Received: 30 January 2024 | Accepted: 6 June 2024

DOI: 10.1002/ajb2.16401

American Journal of Botany

RESEARCH ARTICLE

Ecological differentiation and sympatry of cryptic species in the *Sphagnum magellanicum* complex (Bryophyta)

Marta Nieto-Lugilde[1] | **Diego Nieto-Lugilde**[2] | **Bryan Piatkowski**[3] | **Aaron M. Duffy**[1] | **Sean C. Robinson**[4] | **Blanka Aguero**[1] | **Scott Schuette**[5] | **Richard Wilkens**[6] | **Joseph Yavitt**[7] | **A. Jonathan Shaw**[1]

[1]Department of Biology and L. E. Anderson Bryophyte Herbarium, Duke University, Durham, NC 27708, USA

[2]Departamento de Botánica, Ecología y Fisiología Vegetal, Universidad de Córdoba, Córdoba, Spain

[3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA

[4]Department of Biology, SUNY Oneonta, Oneonta, NY 13820, USA

[5]Pennsylvania Natural Heritage Program, Western Pennsylvania Conservancy, Pittsburgh, PA 15222, USA

[6]Biological Sciences Department, Salisbury University, Salisbury, MD 21801, USA

[7]Department of Natural Resources, Cornell University, Ithaca, NY 14853, USA

Correspondence

Marta Nieto-Lugilde, Department of Biology and L. E. Anderson Bryophyte Herbarium, Duke University, Durham, NC 27708 USA.
Email: marta.nietolugilde@duke.edu

Abstract

Premise: *Sphagnum magellanicum* (Sphagnaceae, Bryophyta) has been considered to be a single semi-cosmopolitan species, but recent molecular analyses have shown that it comprises a complex of at least seven reciprocally monophyletic groups, that are difficult or impossible to distinguish morphologically.

Methods: Newly developed barcode markers and RADseq analyses were used to identify species among 808 samples from 119 sites. Molecular approaches were used to assess the geographic ranges of four North American species, the frequency at which they occur sympatrically, and ecological differentiation among them. Microhabitats were classified with regard to hydrology and shade. Hierarchical modelling of species communities was used to assess climate variation among the species. Climate niches were projected back to 22,000 years BP to assess the likelihood that the North American species had sympatric ranges during the late Pleistocene.

Results: The species exhibited parallel morphological variation, making them extremely difficult to distinguish phenotypically. Two to three species frequently co-occurred within peatlands. They had broadly overlapping microhabitat and climate niches. Barcode- versus RADseq-based identifications were in conflict for 6% of the samples and always involved *S. diabolicum* vs. *S. magniae*.

Conclusions: These species co-occur within peatlands at scales that could permit interbreeding, yet they remain largely distinct genetically and phylogenetically. The four cryptic species exhibited distinct geographic and ecological patterns. Conflicting identifications from barcode vs. RADseq analyses for *S. diabolicum* versus *S. magniae* could reflect incomplete speciation or hybridization. They comprise a valuable study system for additional work on climate adaptation.

KEYWORDS

community assembly, DNA barcoding, peatmoss, Sphagnaceae, *Sphagnum diabolicum*, *Sphagnum divinum*, *Sphagnum magniae*, *Sphagnum medium*

When a plant species is difficult or impossible to distinguish from close relatives in the field it can be challenging to identify morphological, phenological, or ecological characteristics that inform relevant research questions about its evolution and systematics. Once a species is recognizable, field biologists can start to assess general patterns of variation among individuals and regional populations. The problem of field identification of species based on macroscopic features is exacerbated in many aquatic plant groups because of phenotypic plasticity associated with varying water levels and other microenvironmental factors (Vanderpoorten and Jacquemart, 2004).

These issues are especially acute in groups that include so-called cryptic species, where distinguishing morphological characteristics are absent or nearly so (Bickford et al., 2007; Renner, 2020). Morphologically cryptic or near-cryptic species occur in a broad range of both animal and plant clades including ants (Schär et al., 2022; Ward and Blaimer, 2022), bumblebees (Williams et al., 2022), lizards (Melville et al., 2014), chipmunks (Herrera et al., 2022), bryophytes (Grundmann et al., 2006; Renner 2020; Martins et al., 2021), and angiosperms (Michalski and Durka, 2015). Nevertheless, many cryptic species have distinct biological properties and represent important units of biodiversity.

Although there are many criteria for delimiting species, including phenetic, phylogenetic, and reproductive traits, reproductive isolation is of obvious evolutionary importance. While species clearly do not interbreed when allopatric, assessing reproductive isolation when they are sympatric is critical for inferring whether they have inherent barriers to interbreeding. When closely related species cannot be distinguished morphologically, the problem of assessing reproductive isolation is made substantially more difficult. In fact, determining whether such species occur in sympatry can be difficult or impossible without some form of species-specific molecular marker.

DNA barcoding (Hollingsworth et al., 2011; Bączkiewicz et al., 2017) is useful for species identification and classification, especially in the absence (or near absence) of diagnostic morphology. Practical applications of barcoding include species conservation, pest control, and species trade management (Francis et al., 2010; Melville et al., 2014; Piper et al., 2019). Barcoding is also a useful tool for basic scientific research (Chase et al., 2005). The complex of *Sphagnum* (Sphagnaceae; peatmoss) species we focused on here was originally targeted for ecological genomic research (Healey et al., 2023) when we understood it to be a single widespread species across a huge range of climate zones. In the process of that work we learned that the group comprises at least seven clades, and four species occur in eastern North America (Yousefi et al., 2017; Shaw et al., 2022). However, because the species are virtually impossible to distinguish morphologically, we still know very little about the extent to which they occur sympatrically within peatland communities. It has also been challenging to assess other biological differences among the cryptic lineages that might support their nomenclatural recognition as species.

The genus *Sphagnum* L. is globally important for carbon accumulation and nutrient cycling and is considered an “ecosystem engineer” that creates favorable conditions for its own growth through acidification and peat formation (van Breemen, 1995; Yu et al., 2010). Growth of various *Sphagnum* species at different heights above the water table generates a heterogeneous microtopography in many northern peatlands (Rydin and Jeglum, 2013). Hummock-forming species that are raised well above the water table tend to be slow growing but decompose very slowly such that accumulating peat from the plants themselves creates the hummocks. Hollow-inhabiting species, in contrast, grow more rapidly but also decompose relatively quickly and thus do not form hummocks and grow at or near the water table. Hummock and hollow species differ in multiple correlated features, suggesting the evolution of “adaptive syndromes” (Johnson et al., 2015; Piatkowski et al., 2021), and hummock-forming versus hollow-inhabiting clades diverged early in *Sphagnum* evolution (Shaw et al., 2016; Healey et al., 2023). Morphological variation associated with position relative to the hummock-hollow gradient significantly complicates the field identification of species of *Sphagnum*.

Sphagnum magellanicum Brid. in the broad sense has for over 100 years been interpreted as having a nearly global geographic range (Crum, 1984; McQueen and Andrus, 2007). Kyrkjeeide et al. (2016) first showed using microsatellites that *S. magellanicum* comprises multiple, partially allopatric genetic groups. Yousefi et al. (2017) showed using RADseq data that two clades occur in Europe and eastern North America and that S. *magellanicum* s.s. is restricted to South America. Hassel et al. (2018) recognized the European-North American clades as *S. medium* Limpr. and *S. divinum* Flatberg & K. Hassel based on global sampling, Shaw et al. (2022) resolved seven species-level clades in the *S. magellanicum* complex. They confirmed that the European species identified by Yousefi et al. (2017) and Hassel et al. (2018) also occur in North America, and new data resolved two additional species, *S. diabolicum* A.J. Shaw, Aguero & Nieto-Lugilde and *S. magniae* A.J. Shaw, Aguero & Nieto-Lugilde, that appear to be endemic to eastern North America (Shaw et al., 2022, 2023b). Despite strong molecular evidence that the four North American species in this complex are reciprocally monophyletic, they are morphologically very similar and difficult or sometimes impossible to tell apart phenotypically (Shaw et al., 2023b).

Hassel et al. (2018) suggested that in Norway, *S. divinum* and *S. medium* differ ecologically with *S. divinum* most common near the margins of peatlands, whereas *S. medium* is most common in the more open parts of peatlands. This niche difference appears to hold in eastern North America also, although *S. divinum* occurs in the open parts of many peatlands as well as in the more-shaded margins and in surrounding wet forests (Shaw et al., 2022). Because the species are hard to distinguish, especially in the field, it is unclear whether *S. diabolicum* differs ecologically from *S. divinum* and *S. medium*. It has also been difficult to determine how often these three species co-occur at the same sites. The fourth North American species, *S. magniae*, is restricted to warm-temperate to subtropical sites in the southeastern United States and is largely allopatric to the other three species, which have cold-temperate to boreal distributions (Shaw et al., 2022).

Multiple species in the complex have been collected at some of the same sites in both Scandinavia and eastern North America (Yousefi et al., 2017; Hassel et al., 2018; Shaw et al., 2022), but their frequency of co-occurrence and/or in what combinations are largely unknown because of their morphologically cryptic nature. The lack of morphological resolution among species has also made it difficult to assess any niche differences among them when they do co-occur. To address these issues, we developed barcode markers using genome resequencing data that distinguish the four northern (North America, Europe) species in the *S. magellanicum* complex and utilized genetically identified collections to better understand their geographic ranges and ecology. We present these barcode markers, which distinguish the species based on

TABLE 1 Primer sequences to amplify the barcode markers used here to study the *Sphagnum magellanicum* complex in North America.

Name	Forward primer (5′–3′)	Reverse primer (5′–3′)
LG09_S1	AGGGGTGGTCATATGTGGGA	AGCAAGCACCCGTTATGGAA
LG19_S5	ATCTGGAGTCTGCCACAAGC	TGGCGATGAATTCTCTGCTCT
LG01_S6	TACTCTTGGGTTGGGGCAAG	AGGGAAGGATCAACTTAGAAGGC

amplicon sizes from three (nuclear) genomic regions. Nucleotide sequencing is not necessary.

