The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2012, **106**, 137–153. With 6 figures

Genetic analysis of the peatmoss *Sphagnum cribrosum* (Sphagnaceae) indicates independent origins of an extreme infra-specific morphology shift

MATTHEW G. JOHNSON*, BLANKA SHAW, PING ZHOU and A. JONATHAN SHAW

Duke University Biology Department, Box 90338, Durham, NC 27708, USA

Received 28 September 2011; revised 27 November 2011; accepted for publication 27 November 2011

Within *Sphagnum cribrosum*, a dioicous aquatic peatmoss, a unique morphological variant (the 'waveform'), found at only two lakes in North Carolina, has a branching architecture that is extremely differentiated from anything otherwise known in *Sphagnum*, although the plants are microscopically indistinguishable from *S. cribrosum*. At one site where the two morphologies co-occur, 60 years of field observations demonstrate the persistence of each morphology, even where the two forms grow intermixed. We conducted a reciprocal transplant experiment in which waveform and normal plants maintained their divergent morphologies for 8 months. We sampled populations throughout the range and conducted genetic and phylogenetic analyses with microsatellite markers and DNA sequences to investigate the genetic context of the waveform morphology within *S. cribrosum*. Haplotype networks from DNA sequences showed the two waveform populations are separated by 11 substitutions across three loci. Microsatellite analyses using nonparametric clustering and admixture models also indicated genetic dissimilarity between genotypes with waveform morphology at the two lakes. Both molecular datasets suggest that the waveform morphology had at least two independent origins, despite the proximity of the two lakes where it occurs uniquely. Given the clonal nature of the waveform, it is unlikely to form a cohesive evolutionary lineage deserving of taxonomic status. The analysis also revealed a genetically diverse population in Georgia as the potential source of variation found in all other populations of *S. cribrosum*. *Biological Journal of the Linnean Society*, 2012, **106**, 137–153.

ADDITIONAL KEYWORDS: clonality – haploid – morphological shift – population structure – species delimitation.

INTRODUCTION

Evolution can sometimes occur by saltation, and 'hopeful monsters' might be responsible for some of the sudden shifts in body plan evident in the fossil record (Theissen, 2009). Among extant examples, some wild populations of shepherd's purse (*Caspella bursa-pastoris* (L). Medik., Brassicaceae) express a mutation that turns all of its petals into stamens, potentially increasing male reproductive fitness (Hintz *et al.*, 2006). Mutations in the MADS-box family of transcription factors, which control the ABC system of flower arrangement and development, are generally responsible for these abrupt infraspecific phenotypic shifts (Becker & Theissen, 2003). Somatic mutations have long been recognized as sources of variety for horticultural uses; 75 years ago, it was noted that 33% of US patents on plants involved bud mutations (Shamel & Pomeroy, 1936). These morphological shifts need not be caused by genomic mutations because heritable gene expression shifts may occur as a result of epialleles that differ in DNA methylation status despite no difference in coding sequence. For example, epiallelic variation underlies the transformation from bilateral-to-radial symmetry in flowers of *Linaria vulgaris* Hill. (Cubas, Vincent & Coen, 1999). In several species of bryophytes, isolated cases of extreme infraspecific variance have included traits that would otherwise be diagnostic of genera,

*Corresponding author. E-mail: mgj4@duke.edu

yet resolve molecularly within (*Leptodon corsicus* Enroth, A. Sotiaux, Quandt, & Vanderp.; Sotiaux *et al*., 2009) or identical to (*Platyhypnidium mutatum* Ochyra & Vanderp.; Stech & Frahm, 1999; *Thamnobryum angustifolium* (Holt) Nieuwl.; Olsson *et al*., 2009) existing species.

Sphagnum L. (peatmosses) is characterized by free-living gametophytes with a highly distinctive gametophyte bauplan that is consistent among the 200–300 extant species (McQueen & Andrus, 2007). Dense clusters of branches form a distinct capitulum at the apex of a main stem and lower branches are organized in fascicles of (typically) three or four branches (Fig. 1A). The branches in each fascicle are more or less differentiated into 'spreading' and 'pendent' branches that differ in orientation and also in the anatomy of the leaves they bear. Main stem bifurcation is infrequent, although it is assumed to facilitate clonal propagation of gametophytes. The leaves, which are one cell thick as in almost all other mosses, have two cell types in *Sphagnum*: small green chlorophyllose cells, and large, dead, empty hyaline cells that take up water through pores in the cell walls. Species identification in *Sphagnum* generally requires microscopic dissection because related species may be only subtly different at the macroscopic scale. For example, in the subgenus *Subsecunda* (Lindb.) AJ Shaw, *Sphagnum cribrosum* Lindb. and its sister species *Sphagnum macrophyllum* Bernh. ex Brid. have almost identical, typical *Sphagnum* bauplans, and can only be distinguished by the size and arrangement of pores on hyaline cell walls (Zhou, Menzel & Shaw, 2007; Anderson, Shaw & Shaw, 2009).

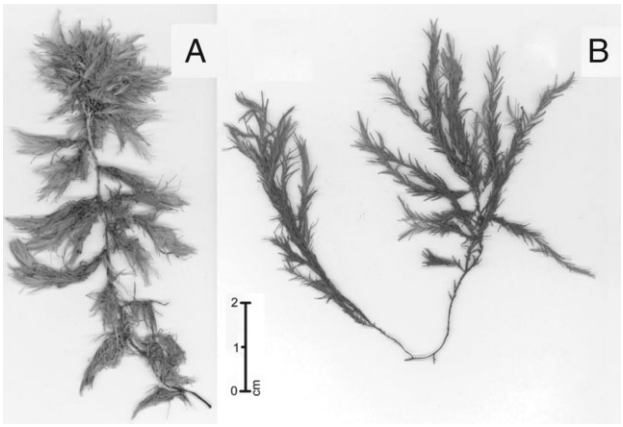

Figure 1. Illustration of the drastic morphological shift between the normal and waveforms of *Sphagnum cribrosum*. A, normal form, with stereotypical *Sphagnum* bauplan: distinct capitulum, infrequent branching of the main stem, and leaves arranged in fascicles (L. E. Anderson 23827). B, waveform, characterized by the lack of distinct capitulum, frequent branching of the main stem, and leaves not arranged in fascicles (B. Shaw 4308). Both samples are from Singletary Lake in Bladen County, NC, USA.

Divergence from the general *Sphagnum* bauplan is rare, yet, in two shallow lakes ('Carolina Bays') in North Carolina, plants of the aquatic, dioicous peatmoss *S. cribrosum* have radically divergent growth architecture relative to normal forms (Fig. 1B). This morphotype has frequent but irregular branching, no fascicles, and no distinct capitulum. It nevertheless has the microscopic features of *S. cribrosum*; it was nicknamed 'waveform' by bryologists who surmised that branching pattern is a plastic response to wave action along the margins of the two Carolina Bays where it occurs (Crum & Anderson, 1981). At one site, Singletary Lake, the waveform co-occurs with the 'normal form' of *S. cribrosum*; in an outlet stream from the lake, both forms grow intimately intermixed. At the nearby (15 km distant) Jones Lake, only the waveform occurs.

