delimitation, and the inherent limitations of taxonomy. Journal of Genetics 97: 811–815.

Zhang, J., P. Kapli, P. Pavlidis & A. Stamatakis. 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29: 2869–2876.

manuscript received October 1, 2021; accepted February 17, 2022.

Supplementary documents online:

Supplementary Table S1. Voucher information for accessions included in this study.

Supplementary Figure S1. Chromosomal distribution map of loci used in this study.

Supplementary Table S2. Statistical results of qualitative and quantitative characters of *Sphagnum majus* specimens included in the biometric study. Number (N) of specimens examined for each group is given. Descriptive statistics (mean ± SD [range])

for quantitative characters are presented. Values with statistically significant difference ($p = 0.05$) are written in bold.

Supplementary Table S3. Pearson correlation coefficients (below the diagonal) and significance tests (above diagonal) between all morphological traits. Significant values ($p \leq 0.05$) are highlighted in bold.

Supplementary Table S4. ANOVAS of morphological characters included in the biometric study. Values with statistically significant difference ($p < 0.05$) are indicated with *.