We addressed four main questions in this study. (1) What are the (genetically) confirmed ranges of the four species in eastern North America? (2) Do species differ in climate niches? (3) Do species co-occur in individual peatlands, and if so, at what relative frequencies? (4) Do the species exhibit niche differences relative to two local-scale variables, hummock vs. hollow microtopography and shade? Our overarching conceptual goal was to assess geographic and ecological differentiation among species in this complex despite the absence or near absence of morphological traits to distinguish them and to evaluate the potential for interspecific mating based on sympatric occurrences within peatlands.

MATERIALS AND METHODS

Plant sampling

A total of 808 collections from which 871 individual gametophyte plants were sampled from the eastern and western USA ($N = 772$ collections) and Canada ($N = 36$) were included in our analyses (Appendix S1). Species determinations were based on phylogenetic analyses using RADseq data ($N = 411$), barcodes ($N = 231$), or both ($N = 229$) (Appendix S2). Some data were from Shaw et al. (2022 [N = 166] and 2023a [N = 404]). Plants were collected from 119 sites (peatlands; mean = 6.8 ± 9.7 collections/site), 40 of them with five samples or more. They represent 24 states and eight Canadian provinces and cover a latitudinal range from 27° to 57°N. Voucher specimens are deposited in the L.E. Anderson Bryophyte Herbarium at Duke University (DUKE). DNA was extracted from only part of a single gametophyte; any remaining parts of that plant were placed in a smaller envelope and returned to the specimen herbarium packet. Morphological variation among field-growing plants was documented photographically.

Barcode design

We used genome resequencing data from 46 samples generated by Shaw and collaborators (Shaw et al., 2022) to design species-specific barcodes. The BWA-MEM algorithm v0.7.12 (Li and Durbin, 2009) was used to align filtered reads to the *S. divinum* v1.1. genome (https://phytozome-next.jgi.doe.gov/info/Smagellanicum_v1_1), and joint genotyping was performed using GATK v4.2.2.0 (Van der Auwera and O'Connor, 2020). Variants were filtered to keep only indels that met the following criteria: $QD \geq 2.0$, $MQ \geq 40.0$, $FS \leq 60.0$, $SOR \leq 3.0$, $MQRankSum \geq -12.5$, and $ReadPosRankSum \geq -8.0$. We then identified biallelic indels that were fixed for either the reference or alternate allele within all samples of each species. PCR primers were developed using NCBI's Primer-BLAST tool (Ye et al., 2012) after identifying candidate loci that distinguish species.

Primers designed for barcoding the *S. magellanicum* complex were based on data from samples collected in the United States. To confirm the applicability of our primers for non-North American samples, plants from North America and Europe were used for testing, with their identification confirmed based on prior phylogenetic analyses with RADSeq sequences.

We here provide PCR primers (Table 1) that can be used to distinguish the four North American species. Use of these loci still requires laboratory work, but identifications just involve DNA extraction, barcode amplification, and electrophoresis of the amplified DNA in an agarose gel. Identification to species is a multistep process. Primer LG09-S1 is used first to distinguish *S. divinum* and *S. medium* from *S. diabolicum* and *S. magniae* (Figure 1A), and depending on the outcome, either primer LG19-S5 is used to distinguish *S. medium* from *S. divinum* or primer LG01-S6 is used to distinguish *S. diabolicum* from *S. magniae* (Figure 1B). These steps are embodied in the following key.

1. Fragment size using LG09-S1 primers = 571 bp.. 2
    2. Fragment size using LG19-S5 primers = 258 bp .... *Sphagnum divinum*
    2. Fragment size using LG19-S5 primers = 433 bp .... *Sphagnum medium*
1. Fragment size using LG09-S1 primers = 457 bp 3
    3. Fragment size using LG01-S6 primers = 137 bp.......................... *Sphagnum magniae*
    3. Fragment size using LG01-S6 primers = 231 bp .... *Sphagnum diabolicum*

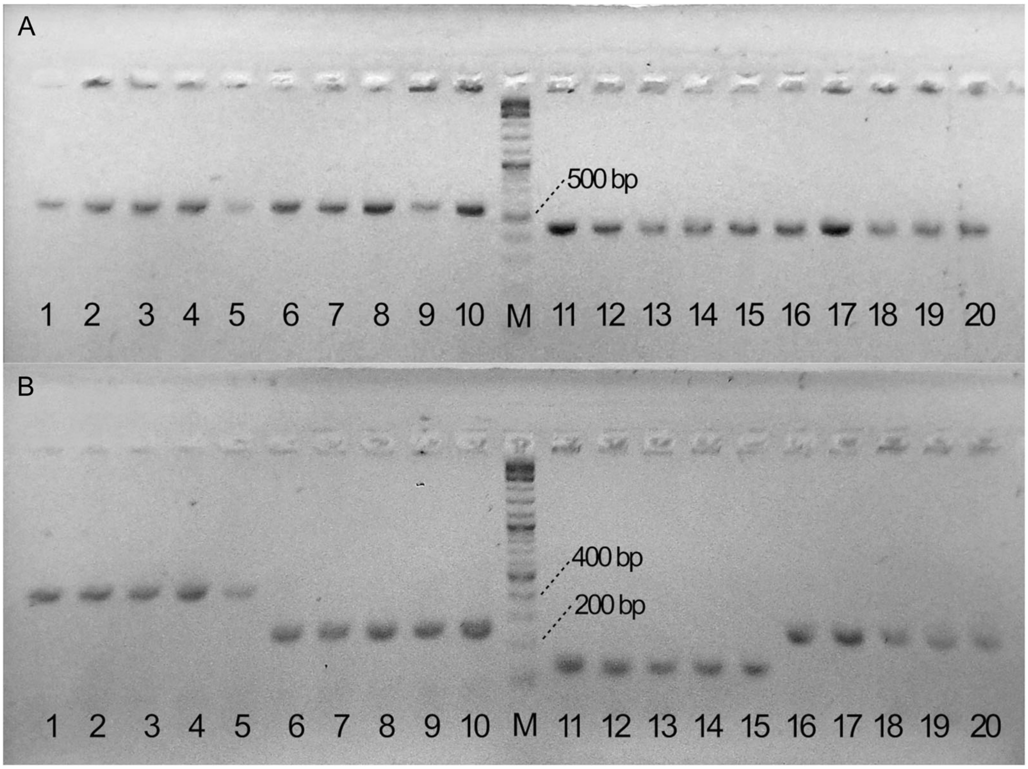

FIGURE 1 Agarose gel of barcode PCR products from 20 samples used to test primers for distinguishing species in the *Sphagnum magellanicum* complex: *S. medium* (lanes 1–5), *S. divinum* (6–10), *S. magniae* (11–15), *S. diabolicum* (16–20). M = 10 kbp ladder marker. (A) PCR products from primer set LG09_S1 with expected amplicon of 571 bp for *S. medium* and *S. divinum* (1–10), and of 457 bp for *S. magniae* and *S. diabolicum* (11–20). (B) PCR products from primer sets LG19_S5 (1–10) and LG01_S6 (11–20) with expected size of 433 bp for *S. medium*, 258 bp for *S. divinum*, 137 bp for *S. magniae*, and 231 bp *S. diabolicum*.

DNA extractions, PCR conditions, and species identifications

Genomic DNA was extracted using a modified CTAB procedure (Nieto-Lugilde et al., 2022). PCR barcode primers designed to amplify species-specific loci that differ in size were used to identify samples. We amplified the three loci in 10-µL reaction mixtures using Choice Taq DNA Polymerase (Denville Scientific, Metuchen, NJ, USA) and the manufacturer's instructions. The cycling conditions were 95°C for 2 min; 35 cycles of 95°C for 30 s, annealing at 57°C for 30 s, and 72°C for 2 min; and terminating with 72°C for 7 min. The banding patterns were visualized after electrophoresis in 3% agarose with SYBR safe (Invitrogen, Eugene, OR, USA) fluorescent DNA stain and scored by hand.

ddRADseq library preparation, sequencing, pipeline, and phylogenetic analysis

Genomic libraries were made using the double-digestion restriction site-associated DNA sequencing (using EcoRI and MseI enzymes) protocol of Parchman et al. (2012), with modifications described by Duffy et al. (2020). The libraries were cleaned and size-selected for fragments of ~350 bp using AMPur XPbeads (Beckman Coulter, Indianapolis, IN, USA), checked for quality on a BioAnalyzer (Agilent, Santa Clara, CA, USA), and sequenced on a single lane of an Illumina NovaSeq. 6000 with 150-bp single-end reads at the Genome Sequencing Shared resource at the Duke Center for Genomic and Computational Biology (https://medschool.duke.edu/research/research-support/service-centers/core-research-facilities/sequencing-and-genomic). For 69 samples that had previously been resequenced (DNA codes with four letters in Appendix S1), "RADseq-like" sequences were generated from the genome assemblies using the in silico digestion method described by Shaw et al. (2022).

Single nucleotide polymorphisms (SNPs) were called using ipyrad v.0.9.50 (Eaton and Overcast, 2020) with default parameters except as noted here. Reads were processed as datatype ddrad to match the library preparation method. A maximum of one mismatched base was allowed in the barcode during demultiplexing. Illumina adapter sequences and low-quality bases were trimmed from the reads, which were trimmed to a maximum of 92 bases after removing the barcode, and trimmed reads less than 35 bases long or with more than five low-quality bases were discarded. Loci were assembled using the reference method against the *S. divinum* genome (Healey et al., 2023). Only loci from chromosomes 1–19 (no sex chromosomes or unidentified linkage groups) present in at least 80% of the samples were kept for final analyses.

To avoid miscalled genotypes resulting from ipyrad's majority-rule haploid genotype calling, samples were initially processed as diploids. Heterozygous genotypes were then converted to missing data and homozygous genotypes were treated as haplotypes, as described by Shaw et al. (2023a).