Sphagnum cribrosum commonly grows floating in roadside ditches, cypress (*Taxodium* L.C. Rich) swamps, and along the margins of lakes in the coastal plain from North Carolina to Florida. In North Carolina, *S. cribrosum* is most commonly found near the margins of moderately sized, shallow lakes known as Carolina Bays. The typical growth form of *S. cribrosum* gametophytes is similar to many *Sphagnum* species; branches in fascicles of two or three originate in tight clusters at the apex, forming a cohesive capitulum (Fig. 1A). The lower fascicles have little or no differentiation of pendent and spreading branches but, aside from the two waveform populations, plants are always fasciculate with a terminal capitulum. Sexual reproduction (evidenced by sporophyte formation) in *S. cribrosum* is very rare and is currently known from only two sites in Georgia. Sporophytes have never been observed in North Carolina despite over 70 years of relatively intensive collecting (Anderson *et al*., 2009; data available at: http://herbarium.duke.edu).

At Singletary Lake (34.5975°N, 78.4548°W), normal, fasciculate and capitulate *S. cribrosum* grows in an outlet stream approximately 75 m from the open lake. However, in the lake itself, the waveform morphotype covers twigs and roots of trees and shrubs that occur below water level (Fig. 2). These waveform plants have no discernable capitulum, lack branch fascicles, and, instead, the gametophyte stem branches so frequently that it does not resemble a *Sphagnum* at all in gross-morphology (Fig. 1B). Samples in the Duke Bryophyte Herbarium (DUKE) were originally misidentified as the aquatic true-moss (class Bryopsida), *Fontinalis* Hedw. At the microscopic level, however, the leaves of these plants possess differentiated hyaline and chlorophyllose

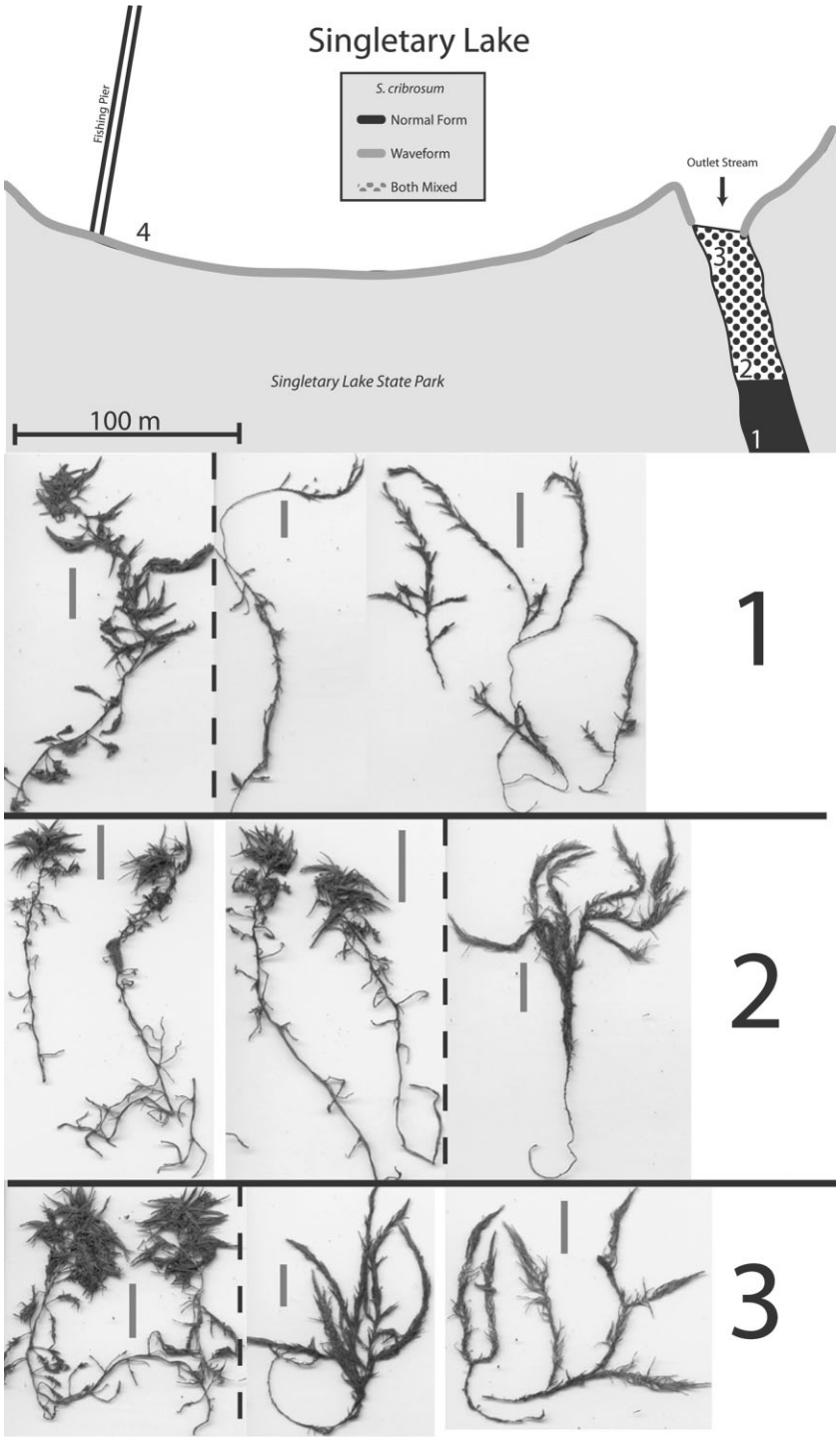

Figure 2. Design and representative results from common growth experiment at Singletary Lake State Park, Bladen County, NC. Top: schematic showing sampling and transplant locations, and the location of waveform and normal form in the lake margin and outlet stream. All waveform plants were sampled from Site 4; all normal form plants were sampled from Site 1. Eight plants of each type were grown at each of the four sites. Bottom: Representative plants showing new growth after 8 months at Sites 1, 2, and 3. Observe the distinct capitulum in normal form plants (left of dashed line) and the frequent branching with no distinct capitulum in waveform plants (right of dashed line). In each subfigure, the grey bar represents one inch (2.5 cm).

cells, with the latter having cell wall pores. Both features are hallmarks of plants in the Sphagnopsida. In all details of leaf and cell structure, the plants match normal growth forms of *S. cribrosum*. This unique waveform morphotype has a long collection record; samples from 1934 at Singletary Lake (*H.L. Blomquist 3105*) and 1974 at Jones Lake (34.6947°N, 78.6089° W; *A. Rushing 116*) are preserved in the Duke herbarium. The label 'waveform' for these plants was coined by L. E. Anderson (pers. comm.) who assumed that the unique growth form was an environmentally-induced response to wave action in these shallow but rather large lakes. We have explored almost all of the Carolina Bays in Bladen County and surrounding areas; the waveform has been found only at Singletary and Jones Lakes. Indeed, intensive exploration of appropriate sites throughout the range of *S. cribrosum* from New Jersey to the Gulf coast has not revealed any other populations with the waveform morphology.

The rarity of the waveform, the proximity of the two waveform populations, and the lack of sexual reproduction all suggest a single origin for this unique morphotype. The most likely scenario is an origin at Singletary Lake (where it is very abundant and co-occurs with the normal form), followed by dispersal to Jones Lake (where it is much less common and more or less restricted to parts of the lake near a public swimming area). If the waveform is monophyletic, and remains reproductively isolated (as a result of obligate asexual reproduction), it may be an incipient species. We tested plasticity of the waveform morphology using a reciprocal transplant experiment at Singletary Lake. We sampled populations throughout the range of *S. cribrosum* to test the genetic relationships of the waveform morphotype with the normal form; monophyly of DNA sequences and low genetic distance between the waveform populations would confirm the single origin hypothesis. Our data are also sufficient to test the hypothesis that Singletary Lake waveform plants originated *in situ* from sympatric normal plants.

MATERIAL AND METHODS

Population sampling

For the genetic study, 206 samples representing 21 populations of *S. cribrosum* were collected from throughout the part of its range where it is most common, from western Florida to eastern North Carolina. All but five populations were sampled with multiple individuals, and more than five plants were sampled from eight populations. One site, near Ludowici, Georgia (31.7236°N, 81.7269°W, designated GA32), was represented by 79 samples, collected as part of an ongoing study of potential hybridization between *S. cribrosum* and *S. macrophyllum*. At Singletary Lake, 26 specimens of normal form and 31 specimens of waveform were sampled, whereas, at Jones Lake, four waveform plants were sampled. (The normal form does not occur at Jones Lake and waveform plants are not as abundant.) All plant material is preserved in the DUKE; collection and GenBank accession numbers are provided in the Appendix. The actual plant from which DNA was extracted was placed in a small envelope and replaced within the larger herbarium packet. Vouchers are identified as such in the bryophyte herbarium database. In addition to plants collected for the present study, four DNA sequences from normal and waveform plants from Singletary Lake, from Zhou *et al*. (2007), were included in the present analyses (see Appendix). For phylogenetic analysis, ten specimens of sister species *S. macrophyllum*, all from the GA32 population, were included as an outgroup.

Reciprocal transplant experiment

To test whether the waveform at Singletary Lake is a plastic response to wave action on the lake margin, a reciprocal transplant experiment was initiated in February 2006. Plants of each form were randomly sampled and grown at four areas. The normal form was sampled far down the outlet stream (Site 1; Fig. 2) where no waveform occurs, whereas the waveform was sampled from the lake margin where no normal form occurs (Site 4; Fig. 2). Samples were then grown at these two sites, as well as at two intermediate sites (Sites 2 and Site 3; Fig. 2). At each experimental location, six plants of each morphotype, trimmed to 2.5 cm in length, were placed in individual small mesh bags that allow free water movement. That is, the plants were subjected to natural wave action at each site. Site 1 experiences little or no wave action, Site 4 experiences maximum wave action, and Sites 2 and 3 experience intermediate levels. The bags containing experimental plants were tied to swimming pool 'fun noodles' that floated on the surface, tied to a polyvinyl chloride pipe anchored in the lake bottom. The plants were re-examined after 8 months, in October 2006. Plants were scored as living if green tissue was visible, and the morphotype of each plant was noted by observing the branching frequency and whether the plant had a distinct capitulum. Plants of intermediate morphology were not observed, so this qualitative scoring was sufficient to record morphological responses to the experimental treatments. Plants were dried and weighed to determine if biomass differences existed between sites for each morphotype. Statistical tests [*t*-test and analysis of variance (ANOVA)] were conducted using R software (R Development Core Team, 2011).

Biomass was selected as the growth response estimator. Ideally, dry weights for each sample would have been measured before the experiment, although this was not feasible because *S. cribrosum* is not desiccation tolerant. Each experimental plant was started from a 2.5 cm length of stem. Waveform plants have more biomass per unit length than normal form plants. Subsequent comparisons were made conservatively within a growth form across in the ANOVA. Plants with no observed new growth (presumed dead) were removed from the analysis.

DNA extraction and amplification

A portion of the apex of selected plants was removed for DNA extraction using a modified CTAB protocol (Shaw, Cox & Boles, 2003). For the nucleotide sequence dataset, polymerase chain reaction amplification of three nuclear DNA regions – the so-called RapdA, RapdB, and RapdF loci (Shaw *et al*., 2003), was accomplished using methods described previously Zhou *et al*. (2007) for 138 samples. The sequences were aligned by eye in PHYDE (Muller *et al*., 2007) and resulted in a total of 2418 bp across the three loci. Genbank accession numbers for all sequences are provided in the Appendix. For the microsatellite dataset, amplification of fourteen *Sphagnum*-specific microsatellite regions for all samples followed the protocol described by Shaw *et al*. (2008) for the loci: 1, 4, 7, 9, 10, 12, 14, 17, 18, 19, 20, 22, 29, and 30. Genotyping was accomplished with an ABI 9600 sequencer (Applied Biosystems), visualized and binned using GENEMARKER (Softgenetics). The microsatellite data set and sample information can be found at doi: 10.5061/dryad.860jq72n (Johnson *et al.*, 2012).

DNA sequence analysis

An initial unweighted pair group method with arithmetic mean (UPGMA) tree was constructed for all samples using PAUP*, version 4.0a109 (Swofford, 2003). Some samples shared identical sequences across the three loci. For each of the unique haploid multi-locus genotypes (UHMG), only one accession per UHMG per population was retained, and further phylogenetic analyses were conducted with 55 *S. cribrosum* UHMGs plus nine UHMGs of the sister species *S. macrophyllum* as outgroups. Maximum likelihood (ML) reconstruction was accomplished with GARLI, version 0.95 (Zwickl, 2006). Substitution, base frequency, and rate variation models were chosen using likelihood ratio tests implemented in MODELTEST, version 3.7 (Posada & Crandall, 1998) with likelihood scores calculated in PAUP*, version 4.0a109 (Swofford, 2003). The best model chosen for the random amplified polymorphic DNA markers using Akaike information criteria was the transversional model, with invariant sites and gamma-distributed rate variation. However, this model is not implemented in GARLI, version 0.95, so the general time reversible model (which had an identical likelihood score, with one additional parameter) was used with a proportion of invariant sites and rates distributed via a gamma distribution with four rate categories (GTR + I + G). Bootstrap values were calculated with 200 pseudoreplicates in GARLI.

TCS, version 1.21 (Clement, Posada & Crandall, 2000) is a statistical parsimony method for reconstructing haplotype networks. It can be used to assess the number of DNA substitutions necessary to represent the distance between all unique haplotypes. With haploid plants, the phase of each gene is known, and plants were sorted into UHMG groups via a 95% sequence similarity cut-off.

Microsatellite analysis

Genetic diversity summary statistics were computed using GENALEX, version 6.3 (Peakall & Smouse, 2006); missing data were not interpolated. Linkage disequilibrium among loci was estimated with MULTILOCUS, version 1.3b (Agapow & Burt, 2001). For these analyses, missing data were fixed, although other missing data options did not change the results.