Phylogenetic relationships among samples were inferred from the best of 10 independent runs of IQ-TREE2v2.1.2 (Minh et al., 2020) using concatenated SNPs. The maximum likelihood tree was estimated using random starting trees and the ultrafast bootstrap approximation (UFBoot: Minh et al., 2013; Hoang et al., 2018). The best nucleotide substitution model (GTR+F+ASC+R7) was estimated using the ModelFinder algorithm with corrections for ascertainment bias (Lewis, 2001; Kalyaanamoorthy et al., 2017) as implemented in IQ-TREE, 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. The phylogenetic tree from RADseq data was used here only to infer species identifications (Appendix S2). Additional analyses of population structure, along with estimates of interspecific introgression and gametophyte sex ratios will be reported in a subsequent publication.

Assessing microhabitats for each species

The four species occur at sites with or without open water (Figure 2A, B) and with variable amounts of shrub cover (Figure 2C). They occurred in open peatlands (Figure 2A–C), along the semi-open forested edges of peatlands (Figure 2D) or in surrounding forests (Figure 2E). The warm temperate-subtropical species, *S. magniae*, most commonly occurred in ecological communities with no counterparts in the boreal zone, sometimes in coastal pine forests dominated by palmetto (*Sabal* sp., Figure 2F), which reaches its northern range limit in North Carolina. For a preliminary assessment of whether species in the complex differ in microniches within peatlands, samples were subjectively categorized into three levels along the hummock–hollow gradient: hollows/lawns, low hummocks, and high hummocks. Hollows/lawns were close to or at the water table; hummocks were subjectively classified as low or high. Samples were also classified with regard to microhabitat shadiness: shade, partial shade, full sun. Plants growing in forests surrounding the peatlands were coded as shade, those in microsites along the margins of the peatlands with scattered trees and/or shrubs as partial shade, and those in the open parts of peatlands as full sun. In some peatlands, dense shrubs occurred throughout the open portions of the sites, and these were coded as partial shade. Species identifications were unknown when ecological data were scored; each plant was vouchered and later identified using molecular markers. Not all samples were included in the ecological analyses. The microniche data sets included 346 samples from 29 peatlands (mean = 11.9 ± 7.9 collections/site). To assess significance of variation in these two qualitative microniche variables—hydrology and shadiness/vegetation cover—we made pairwise species comparisons using the fisher.test function in R v.4.2.3 (R Core Team, 2023). Fisher's exact test does not necessitate assuming a minimum of five expected counts in the contingency table (Agresti, 2007), making it best suited for our data set.

Modelling species distributions and co-occurrence

To study the effect of climate and microhabitat variables on the distribution and co-occurrence of the four *Sphagnum* species, we used hierarchical modelling of species communities (HMSC; Ovaskainen et al., 2017), a hierarchical Bayesian approach that fits multiple species jointly while taking phylogeny into account. This method can also be used to estimate the residual of the co-occurrence between species after accounting for the effects of the environmental variables. To do so, we used the species occurrences (presence–absence) from the 38 sites where five or more samples were identified using genetic markers to build a community matrix where each sample was kept as a row in the community matrix to capture microhabitat variability. Samples with conflicting identifications between molecular methods (barcoding vs. RADseq) were excluded. Current climate information for the 38 sites were extracted from the 19 bioclimatic variables in the Chelsa climate database (v1.2; https://chelsa-climate.org/) and summarized by the first four axes of a PCA. The number of axes retained was proposed by a broken stick model applied to the resulting PCA (King and Jackson, 1999). The community matrix was the dependent variable in our HMSC model and the four climate variables plus the two microhabitat parameters were our independent variables. To account for the nested nature of the sampling scheme, two random variables were defined: one at the site level and a second one at the sample level. The structure of the model was completed with the phylogenetic tree, which was used to estimate phylogenetic signal on the species responses to environmental variables. Model calibration was carried out with two Markov chain Monte Carlo runs, with a warming phase of 10,000 samples and 100,000 samples of posteriors, which were then thinned by a factor of 100 to avoid autocorrelation in the resulting posteriors. We used the effective sample size (ESS) and Gelman diagnostic over the beta parameters estimated by the Bayesian model (the effect of the environmental variables) to test convergence.

We used species distribution models (SDMs) to assess climatic niches and their effects on present and past potential distributions of the species. To do so, we used an ensemble of five different algorithms (Velazco et al., 2022): generalized linear models, Gaussian process models, generalized boosted regression models, neural networks and maxent. To avoid pseudoreplication, we applied environmental filtering to remove samples with similar climatic conditions (Varela et al., 2014). Then, a calibration area of 500 km was defined around the observations for each species, which allowed

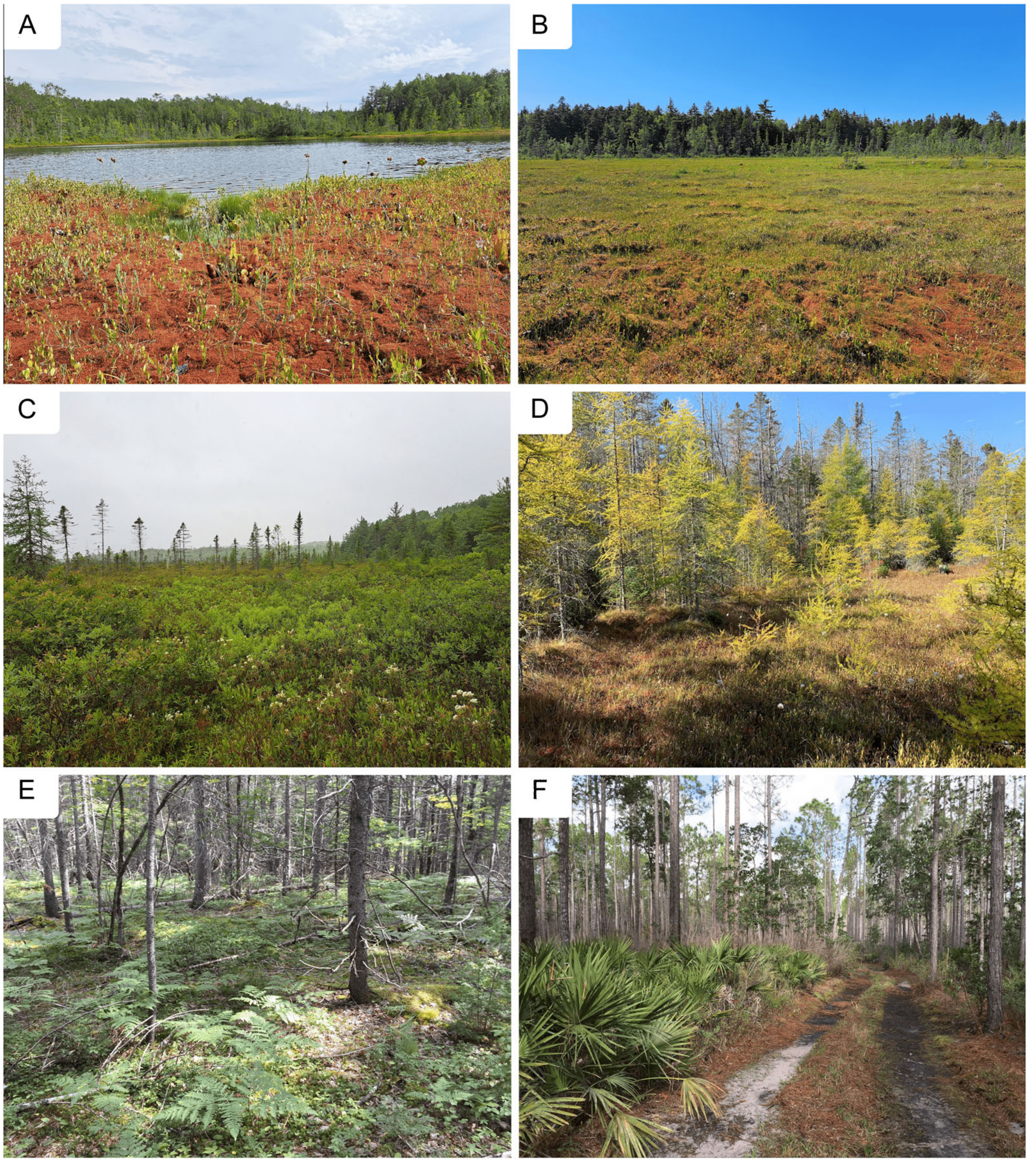

FIGURE 2 Habitats of species in the *Sphagnum magellanicum* complex. (A, B) Open peatlands with or without open water, scored as "Full sun" sites. (C, D) Shrubby peatlands (C) and peatland margins (D) scored as "Partial shade". (E, F) Forest habitats in Maine (E) and Florida (F) scored as "Shade".

completion of the occurrence data set (by selecting pseudoabsences and background data within the calibration area) and avoided biasing the models because of nonsampled regions. For each species, the same number of pseudoabsences as presences were selected, while the background was made up of 10 times the number of presences. To evaluate the models, we calculated the area under the receiver operating characteristic curve (AUC) and the true skill statistics (TSS) over four partitions of spatial blocks. By doing so, we avoided spatial autocorrelation between training and testing data sets.

The ensemble model was projected to the climate conditions during the Last Glacial Maximum (22 kyr BP) obtained from the Chelsa-Climate v1.2 repository (https://chelsa-climate.org/). Past climate projections are from the CCSM4 climate model and were used to assess likely range overlap and the possibility of past interspecific hybridization.

RESULTS

Species identification

Of the 871 samples, 293 were identified as *S. diabolicum*, 373 as *S. divinum*, 83 as *S. magniae*, and 109 as *S. medium*. Of the 229 barcode-identified samples that were also subjected to RADseq (2140 loci with 14388 SNPs present in 80% of samples; 16.8% total missing data), 13 samples (6%) had barcodes and RADseq data that yielded conflicting inferences about species identification. Conflicting identifications always involved *S. diabolicum* and *S. magniae*; barcodes and RADseq data always yielded the same results for the other five species pairs. Samples with conflicting identifications came from seven localities (Table 2), two populations from North Carolina (Alligator River and Green Swamp), two from Maryland

TABLE 2 Samples of *Sphagnum diabolicum* and *S. magniae* with conflicting species identifications from barcodes and RADseq data. See full sample information in Appendix S1.