Cluster analyses were undertaken using both parametric (STRUCTURE, version 2.2.1; Pritchard, Stephens & Donnelly, 2000) and nonparametric (AWCLUST; Gao & Starmer, 2008) methods. For STRUCTURE, ten replicate runs of one million generations followed a 250 000 generation burn-in, for several values of *K*. Cluster assignment and admixture across the ten replicate runs was summarized using CLUMPP, version 1.2.2 (Jakobsson & Rosenberg, 2007), and visualized with DISTRUCT, version 1.1 (Rosenberg, 2004). The analysis was conducted with all 206 samples and with a subset representing 109 UHMGs. As a haploid-dominant organism capable of extensive clonal growth, allele frequencies in the gamete pools of *S. cribrosum* will be affected by the relative growth rates of each clone. Although the manual for STRUCTURE indicates 'family members' should not be included, sampling within populations was random and therefore reflects the genetic diversity. The utility of STRUCTURE analysis is two-fold: (1) to examine what genetic context the waveform occupies within *S. cribrosum* and (2) to determine the overall pattern of admixture in the species. The complete dataset is more appropriate for the former objective, whereas using only UHMG can help answer the latter.

The nonparametric clustering technique employed in the R-module AWCLUST (Gao & Starmer, 2008)

has the advantage of not requiring assumptions about Hardy–Weinberg equilibrium and is based not on allele frequency variation, but on hierarchical clustering of a distance matrix, visualized through non-metric multi-dimensional scaling. Microsatellite alleles were converted to a presence/absence dataset where each allele was treated as a 'locus,' with a 0 or 1 corresponding to the allele an individual carries. After the distance matrix was calculated, each individual was assigned to exactly one cluster. The optimal number of clusters was determined by a gap statistic that for each value of the *K*-statistic is the difference between the pooled within-cluster sum of square distance between individuals, and a null reference distribution for that value of *K*. The statistic is greatest at the optimal *K*-value. Cluster assignment and geographical distribution of the clusters were visualized using custom R scripts and the MAPS package (Becker & Wilks, 1993).

RESULTS

Reciprocal transplant experiment

After 8 months, the biomass, vitality and morphotype of each sample was assessed. Unfortunately, survival at several sites was poor, with some plants of both types showing no new growth. Biomass (dry weight) was measured after the experiment (Table 1), and indicates that normal form plants were significantly larger (Welch two sample *t*-test: $t = 3.3933$, $P < 0.01$) in their native habitat (mean ± SD; Site 1, 0.323 ± 0.117 g) than in the waveform habitat (Site 4, 0.130 ± 0.031 g). However, waveform plants did not show this differentiation between extreme sites (Welch two sample *t*-test: $t = 1.1567$, $P > 0.2$). Direct comparisons of biomass between morphotypes is not possible because waveform plants are less bushy and generally have less biomass per unit length than normal form plants. However, an ANOVA demonstrated a significant morphotype × experimental site interaction ($F = 4.10$, $P < 0.05$). This indicates that the two morphotypes responded differently to the four sites, although additional experiments are necessary to assess whether there is local adaptation at Singletary Lake.

In the plants with new (green) growth at each site, morphology (normal or waveform) was maintained. A representative selection of experimental plants from three of the sites (Sites 1, 2, and 3; Fig. 2) shows no indication of morphotype reversal. Normal form plants maintained capitula through new growth. Critically, waveform plants from Site 4 maintained waveform morphology after 8 months at Site 1 (where only normal form occurs). None of the waveform plants began to form a cohesive capitulum, and none of the normal form plants began to branch irregularly. An environmental carryover effect is possible because of the limited duration of the experiment, although waveform plants grow side-by-side with normal form plants in the outlet stream near Sites 2 and 3, and these mixtures have been observed for many years (A. J. Shaw & L. E. Anderson, pers. observ.).

DNA sequence dataset

Combined analyses for RapdA, RapdB, and RapdF under ML indicate that waveform UHMGs from the two lakes where they occur are not monophyletic (Fig. 3). However, only the branch leading to the in-group has likelihood bootstrap support exceeding 75% in 1000 replicates. In an analysis of in-group (*S. cribrosum*) specimens only, the maximum log-likelihood score of an unconstrained tree was –3619.87. When the waveform samples from Jones and Singletary Lakes were constrained to be monophyletic, the maximum log-likelihood decreased to –3637.20. A Kiroshima–Hasegawa test for significance of the difference between the ML tree versus the constraint tree had a *P*-value of 0.04, indicating that the constraint tree was significantly less likely. The less biased Shimodaira–Hasegawa test, with 10 000 RELL bootstrap replicates, was also calculated

Table 1. Dried biomass (g) of plants following an 8-month common garden experiment

	Site 1	Site 2	Site 3	Site 4
Normal form	0.342	0.342	0.252	0.139
	0.383	0.383	0.167	0.164
	0.442	0.442	0.430	0.106
	0.409	0.409	0.072	0.090
	0.148	0.148	0.187	0.149
	0.213	0.213	0.195	(d)0.154
Average	0.323	0.323	0.217	0.130
Waveform	0.054	(d)0.003	(d)0.009	0.024
	0.045	(d)0.009	0.019	0.045
	0.035	(d)0.008	0.144	0.150
	0.067	(d)0.008	0.023	0.168
	0.011	0.038	0.007	0.035
	0.023	0.032	0.047	0.014
Mean	0.039	0.035	0.048	0.073

The identity of each site is shown in Fig. 2: Sites 1 and 2 typically have only normal form plants; Site 3 has a mixture of normal and waveform plants, and Site 4 (lake margin) has only waveform plants. At the beginning of the experiment, plants of each type were taken from Sites 1 and 4 and transplanted at each site, trimmed to 2.5 cm. (d) represents plants that showed no new growth (presumed dead) after the experiment, and were removed from analysis.

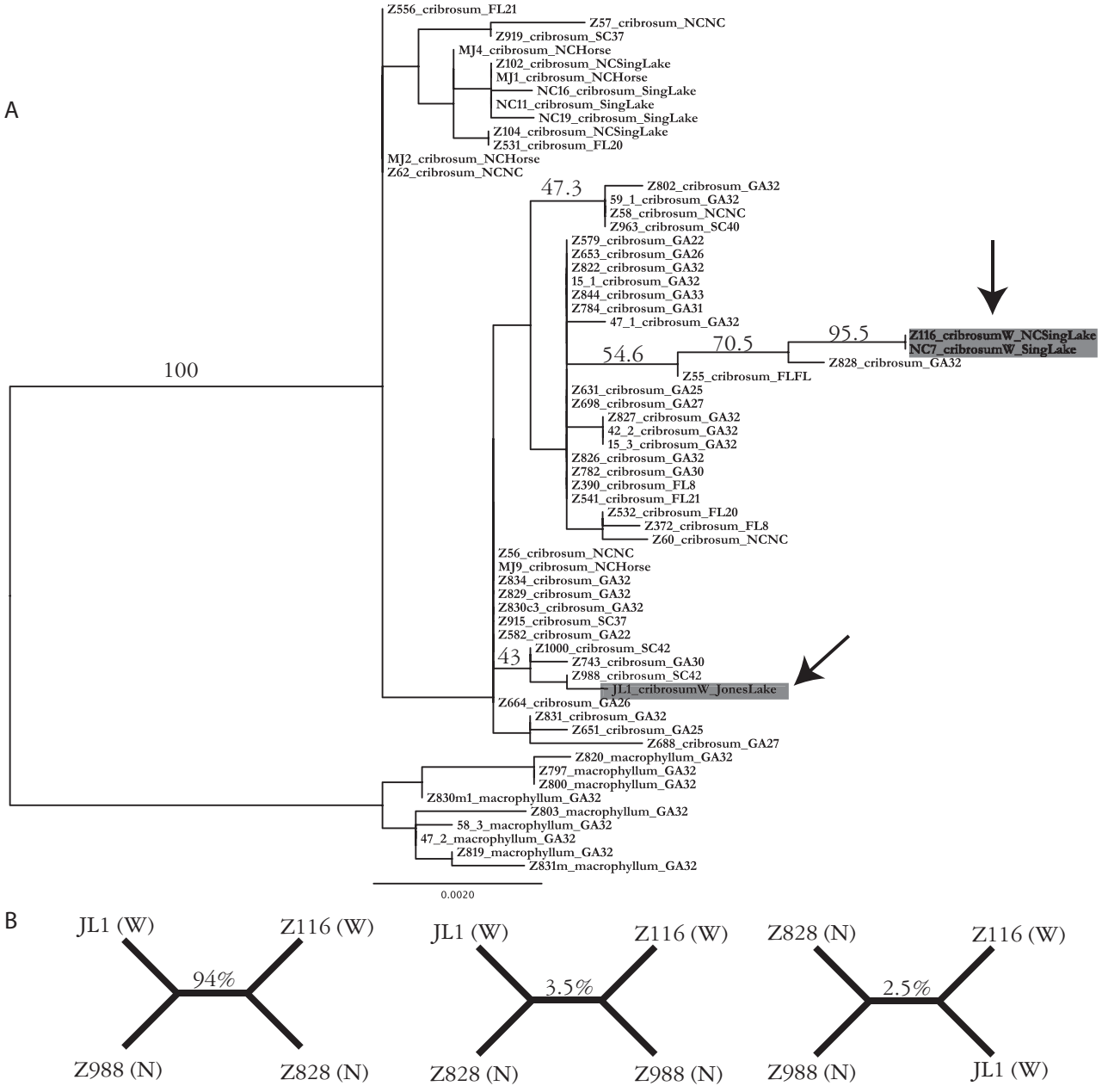

Figure 3. A, maximum likelihood phylogeny of 137 samples of *Sphagnum cribrosum*. Constructed with GARLI, using three anonymous nuclear markers (RapdA, RapdB, RapdF, 2418 bp). Sister-species *Sphagnum macrophyllum* is used as the outgroup. Numbers indicate likelihood bootstrap support (1000 replicates) exceeding 40% within *S. cribrosum*. Grey boxes and arrows indicate waveform samples. B, reduced consensus trees showing rejection of waveform monophyly. For each waveform sample, a related normal form sample was chosen; all other samples were pruned from each of the 1000 bootstrap trees using PAUP. Percent support is indicated above the internal branch; the tree showing monophyly of the waveform samples (right) has support of only 2.5%.

using PAUP*, version 4.0a109 (Swofford, 2003). This analysis compared the 100 best trees from a likelihood search with the constraint tree, and the test indicated that the difference was (barely) not significant ($P = 0.065$). The hypothesis that Jones and Singletary Lake waveforms form a single monophyletic group cannot be rejected at $P \leq 0.05$ by this more conservative, although less-biased test.

Although none of the informative branches on the ML tree show bootstrap support exceeding 75% within *S. cribrosum* (Fig. 3A), several branches show moderate support (between 40% and 70%). A

further test of waveform monophyly is possible using a reduced consensus approach (Wilkinson, 1996), as well as by pruning 'problematic' taxa. Should there be high support for any branch between the waveform samples, waveform monophyly can be rejected. Moderate support groups a normal form sample from Georgia (Z988) with Singletary Lake waveform (Z116 and NC7); this support increases to 75% if the nearby normal form sample Z55 is pruned from the bootstrap trees. To fully illustrate the support for the branch located between the waveform samples (grey boxes in Fig. 3A), two waveform samples (Z116 from Singletary Lake and JL1 from Jones Lake) were chosen along with a normal form sample closely related to each (Z988 and Z828, respectively). In each of the 1000 bootstrap trees, all other samples were pruned from the tree in PAUP, and support was assessed on each of the three possible four-taxon trees (Fig. 3B). The ML topology, with waveform non-monophyletic, was found in 94% of the bootstrap trees (Fig. 3B, left). The topology where waveform was still non-monophyletic but the normal form samples switched positions (Fig. 3B, centre) was found in 3.5% of the bootstrap trees. Finally, the reduced tree showing monophyly of waveform (Fig. 3B, right) was found in just 2.5% of the replicates. The high support for non-monophyly of waveform is a result of combining moderate support on several branches separating the waveform samples (Fig. 3A). Selection of alternative samples of *S. cribrosum* normal form reduced the strength of the evidence for non-monophyly (results not shown). Nevertheless, these analyses suggest that monophyly of waveform is highly unlikely.

Reconstruction of the UHMG networks using statistical parsimony in TCS revealed large distances (in DNA substitutions) between waveform UHMGs from Jones versus Singletary Lakes (Fig. 4). Waveform samples comprise two groups: one at Singletary Lake and one at Jones Lake. The Jones Lake waveform genotype is separated from the Singletary Lake waveform genotype by at least 11 substitutions. Four UHMGs detected among Singletary Lake normal form plants are more closely related to the Jones Lake waveform than to the sympatric Singletary Lake waveform genotype.

Further evidence for the genetic distance between waveform and normal form at Singletary Lake was evidenced by a 25-base minisatellite within the RapdF locus (Table 2). Three samples (Z57, from North Carolina; Z531 and Z556, both from Florida) have three copies of the repeat. Every normal form sample at Singletary Lake contains two copies of the repeat, a trait shared with several other normal form samples in North and South Carolina (MJ1, MJ2, MJ4, Z62, Z919). Singletary Lake waveform samples, however, contain just one copy of the repeat, which is the most common repeat number in *S. cribrosum*. The waveform at Singletary Lake is additionally differentiated from nearly all other samples of *S. cribrosum* by a single base substitution (i.e. single-nucleotide polymorphism) in the fourth position of the repeat region. Only Z828 from GA32 and Z55 from Florida share the substitution. Jones Lake waveform plants also have just one copy of the repeat, although they have the more common form of the single-nucleotide polymorphism. Because TCS treats gaps as missing data, the minisatellite variation was not reflected in the network, and therefore this genome structural feature provides additional evidence of genetic distance between normal form and waveform at Singletary Lake not reflected in Figure 4, nor reflected in the phylogenetic tree (Fig. 3).

Ten UHMGs from the population near Ludowici, GA (GA32), are scattered across the network (Fig. 4). Five UHMGs from this population also contain plants found in all four states sampled. Thirty-six plants (35% of the normal form sequences) have genotypes identical to plants in GA32. The other five UHMGs found at GA32 are unique to that population, although none of these UHMGs differ by more than two substitutions from plants found in other populations. Samples from GA32 can also be found in all but one clade resolved in the ML phylogeny (Fig. 3).