Locality	DNA sample	Barcode ID	RADseq ID
NC-GreenSwamp	SB5603	*S. diabolicum*	*S. magniae*
NC-AlligatorRiver	SB5518	*S. diabolicum*	*S. magniae*
MD-PokomokeSwamp3	SB8155	*S. diabolicum*	*S. magniae*
MD-PokomokeSwamp3	SB7892	*S. diabolicum*	*S. magniae*
MD-PokomokeSwamp3	SB7893	*S. diabolicum*	*S. magniae*
MD-PokomokeSwamp3	SB7718	*S. diabolicum*	*S. magniae*
MD-PokomokeSwamp3	SB7897	*S. diabolicum*	*S. magniae*
MD-PokomokeSwamp1	SB8153	*S. magniae*	*S. diabolicum*
NY-KibbieLake	SB6720	*S. magniae*	*S. diabolicum*
ME-GreatWassIsland	SB5430	*S. magniae*	*S. diabolicum*
ME-EastHancock	SB7927	*S. magniae*	*S. diabolicum*
ME-EastHancock	SB7808	*S. magniae*	*S. diabolicum*
ME-EastHancock	SB7919	*S. magniae*	*S. diabolicum*

(Pokomoke Swamp 1 and 3), one from New York (Kibbie Lake), and two from Maine (Great Wass Island and East Hancock). At the Green Swamp, all 16 samples were unambiguously *S. magniae* based on RADseq data, the expected species at this coastal plain site, but one sample barcoded as *S. diabolicum*. Similarly, four of five samples from Alligator River (NC) were identified as *S. magniae* by both barcodes and RADseq data, but one sample that RADseq data resolved within the *S. magniae* clade was identified as *S. diabolicum* by barcodes. Five samples from the Maryland site, Pokomoke Swamp 3, also barcoded as *S. diabolicum*, whereas RADseq data placed them within *S. magniae*. At this site, five of eight plants sampled had conflicting identifications from barcodes and RADseq data. The remaining three samples were identified by barcodes only, one as *S. diabolicum* and two as *S. magniae*. A plant from a nearby site in coastal Maryland (Pokomoke Swamp 1) barcoded as *S. magniae* but was phylogenetically nested within *S. diabolicum* (Table 2). For the Kibbie Lake (NY) samples, both methods identified two samples as *S. diabolicum*, one as *S. magniae* based on barcodes only, and one sample was phylogenetically (RADseq data) *S. diabolicum* but barcoded as *S. magniae*. The four conflicting samples from Maine (two sites; Table 2) were also phylogenetically *S. diabolicum* but barcoded as *S. magniae*.

Morphological variation

Previous research indicated that the four species have broadly overlapping microscopic morphological traits, though *S. medium* can sometimes be distinguished by branch leaf hyaline cell pore structure (Hassel et al., 2018; Shaw et al., 2023b). In the field, each species varied in color and colony density. Plant pigmentation in each species ranged from green to variously red (Figures 3–6). With the exception of *S. magniae*, which usually occurred in warm temperate communities that differ in ecology and physiognomy from typical "boreal peatlands", each of the northern species occurred across a broad hummock–hollow gradient within peatlands (see below). Plants growing close to the water table formed loose hummocks, whereas those forming high hummocks were denser. The species exhibited parallel clinal variations relative to this microtopological gradient and were virtually impossible to distinguish in the field. Dense, high-hummock phenotypes (e.g., *S. diabolicum*, Figure 3H; *S. divinum*, Figure 4H, I; *S. medium*, Figure 6A, D) were especially similar.

Geographic ranges and sympatric populations

Three of the species, *S. diabolicum*, *S. divinum*, and *S. medium*, had northern ranges and were most abundant from Pennsylvania northward to Minnesota, Michigan, and Quebec, whereas *S. magniae* extended from New Jersey southward to Florida and along the Gulf Coast (Figure 7). One sample from Michigan was identified as *S. magniae* by RADseq data; replicate DNA extractions from these plants confirmed it as a disjunct occurrence and not a laboratory error (Shaw et al., 2022). Both *S. divinum* and *S. medium* also occurred in western Europe, but *S. diabolicum* and *S. magniae* appeared to be endemic to eastern North America (Shaw et al., 2022; and unpublished data). *Sphagnum divinum* is the only species in the complex that occurred in western and eastern North America. Both *S. diabolicum* and *S. magniae* were recorded from Maryland and New York, and based on the barcode markers, occurred sympatrically at one site in each state (Pokomoke Swamp 3 in MD and the Kibbie Lake in NY; Appendix S1). No sympatric populations of *S. diabolicum* and *S. magniae* were confirmed with RADseq data.

Niche modeling showed that in terms of climate, the four species have potential geographic ranges much broader than their currently documented distributions indicate (Figure 8). The potential ranges of all four species, including the southeastern *S. magniae*, extend northward to maritime Canada including Newfoundland and eastern Quebec/Labrador. The only species whose potential range includes peninsular Florida is *S. magniae*, which is recorded south to the Lake Okeechobee area (Figure 7). The climate of northwestern North America also appears to be appropriate for all four species, although only *S. divinum* occurs there.

Many peatlands harbored multiple sympatric species of the complex (Appendix S3), and there was no clear relationship to latitude except that *S. magniae* predominated at the southernmost sites (Figure 9). At the southernmost sites (e.g., Georgia, Florida, Gulf Coast), populations were monospecific *S. magniae*. Among the northern species, *S. diabolicum* was recorded farther south (to North Carolina)

FIGURE 3 Field morphological variation in *Sphagnum diabolicum*. (A) *B. Aguero 19832*. (B) *Shaw 2022-34*. (C) *Shaw 2022-64*. (D) *Shaw 2022-224*. (E) *Shaw 2022-177*. (F) *Shaw 2022-243*. (G) *B. Aguero 20126*. (H) *B. Aguero 19843*. (I) *B. Aguero 20182*.

than either *S. divinum* or *S. medium*. Communities containing *S. divinum* and *S. medium* were predominant at the northernmost sites (Figure 9).

All three of the northern species, *S. diabolicum*, *S. divinum*, and *S. medium*, were recorded together at five sites (Figure 9; Appendix S3). Eighteen other sites harbored two species. Of those 18, six yielded *S. diabolicum* and *S. divinum*, six had sympatric *S. divinum* and *S. medium*, five had sympatric *S. diabolicum* and *S. medium*, and one had *S. diabolicum* and *S. magniae*. Some plants at the Maryland site, Pokomoke Swamp 3 mentioned above, were identified by barcode markers as *S. diabolicum* but were placed in the *S. magniae* clade based on phylogenetic analyses of RADseq data, so it is unclear whether this occurrence should be considered sympatric. That is the only site with possible sympatric plants of these two species, and *S. magniae* does not otherwise co-occur with any other species in the complex. Only a single species in the complex was recorded from 90 sites, but 46 of those were represented by one sample and another 10 by two samples (Appendix S1). Sixteen sites with five or more samples yielded just one species. *Sphagnum medium* was never encountered without another species in the complex at any site.

When species co-occurred at a site their relative abundances varied from site to site (Appendix S4). It is impossible to accurately estimate the relative frequencies among species at sites where sample sizes were limited. Among the most intensively sampled sites with three sympatric species (>20 genetically identified samples), two yielded 80–90% *S. diabolicum* plants with just a few collections of *S. divinum* and *S. medium* (Orono

FIGURE 4 Field morphological variation in *Sphagnum divinum*. (A) *Shaw 2022-41*. (B) *B. Aguero 20216*. (C) *Shaw 2022-105*. (D) *B. Aguero 20248*. (E) *M. Nieto-Lugilde 2022-319*. (F) *B. Aguero* 20214. (G) *B. Aguero 20181*. (H) *B. Aguero 20160*. (I) *Shaw 2022-98*.

Bog, Indian River Fen; Appendix S4). At a third site (Whalesback Ridge), plant samples were 48% *S. medium*, 43% *S. diabolicum*, and 9% *S. divinum*. Plants that barcoded as S. *magniae*, which is mostly a species of southeastern and Gulf coastal areas, was only recorded at more northern sites where its close relative *S. diabolicum* also occurred.

At sites with two species and sample sizes of at least 20 plants, *S. diabolicum* and *S. medium* grew together at three sites, *S. diabolicum* and *S. divinum* at two, and *S. divinum* and *S. medium* at one. Relative abundances varied across sites and did not appear to be related to which pair of species were sympatric. For those sites with sample sizes >20, *S. diabolicum* co-occurred with every other species and was often but not always the most abundant.

Microhabitat niches

The positions of sympatric species within two peatlands in Maine (Appendix S5) suggest that the species co-occur at scales that would allow interbreeding. *Sphagnum divinum* tended to grow near the peatland margins (Appendix S5a) but also occurred in more open areas up to the edges of open water when open water was present (Appendix S5b). *Sphagnum diabolicum* and *S. medium* tended to be most common in the open parts of the peatlands but also grew along the margins of the peatlands among shrubs and trees as well.

The distributions of species relative to microhabitat parameters are listed in Appendices S6 and S7 and summarized in Figure 10. The species exhibited significant differences ($P < 0.05$) in pairwise comparisons for both hydrology

FIGURE 5 Field morphological variation in *Sphagnum magniae*. (A) *Shaw 2018-140*. (B) *B. Aguero 19608*. (C) *Shaw 2017-306*. (D) *B. Aguero 19956*. (E) *Shaw 2018-66*. (F) *Shaw 2018-96*. (G) *B. Aguero 20005*. (H) *B. Aguero 19977*. (I) *B. Aguero 20041*.

and shadiness/vegetation cover, except for the comparison between *S. diabolicum* and *S. divinum* ($P = 0.6624$ and $P = 0.06267$, respectively). Most *S. magniae* collections were scored as forming low hummocks, reflecting its occurrence in forests of the southern United States. All three northern species formed medium or high hummocks, but also sometimes occurred in hollows (Figure 10A). *Sphagnum medium* seemed to commonly occur in hollows or form high hummocks, with low hummocks less common. The other two northern species, *S. diabolicum* and *S. divinum*, more commonly formed low hummocks rather than high hummocks or occupying hollows (Figure 10A).