Microsatellite dataset

Genetic diversity within populations varied from site to site across the range of *S. cribrosum* (Table 3) but was clearly highest at the GA32 site (Information Index = 0.717, Shannon's Haploid Diversity = 0.392). The sample size in this population was much higher than for others (79 accessions compared to 21 for the next most abundantly sampled population), so to correct for this, data from the GA32 site were randomly 'subdivided' into eight populations of nine or ten individuals, and the diversity measures were repeated on the subdivisions. This subdivision process was repeated five times and the mean for each of the genetic diversity measures within each random subdivision was still higher than the next most diverse population (Table 3). This population was also the only one characterized by private alleles; seven private alleles at six loci occurred at frequencies between 0.013 and 0.203 within the GA32 population.

Clonality is high in many populations, including GA32, where 79 specimens can be represented by 29 UHMGs. Clonality is also high at Singletary Lake, where the 21 normal form samples comprise three UHMGs, whereas the 28 waveform samples belong to just two UHMGs. By contrast, two populations in Georgia have as many UHMGs as samples (Table 3).

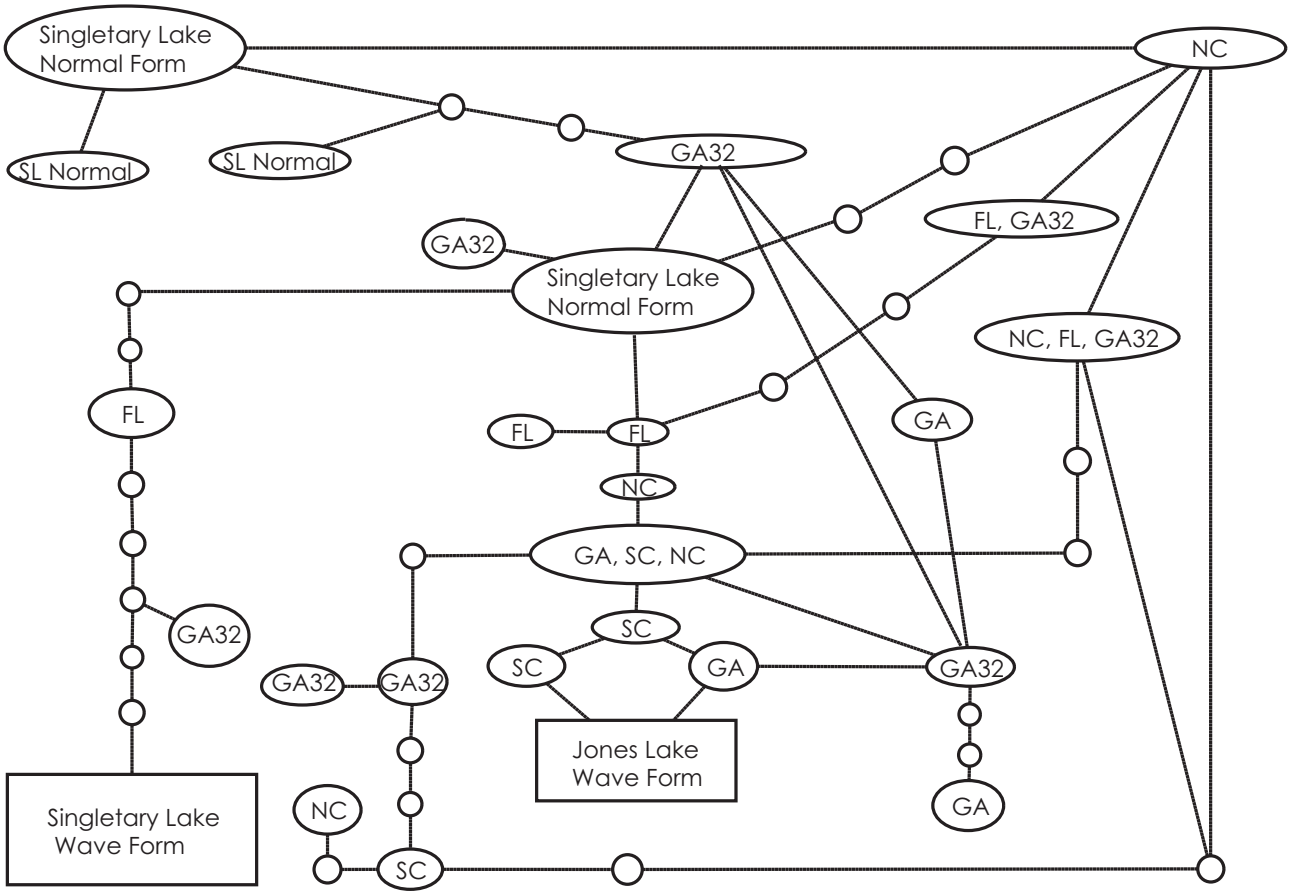

Figure 4. Haplotype network of DNA sequence substitutions for RapdA, RapdB, and RapdF. Ovals indicate populations or individual samples of *Sphagnum cribrosum*; the size of the oval corresponds to the number of samples within a haplotype group. Rectangles indicate haplotype groups containing one or more samples of waveform morphology. The number of DNA substitutions between haplotypes is shown as the number of ovals, rectangles or circles between the haplotypes. Generated using the statistical parsimony software TCS. FL, Florida; GA32, Ludowici, Georgia population; GA, Georgia (GA32 and other populations); NC, North Carolina populations (other than Singletary and Jones Lakes); SC, South Carolina.

Table 2. Separation of waveform and normal form at Singletary Lake by RapdF minisatellite (25 bp)

	Number of repeats	First repeat, fourth base	Other samples with same type
Jones Lake waveform	1	A	All other *S. cribrosum*
Singletary Lake waveform	1	T	GA32: Z828; FL: Z55
Singletary Lake normal form	2	A	NC: MJ1, MJ2, MJ4, Z62; SC: Z919
3-Repeat type	3	A	FL: Z531, Z536; NC: Z57

A substitution in the fourth base of the first repeat occurs in Singletary Lake waveform; all other samples have an A at this position. The location and sample identity of other specimens with a minisatellite type are also listed.

Despite clonality within most populations, multilocus linkage disequilibrium across the range of *S. cribrosum* was low, albeit significantly above zero using both the full datasets (rBarD = 0.0933, $P < 0.0001$) and unique genets only (rBarD = 0.0781, $P < 0.0001$). No within-plant heterozygosity was observed, consistent with the haploid cytological condition of *S. cribrosum* gametophytes.

Results from STRUCTURE with the complete dataset indicate that the optimal value of $K = 7$

Table 3. Genetic diversity indices for the six populations of *Sphagnum cribrosum* for which there were at least five samples

Population	N	N_A	N_E	I	h	UHMG
GA32	75.714 ± 2.754	3.357 ± 0.509	2.086 ± 0.317	0.717 ± 0.144	0.392 ± 0.074	29
GA32 subsamples	9.464 ± 0.342	2.407 ± 0.180	1.900 ± 0.107	0.612 ± 0.052	0.357 ± 0.027	
SL_Normal	19.643 ± 0.357	1.357 ± 0.133	1.305 ± 0.116	0.231 ± 0.086	0.163 ± 0.061	3
GA25	8.000 ± 0.000	1.714 ± 0.244	1.613 ± 0.209	0.402 ± 0.122	0.261 ± 0.075	8
GA30	11.214 ± 0.786	1.857 ± 0.275	1.454 ± 0.144	0.369 ± 0.111	0.226 ± 0.068	11
FL8	7.000 ± 0.000	1.786 ± 0.261	1.443 ± 0.161	0.351 ± 0.118	0.207 ± 0.070	5
SLWave	27.571 ± 0.228	1.071 ± 0.071	1.011 ± 0.011	0.018 ± 0.018	0.009 ± 0.009	2
JLWave	4.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	1

Data are the mean ± SE values across 14 microsatellite loci. N, sample size (corrected for missing data); N_A, number of different alleles per locus; N_E, effective number of alleles per locus; I, Information Index; h, Shannon's Haploid Diversity Index; UHMG, unique haploid multilocus genotypes. For the population GA32, values are for all 79 samples (first line) and for the mean from forty 'populations' of ten specimens randomly subsampled from GA32 (second line).

(Fig. 5, left), from which a lack of geographical structure is apparent (Fig. 5, right). Groups A, B, D, and E represent populations from Florida, Georgia (other than GA32), South Carolina, and North Carolina (other than Singletary and Jones Lakes), respectively. Admixture among these groups is high, indicating a lack of geographical structure. The remaining groups in Figure 5 (Groups C, F, G, and H) are from individual populations with extensive sampling. The uniformity of samples in Group G indicate that waveform samples at Singletary Lake are genetically very similar, whereas the high amount of admixture in Group C is indicative of high genetic diversity and sexual reproduction at GA32.

Waveform samples from the two lakes (individuals in groups G and H in Fig. 5) are not in the same cluster, whereas normal form samples from Singletary Lake (F) show more similarity with Jones Lake waveform than with Singletary Lake waveform. The genetic group that includes Jones Lake waveform plants (pink in Fig. 5) is found in every geographical region, although it is differentiated from Singletary Lake waveform plants. Admixed individuals possessing significant contributions from each of the seven clusters were found in the GA32 population, and individuals belonging predominantly to the blue cluster in Figure 5 were found exclusively at GA32.

When the STRUCTURE analysis was repeated with only the 109 UHMG, the optimal K as determined by the delta-K method was $K = 4$ (results not shown). However, at all values of K, individual cluster assignment was evenly split between each cluster; for example, at $K = 2$, all but four UHMG showed between 45% and 55% membership in the two clusters (results not shown). This pattern is consistent with the low levels of linkage disequilibrium found in the UHMG dataset, and indicates that across the south-eastern USA, there is no evidence of geographical structure within *S. cribrosum*.

The nonparametric clustering technique, AWCLUST, separated the 187 individuals into 11 clusters (Fig. 6). AWCLUST uses a 'gap statistic' to calculate the optimal number of clusters; however, the software was unable to find an optimal number because the gap statistic was still rising at the software's maximum value of K = 8. For illustrative purposes, K = 11 was chosen, and is reflected in the unrooted UPGMA tree as a horizontal line (Fig. 6B). The chief observation from the hierarchical clustering is the genetic distance between waveform at Singletary Lake and Jones Lake. If the horizontal line in Figure 6B were moved up the tree (reflecting fewer clusters), the waveform populations would not be part of the same cluster until K = 1.

With two exceptions (including waveform plants from Singletary Lake), every population with multiple samples had members in multiple genetic clusters. Waveform plants at Jones Lake belong to a cluster that also occurs in Georgia (including GA32) and Florida. Individuals from the GA32 population were assigned to ten of the eleven clusters; two of these clusters (shades of green in Fig. 6A) are found only at GA32. Other values of K in AWCLUST produced similar results; all showed that population GA32 includes specimens that belong to many different genetic clusters (results not shown).

DISCUSSION

Multiple sources of evidence together indicate the waveform is both genetically based and has at least two independent origins: (1) No reversal to normal form was observed when waveform plants were grown in a common garden with normal form plants at

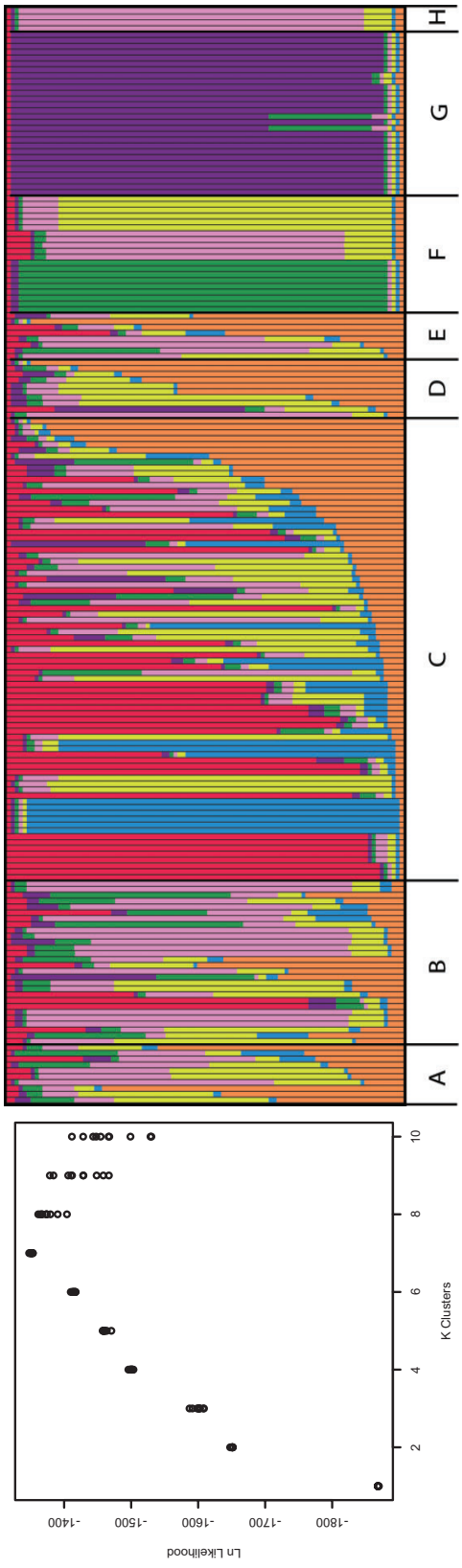

Figure 5. STRUCTURE, version 2.3.3 results from ten replicate runs of one million post-burn-in generations with the microsatellite dataset. Left: likelihood plot, indicating that $K = 7$ is the most probable estimate for *Sphagnum cribrosum*. Right: *S. cribrosum* samples with $K = 7$, organized by geographical location; each sample is a vertical bar, and its admixture in the seven clusters is indicated by colours. Letters indicate samples from: Florida (A); other Georgia Populations (B); GA32 (Ludowici, GA); South Carolina (D); other North Carolina Populations (E); Singletary Lake Normal Form (F); Singletary Lake Waveform (G); Jones Lake Waveform (H). STRUCTURE replicates summarized using CLUMPP and visualized using DISTRUCT.