Relative to light levels, *S. medium* was more common in open sunny microsites than in shaded places and was rare in the deep shade of forests (Figure 10B). Both *S. diabolicum* and *S. divinum* occurred in full sun to shade, perhaps most commonly in partial shade. *Sphagnum magniae* rarely occurred in full sun and typically grew in partial shade within somewhat open pine forests, or in more densely shaded forests.

Integrated effects of climate and microniche variation

The first three PCs explained 80% of the variance in all climatic variables (Appendix S8). Initially, each species was assumed to have a unimodal (quadratic) response to the three PCs. However, no species had significant coefficients for the quadratic terms of the climate variables. Hence, we rebuilt models considering only monotonic (binomial-linear) responses.

FIGURE 6 Field morphological variation in *Sphagnum medium*. (A) *B. Aguero 20082*. (B) *Shaw 2022-87*. (C) *Shaw 2022-88*. (D) *B. Aguero 20265*. (E) *B. Aguero 20120*. (F) *B. Aguero* 20174. (G) *B. Aguero 20155*. (H) *B. Aguero 20244*. (I) *B. Aguero 20239*.

The ESS and Gelman diagnostic in the HMSC model indicated a good convergence of beta parameters in the model (Appendix S9). A relatively low value for root mean squared error (RMSE), high values for the AUC, and adjusted R^2 for binomial data (TjurR2) in the model indicated that the model adjusted well to the training and testing (using cross validation with four folds) data and is hence able to explain and predict species occurrences accurately (Appendix S10).

Species differed in terms of the partitioning of variance explained by ecological variables (Figure 11). Climate variables explained a large proportion (>75%) of occurrences in *S. magniae*, which showed small but significant negative responses to several climatic gradients (PCs 1–2, 4; Appendix S11), whereas for *S. diabolicum*, *S. divinum*, and *S. medium* climate variables were far less influential (<10, 20, and 20%, respectively). Microhabitat variables explained, in general, a small amount of occurrence variation (<20%) for all species. In contrast, random variables (sample and locality) combined explained the greatest proportion of variation for *S. diabolicum*, *S. divinum*, and *S. medium* (>50%). Co-occurrence residuals estimated by the HMSC model at the locality level indicated that once the effects of climate and habitat were removed, both *S. diabolicum and S. medium* co-occurred less frequently than expected by random with *S. divinum* (Figure 12). The negative association of *S. divinum* with *S. diabolicum* is stronger than with *S. medium*, and both are statistically significant.

Projecting geographic ranges

Our ensemble distribution models also had good predictive abilities (Appendix S10). When the geographic ranges were projected back 22,000 years to the Last Glacial Maximum, all four species could have had overlapping ranges in the

FIGURE 7 North American distributions of species in the *Sphagnum magellanicum* complex based on genetically confirmed collections.

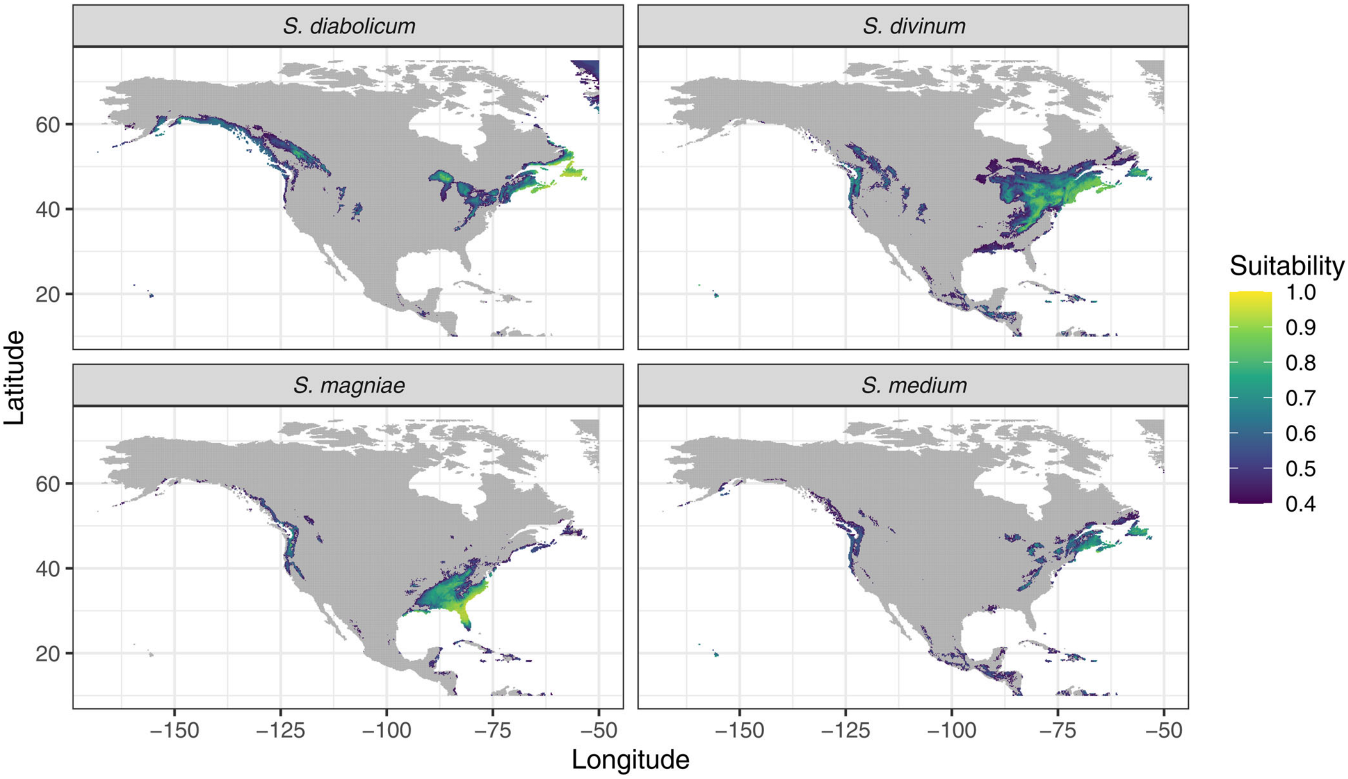

FIGURE 8 Current potential distribution maps of species in the *Sphagnum magellanicum* complex based on climate niche modeling (see text for additional information).

south-central coastal areas of the eastern United States (Figure 13). Our projection suggests that all four species could have occurred in the Pacific Northwest of North America with the greatest suitable area for *S. diabolicum* and the least for *S. divinum*, yet *S. divinum* is the only species currently in that region.

DISCUSSION

Our results clearly showed that species in the *S. magellanicum* complex frequently grow sympatrically within peatlands. They exhibited limited (but statistically significant) ecological differentiation relative to hummock

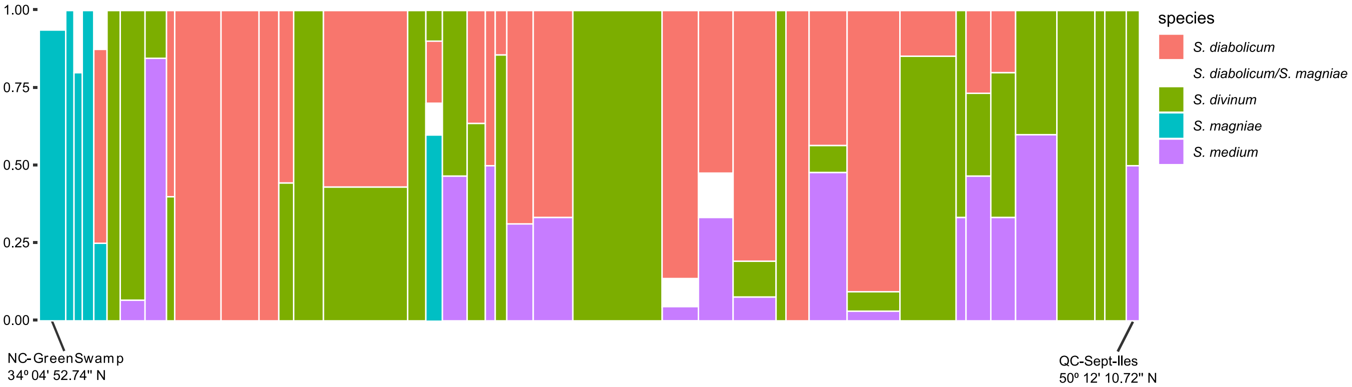

FIGURE 9 Occurrences of species in the *Sphagnum magellanicum* complex within sites, sorted by latitude (south–north). Each column is a peatland site; the widths of bars are proportional to number of samples from each site. *S. diabolicum/S. magniae* includes samples for which barcodes and RADseq yielded conflicting identifications.

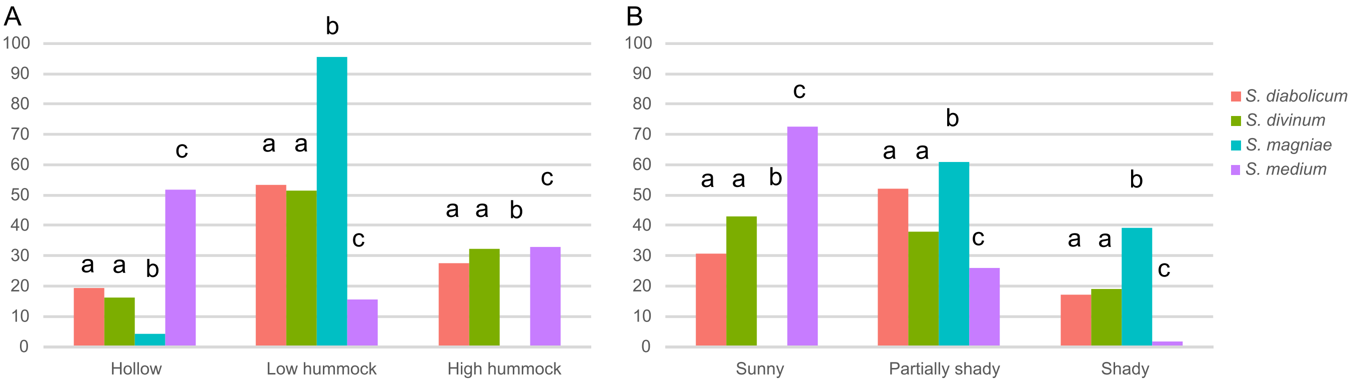

FIGURE 10 Microhabitat distributions of species in the *Sphagnum magellanicum* complex. (A) Hummock–hollow gradient in relation to the water table. (B) Shadiness gradient in relation to the vegetation cover. Lowercase letters denote groups established based on results of Fisher's test, indicating pairwise significantly different species.

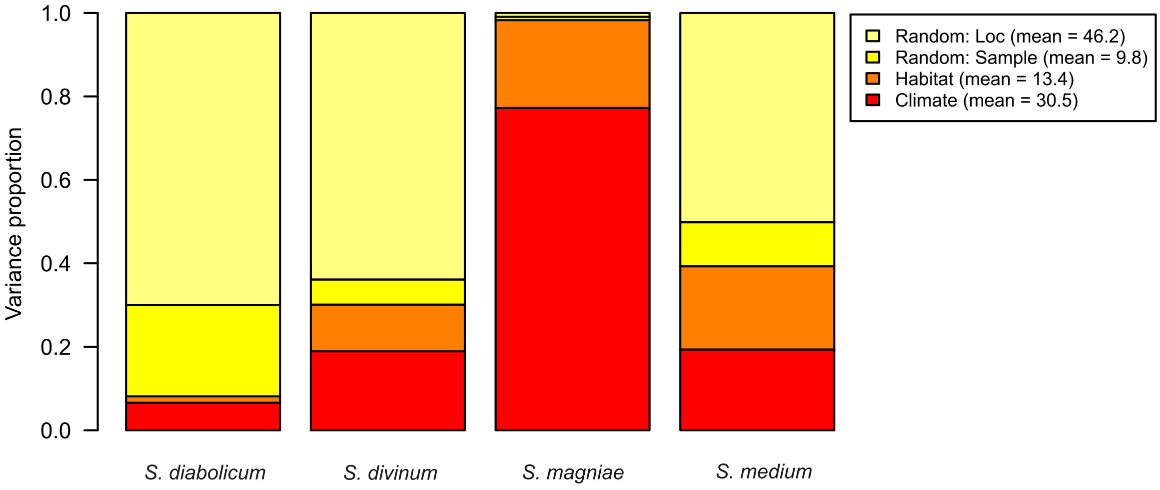

FIGURE 11 Overall variance partitioning of the hierarchical model of species communities in the *Sphagnum magellanicum* complex.

formation and shadiness of the microhabitat, but at many sites, they grew close enough to permit hybridization. Additional studies of mating patterns in natural *Sphagnum* populations are needed, but sexual reproduction appears to be common (Sundberg and Rydin, 2000; Shaw et al., 2023a). Like other bryophytes, *Sphagnum* requires free water for sperm to reach eggs, and water is often abundant in these peatland habitats. The degree of

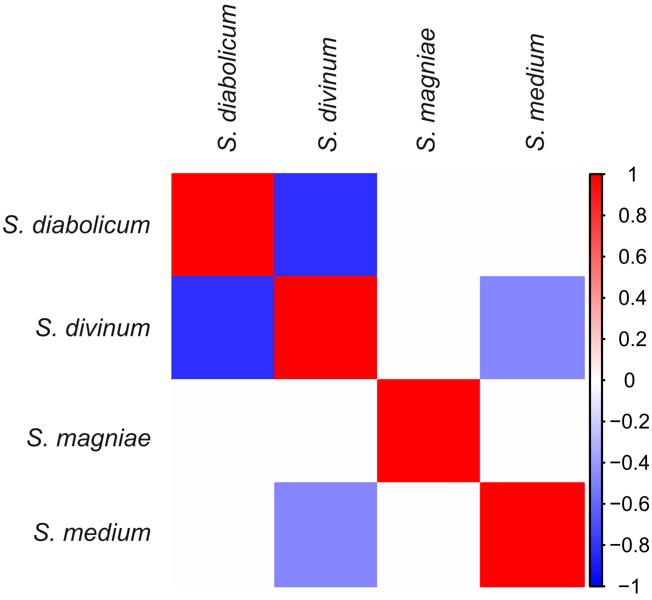

FIGURE 12 Species residual association structure (Ω parameters) in the *Sphagnum magellanicum* complex. Positive associations with high (at least 95% posterior probability) support are shown in red and negative associations in blue.

outcrossing in *Sphagnum* is impacted by position along the hummock–hollow gradient (Johnson and Shaw, 2015), but the physical scale of gamete dispersal is presently unclear (Johnson and Shaw, 2016). We nevertheless detected multiple species separated by distances that would very likely allow interbreeding.

Phylogenetic relationships among the four North American species in the *S. magellanicum* complex have been reconstructed from whole-genome sequencing and RADseq data (Shaw et al., 2022). Among North American species in the complex, *S. medium* is sister to *S. divinum*, *S. diabolicum*, and *S. magniae*, and then *S. divinum* is sister to *S. diabolicum* and *S. magniae*. The last two taxa are especially closely related and are in fact not reciprocally monophyletic across their genomes. Some genomic regions suggest that *S. diabolicum* is paraphyletic because *S. magniae* is nested within it (Shaw et al., 2022). Genome sequencing provides some evidence of introgression among all pairwise combinations of the North American species (Shaw et al., 2022; Piatkowski et al., 2023), but introgression does not appear to be sufficient to scramble the dominant pattern of reciprocal monophylly. Notwithstanding some introgression, the species are strongly resolved as distinct clades.

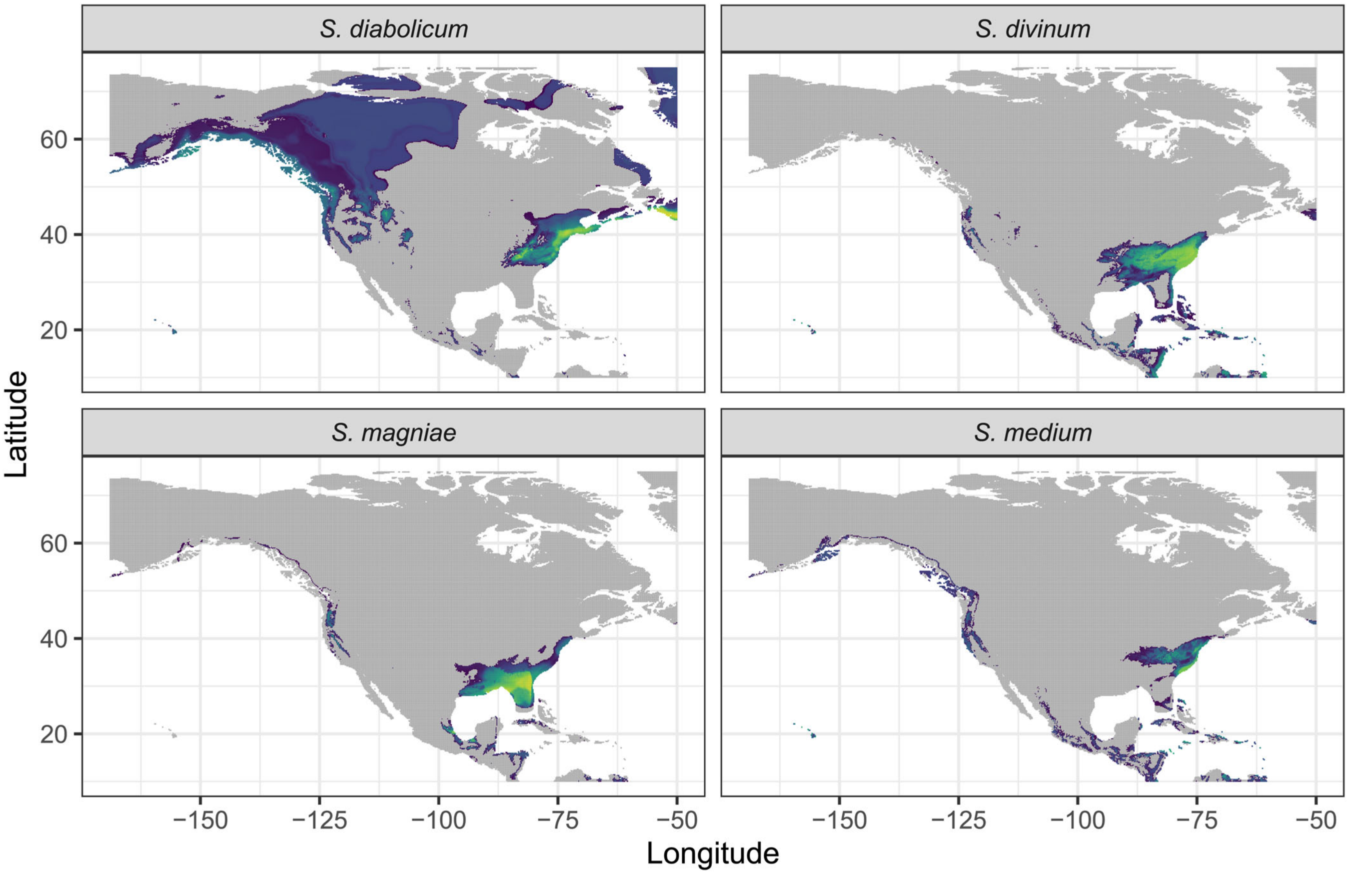

FIGURE 13 Last Glacial Maximum (LGM) distribution maps of species in the *Sphagnum magellanicum* complex in North America. Areas with high suitability in yellow and low in dark blue.

Whether hybridization in the complex is ancient or ongoing is unclear. Focusing on the especially closely related sister species, *S. diabolicum* and *S. magniae*, Piatkowski et al. (2023) used demographic modeling to infer that secondary contact and interspecific hybridization subsequent to speciation (initial reproductive isolation) most likely explains genetic admixture where their ranges overlap. Because species of the complex are so difficult to distinguish morphologically, it is virtually impossible to detect hybrid swarms based on morphology. Genetic analyses of the species at sites where they occur sympatrically are needed to assess whether they hybridize when growing in proximity. We have now identified several sites, and our results indicate that the Pokomoke (MD) populations in Maryland may be critical for future studies of *S. diabolicum* and *S. magniae*, since in these two proximate populations (less than 1 km apart), both species were detected with RADSeq data.

Species identifications in our study were based on barcodes alone for 231 samples, only RADseq data for 413 samples, and both barcodes and RADseq data for 229 samples. Of the 229 samples for which we had both barcode and RADseq data, we obtained conflicting identifications in just 13 cases. The barcodes appear to be very reliable. We were therefore surprised that two samples from Kibbie Lake in central New York were identified as *S. magniae* by our barcode markers. Central New York is well outside the main range of *S. magniae*, latitudinally and climatically. At present, we have RADseq data for one of these plants, which appears to be *S. diabolicum* based on phylogenetic analyses. Similar cases occurred in two populations in Maine (East Hancock and Great Wass Island), where three and one sample(s), respectively, were barcoded as *S. magniae*, but the RADseq data identified them as *S. diabolicum*. We have three alternative hypotheses to explain our observations to date. (1) As appears to be the case with the Michigan sample identified by RADseq data as *S. magniae*, this species really does have isolated disjunct occurrences well outside its main range. (2) Homoplastic mutations at the barcode loci can rarely result in erroneous species identifications. (3) These plants are genetically admixed between *S. diabolicum* and *S. magniae*, but may be predominantly *S. diabolicum* (as expected because of their geographic location) and happen to have the barcode marker of *S. magniae* at locus LG01-S6, which we use to distinguish the two species. These hypotheses are currently being tested.

Sphagnum diabolicum reaches North Carolina in the Appalachian Mountains, but *S. magniae* is widespread in the coastal plain of eastern North America, as it is throughout southeastern coastal areas. RADseq and whole-genome resequencing data showed that plants from the Appalachian Mountains are genetically admixed between *S. magniae* and *S. diabolicum* but are mostly *S. diabolicum* (Piatkowski et al., 2023). Admixed plants in coastal areas are predominantly *S. magniae*. We note that both sites in Maine and New York where barcodes identified plants as *S. magniae* well outside its main range also included unambiguous *S. diabolicum*.

Multitudes of different criteria for inferring species delimitation have been proposed (e.g., Zachos, 2018). We, like most others over the last two decades working on mosses and other plant groups, have relied heavily on phylogenetic inferences to identify clades worthy of species status (Shaw et al., 2023b). Reproductive isolation is the criterion embodied in the biological species concept (BSC; Mayr, 1942), and while using the BSC to delimit taxonomic species can be fraught, information about mating patterns and hybridization (or lack thereof) provides important biological/evolutionary information.

Piatkowski et al. (2023) found that even the closely related sister species, *S. diabolicum* and *S. magniae*, are differentiated by 153 genes that include a total of 74,344 fixed differences, that these genes are physically linked in "islands of differentiation" concentrated on chromosomes 1, 11, and 16, that many of the genes contain signals of divergent natural selection, and that differentiation is significantly correlated with climate variables mostly related to temperature. Many of the fixed differences appear to be under divergent selection, indicating that differential adaptation may underlie the process of speciation. The two species also differ in sex ratios, with *S. diabolicum* having a male bias and *S. magniae* a female bias (Shaw et al., 2023a). Interestingly, *S. diabolicum* and *S. magniae* are the most genetically variable in the complex (Shaw et al., 2022). Sites where species occur sympatrically provide natural laboratories for downstream investigations of hybridization and its consequences, including adaptive introgression. The fact that we detected multispecies communities across a range of environments further enriches the potential value of this group as a "natural experiment".

Although species in the *S. magellanicum* complex are (at present) nearly impossible to distinguish morphologically, they differ in other important biological and ecological traits. *Sphagnum medium* and *S. divinum* occur in both Europe and North America, whereas *S. diabolicum* and *S. magniae* are endemic to eastern North America. *Sphagnum divinum* occurs in both eastern and western North America, whereas *S. medium* has an amphi-Atlantic range. The two most closely related and recently diverged species, *S. diabolicum* and *S. magniae*, have the most divergent ecologies relative to climate.

This study sets the stage for additional research in several directions. More genetic data are needed, RADseq or genome resequencing data, to further characterize many of the plants for which we currently have only barcode data. The barcode data identified peatland sites with multiple species, and sites that harbor species well outside their main geographic ranges. There are many potential populations for further study with three species, and even more with various pairwise combinations of two species. More data will permit the identification of genetically admixed plants at sympatric sites, possible evidence of ongoing interspecific gene flow. Admixed allopatric plants could reflect past introgression. Genetic admixture can reflect incomplete lineage sorting (ILS) or introgression, or some combination of processes. Explicit tests to distinguish ILS from introgression are widely available (e.g., Malinsky et al., 2018, 2021). Experimental work on pure and genetically admixed plants are needed to assess the functional correlates of admixture. The availability of a reference-quality genome and chromosome-level genetic map for *S. divinum* increases the value of this group as a model for linking ecology and genome evolution, and the taxonomic problem of how to deal with morphologically cryptic lineages adds an additional level of intrigue to the group. We are currently generating new RADseq data from many of the accessions included in our present barcode-only data set, but there are many unanswered questions pertaining to both North American and European plants. Two additional species in the complex occur in South America (Shaw et al., 2022), and comparable questions pertain to those species and populations.

AUTHOR CONTRIBUTIONS

M.N.-L., B.A., and A.J.S. conceived and undertook the project. M.N.-L., B.A., S.C.R., S.S., R.W., A.M.D., J.Y., and A.J.S. contributed to fieldwork and making collections. D.N.-L. conducted the niche modeling. M.N.-L., and A.M.D. contributed to lab work and statistical analyses. B.P. developed the barcode makers from genome resequencing data. M.N.-L. and A.J.S. led writing the manuscript.

ACKNOWLEDGMENTS

This research was supported by U.S. National Science Foundation grants DEB-1737899 and DEB-1928514 (PI A. J. Shaw). We thank two anonymous reviewers and Associate Editor for their valuable comments, which helped to improve the manuscript.

DATA AVAILABILITY STATEMENT

Species microniche data are deposited on GitHub: https://github.com/dinilu/Sphagnum_magellanicum_complex. RADseq data are available in the Dryad Digital Repository: https://doi.org/10.5061/dryad.37pvmcvtk (Nieto-Lugilde et al., 2024).

ORCID

Marta Nieto-Lugilde http://orcid.org/0000-0002-1593-3853
Diego Nieto-Lugilde http://orcid.org/0000-0003-4135-2881
Bryan Piatkowski http://orcid.org/0000-0002-1334-8431
Aaron M. Duffy http://orcid.org/0000-0003-0530-6191
Blanka Aguero http://orcid.org/0000-0001-8442-5409
Scott Schuette http://orcid.org/0000-0002-7185-6178
A. Jonathan Shaw http://orcid.org/0000-0002-7344-9955

REFERENCES

Agresti, A. 2007. An introduction to categorical data analysis. John Wiley, Hoboken, NJ, USA.

Bączkiewicz, A., M. Szczecińska, J. Sawicki, A. Stebel, and K. Buczkowska. 2017. DNA barcoding, ecology and geography of the cryptic species of *Aneura pinguis* and their relationships with *Aneura maxima* and *Aneura mirabilis* (Metzgeriales, Marchantiophyta). *PLoS One* 12: e0188837.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* 22: 148–155.

Chase, M. W., N. Salamin, M. Wilkinson, J. M. Dunwell, R. P. Kesanakurthi, N. Haidar, and V. Savolainen. 2005. Land plants and DNA barcodes: Short-term and long-term goals. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360: 1889–1895.

Crum, H. A. 1984. Sphagnopsida, Sphagnaceae. North American flora, series II, part 11, 1–180. New York Botanical Garden, Bronx, NY, USA.

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

Eaton, D. A. R., and I. Overcast. 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* 36: 2592–2594.

Francis, C. M., A. V. Borisenko, N. V. Ivanova, J. L. Eger, B. K. Lim, A. Guillén-Servent, S. V. Kruskop, et al. 2010. The role of DNA barcodes in understanding and conservation of mammal diversity in southeast Asia. *PLoS One* 5: e12575.

Grundmann, M., H. Schneider, S. Russell, and J. Vogel. 2006. Phylogenetic relationships of the moss genus *Pleurochaete* Lindb. (Bryales: Pottiaceae) based on chloroplast and nuclear genomic markers. *Organisms Diversity and Evolution* 6: 33–45.

Hassel, K., M. Kyrkjeeide, N. Yousefi, T. Prestø, H. Stenøien, A. J. Shaw, and K. I. Flatberg. 2018. *Sphagnum divinum* (sp. nov.) and *S. medium* Limpr. and their relationship to *S. magellanicum* Brid. *Journal of Bryology* 40: 1–26.

Healey, A. L., B. T. Piatkowski, J. T. Lovell, A. Sreedasyam, S. B. Carey, S. Mamidi, S. Shu, et al. 2023. Newly identified sex chromosomes in the *Sphagnum* (peat moss) genome alter carbon sequestration and ecosystem dynamics. *Nature Plants* 9: 238–254.

Herrera, N. D., K. C. Bell, C. M. Callahan, E. Nordquist, B. A. J. Sarver, J. Sullivan, J. R. Demboski, and J. M. Good. 2022. Genomic resolution of cryptic species diversity in chipmunks. *Evolution* 76: 2004–2019.

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

Hollingsworth, P. M., S. W. Graham, and D. P. Little. 2011. Choosing and using a plant DNA barcode. *PLoS One* 6: e19254.

Johnson, M. G., G. Granath, T. Tahvanainen, R. Pouliot, H. K. Stenøien, L. Rochefort, H. Rydin, and A. J. Shaw. 2015. Evolution of niche preference in *Sphagnum* peat mosses. *Evolution* 69: 90–103.

Johnson, M. G., and A. J. Shaw. 2015. Genetic diversity, sexual condition, and microhabitat preference determine mating patterns in *Sphagnum* (Sphagnaceae) peat-mosses. *Biological Journal of the Linnean Society* 115: 96–11.

Johnson, M. G., and A. J. Shaw. 2016. The effects of quantitative fecundity in the haploid stage on reproductive success and diploid fitness in the aquatic peat moss *Sphagnum macrophyllum*. *Heredity* 116: 523–530.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–589.

King, J. R., and D. A. Jackson. 1999. Variable selection in large environmental data sets using principal components analysis. *Environmetrics* 10: 67–77.

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

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50: 913–925.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760.

Malinsky, M., H. Svardal, A. M. Tyers, E. A. Miska, M. J. Genner, G. F. Turner, and R. Durbin. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology and Evolution* 2: 1940–1955.

Malinsky, M., M. Matschiner, and H. Svardal. 2021. Dsuite - Fast *D*-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources* 21: 584–595.

Martins, S., M. Sim-Sim, and M. Stech. 2021. The Macaronesian endemic moss *Andoa berthelotiana* (Myuriaceae, Bryophyta): Phylogenetic relationships and cryptic speciation. *Nova Hedwigia* 112: 335–357.

Mayr, E. 1942. Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press, Cambridge, MA, USA.

McQueen, C. B., and R. E. Andrus. 2007. Sphagnaceae. *In* Flora of North America Editorial Committee [eds.], Bryophytes: Mosses, part 1. Flora of North America, 27, 45–101. Oxford University Press, NY, NY, USA.

Melville, J., K. Smith, R. Hobson, S. Hunjan, and L. Shoo. 2014. The role of integrative taxonomy in the conservation management of cryptic species: The taxonomic status of endangered earless dragons (Agamidae: Tympanocryptis) in the grasslands of Queensland, Australia. *PLoS One* 9: e101847.

Michalski, S. G., and W. Durka. 2015. Separation in flowering time contributes to the maintenance of sympatric cryptic plant lineages. *Ecology and Evolution* 5: 2172–2184.

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

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

Nieto-Lugilde, M. 2024. Molecular data of cryptic species in the Sphagnum magellanicum complex (Bryophyta) in North America [Dataset]. *Dryad*. https://doi.org/10.5061/dryad.37pvmcvtk

Nieto-Lugilde, M., S. Robinson, B. Aguero, A. Duffy, K. Imwattana, K. Hassel, K. I. Flatberg, et al. 2022. Morphological–molecular incongruence in *Sphagnum majus* ssp. *majus* and ssp. *norvegicum*. *Bryologist* 125: 294–310.

Ovaskainen, O., G. Tikhonov, D. Dunson, V. Grøtan, S. Engen, B. E. Sæther, and N. Abrego. 2017. How are species interactions structured in species-rich communities? A new method for analysing time-series data. *Proceedings of the Royal Society, B, Biological Sciences* 284: 20170768.

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

Piatkowski, B., D. J. Weston, B. Aguero, A. Duffy, K. Imwattana, A. L. Healey, J. Schmutz, and A. J. Shaw. 2023. Divergent selection fuels genomic differentiation between incipient species of *Sphagnum* (peat moss). *Annals of Botany* 132: 499–512.

Piatkowski, B. T., J. B. Yavitt, M. R. Turetsky, and A. J. Shaw. 2021. Natural selection on a carbon cycling trait drives ecosystem engineering by *Sphagnum* (peat moss). *Proceedings of the Royal Society, B, Biological Sciences* 288: 20210609.

Piper, A. M., J. Batovska, N. O. I. Cogan, J. Weiss, J. P. Cunningham, B. C. Rodoni, and M. J. Blacket. 2019. Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. *GigaScience* 8: giz092.

R Core Team. 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website https://www.R-project.org/

Renner, M. A. M. 2020. Opportunities and challenges presented by cryptic bryophyte species. *Telopea* 23: 41–60.

Rydin, H., and J. Jeglum. 2013. The biology of peatlands, 2nd ed. Oxford University Press, Oxford, UK.

Schär, S., G. Talavera, J. D. Rana, X. Espadaler, S. P. Cover, S. O. Shattuck, and R. Vila. 2022. Integrative taxonomy reveals cryptic diversity in North American *Lasius* ants, and an overlooked introduced species. *Scientific Reports* 12: 5970.

Shaw, A. J., N. Devos, Y. Liu, C. J. Cox, B. Goffinet, K. I. Flatberg, and B. Shaw. 2016. Organellar phylogenomics of an emerging model system: *Sphagnum* (peatmoss). *Annals of Botany* 118: 185–196.

Shaw, A. J., A. Duffy, M. Nieto-Lugilde, B. Aguero, S. Schuette, S. Robinson, J. Loveland, et al. 2023a. Clonality, local population structure and gametophyte sex ratios in cryptic species of the *Sphagnum magellanicum* complex. *Annals of Botany* 132: 77–94.

Shaw, A. J., M. Nieto-Lugilde, B. Aguero, A. M. Duffy, B. T. Piatkowski, J. Jaramillo-Chico, S. Robinson, et al. 2023b. *Sphagnum diabolicum* sp. nov. and *S. magniae* sp. nov.; morphological variation and taxonomy of the "*S. magellanicum* complex". *Bryologist* 126: 69–89.

Shaw, A. J., B. T. Piatkowski, A. M. Duffy, B. Aguero, K. Imwattana, M. Nieto-Lugilde, A. Healey, et al. 2022. Phylogenomic structure and speciation in an emerging model: The *Sphagnum magellanicum* complex (Bryophyta). *New Phytologist* 236: 1497–1511.

Sundberg S., and H. Rydin. 2000. Experimental evidence for a persistent spore bank in *Sphagnum*. *New Phytologist* 148: 105–116.

van Breemen, N. 1995. How *Sphagnum* bogs down other plants. *Trends in Ecology and Evolution* 10: 270–275.

Van der Auwera, G. A., and O'Connor, B. D. 2020. Genomics in the cloud: Using Docker, GATK, and WDL in Terra. O'Reilly Media, Sebastopol, CA, USA.

Vanderpoorten, A., and A. L. Jacquemart. 2004. Evolutionary mode, tempo, and phylogenetic association of continuous morphological traits in the aquatic moss genus *Amblystegium*. *Journal of Evolutionary Biology* 17: 279–287.

Varela, S., R. P. Anderson, R. García-Valdés, and F. Fernández-González. 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* 37: 1084–1091.

Velazco, S. J. E., M. B. Rose, A. F. A. de Andrade, I. Minoli, and J. Franklin. 2022. flexsdm: An R package for supporting a comprehensive and flexible species distribution modelling workflow. *Methods in Ecology and Evolution* 13: 1661–1669.

Ward, P. S., and B. B. Blaimer. 2022. Taxonomy in the phylogenomic era: Species boundaries and phylogenetic relationships among North American ants of the *Crematogaster scutellaris* group (Formicidae: Hymenoptera). *Zoological Journal of the Linnean Society* 194: 893–937.

Williams, P. H., P. Dorji, Z. Ren, Z. Xie, and M. Orr. 2022. Bumblebees of the *hypnorum*-complex world-wide including two new near-cryptic species (Hymenoptera: Apidae). *European Journal of Taxonomy* 847: 46–72.

Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, and T. L. Madden. 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13: 134.

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

Yu, Z., J. Loisel, D. P. Brosseau, D. W. Beilman, and S. J. Hunt. 2010. Global peatland dynamics since the Last Glacial Maximum. *Geophysical Research Letters* 37: L13402.

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

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Voucher information for accessions included in this study.

Appendix S2. Phylogenetic relationships among samples of species in the *Sphagnum magellanicum* complex based on RADseq loci.

Appendix S3. Numbers of species in the *S. magellanicum* complex detected within sites.

Appendix S4. Eastern North American peatland sites with five or more genetically confirmed samples of species in the *S. magellanicum* complex (666 of 808 total samples).

Appendix S5. Individual genetically confirmed samples of sympatric species in the *Sphagnum magellanicum* complex at two Maine peatland sites.

Appendix S6. Scoring of microhabitat variation relative to light levels among species of the *Sphagnum magellanicum* complex.

Appendix S7. Scoring of microhabitat variation relative to height above the water table (HWT) among species of the *Sphagnum magellanicum* complex.

Appendix S8. Results of the PCA of climatic variables in the *Sphagnum magellanicum* complex.

Appendix S9. Frequency distribution of effective sample size and Gelman diagnostic over the β parameters estimated by the Bayesian model.

Appendix S10. Hierarchical model of species communities evaluation.

Appendix S11. Species responses to environmental covariates (β parameters) in the *Sphagnum magellanicum* complex.

How to cite this article: Nieto-Lugilde, M., D. Nieto-Lugilde, B. Piatkowski, A. M. Duffy, S. C. Robinson, B. Aguero, S. Schuette, et al. 2024. Ecological differentiation and sympatry of cryptic species in the *Sphagnum magellanicum* complex (Bryophyta). *American Journal of Botany* 111(9): e16401. https://doi.org/10.1002/ajb2.16401