Singletary Lake. (2) The two distinct growth forms have been recorded at Singletary Lake for at least 70 years, and they grow intimately mixed within centimeters of one another. (3) Haplotype networks and reduced consensus approaches with sequence data show that Singletary Lake waveform and Jones Lakes waveform are separated by many DNA sequence substitutions, and monophyly is unlikely in the extreme. (4) Microsatellite variation across the range of *S. cribrosum* shows the two waveform populations have different genetic backgrounds. Although no single observation 'proves' the two waveform populations are non-monophyletic, every observation and method of data analysis supports this interpretation.

Maintenance of morphological differences during the 8-month experiment and long-term coexistence of the two growth forms at Singletary Lake, along with strong differentiation in nucleotide sequences and microsatellite repeat profiles, suggests that the waveform morphology has a genetic component. Multiple types of genetic analysis, including phylogenetic reconstructions, haplotype networks, and clustering methods, fail to group the two waveform populations from Jones and Singletary Lakes together. These results strongly suggest that the two groups of individuals with the waveform morphology had independent origins despite sharing an indistinguishable, highly aberrant branching pattern, and despite their close geographical proximity (approximately 15 km apart). The waveform morphology is clearly present in two different genetic backgrounds (Figs 5, 6), which are unlikely to have undergone recombination as a result of the rarity of sexual reproduction in most populations of *S. cribrosum*. The indication that the waveform must have independent origins is remarkable because the only two places where the aberrant morphology occurs are only a few kilometers apart, considering the range of *S. cribrosum* from New Jersey to the Gulf coast. Our genetic analyses of *S. cribrosum* throughout its geographical range in the eastern USA also revealed that one population, referred to here as GA32, is exceptionally diverse and appears to be critical to understanding the phylogeographical history of the species, including the waveform.

Microsatellite variation supports a relationship of the Jones Lake waveform to *S. cribrosum* samples with normal morphology from Georgia and Florida, rather than to waveform plants at Singletary Lake. At Singletary Lake, sympatric waveform and normal plants are strongly differentiated genetically. Both microsatellite clustering techniques place the two morphologies at Singletary Lake in separate clusters. In the sequence dataset, nine DNA base substitutions and a 25-nucleotide duplication in RapdF distinguish the normal form from waveform at Singletary Lake. These genetic patterns make it highly unlikely that

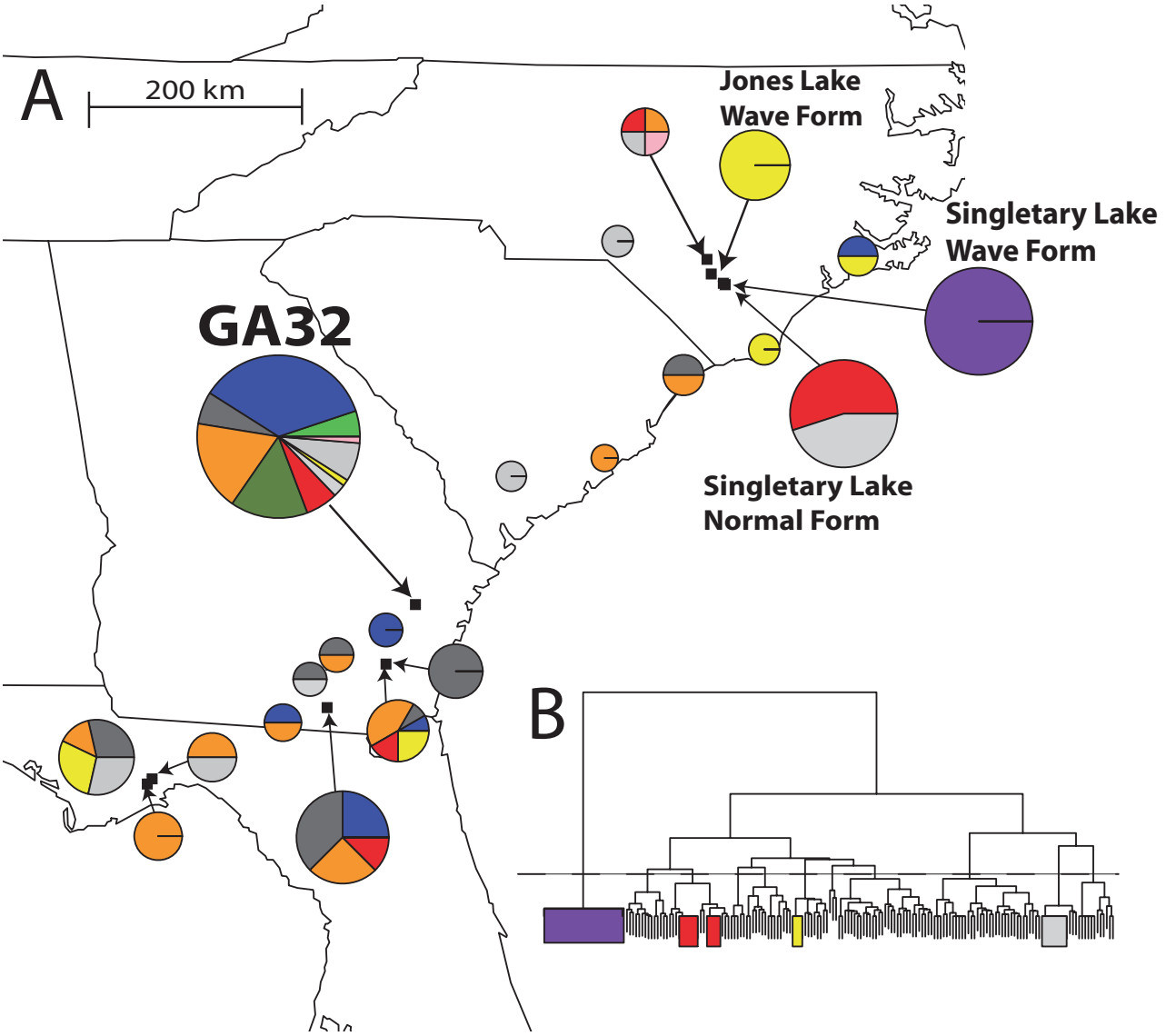

Figure 6. Geographical presentation of nonparametric cluster analysis of the microsatellite dataset, using the R module AWclust, shown for $K = 11$. A, size of the pie charts corresponds to the sample size for each population. Colours indicate the 11 clusters, and shared colours among populations indicate unique haploid multilocus genotypes with broad geographical ranges. B, unrooted unweighted pair group method with arithmetic mean tree of multilocus microsatellite distances, created by the R nonparametric clustering module AWCLUST. The horizontal line indicates a value of $K = 11$. Samples from Singletary Lake (normal form: red, grey; waveform: purple) and Jones Lake (yellow) are indicated by boxes. Moving the horizontal line up or down the tree (reflecting fewer or more clusters) would not unite the waveform samples unless $K = 1$.

the waveform arose from normal form at Singletary Lake. A more likely possibility is that the genetic or epigenetic changes responsible for the waveform morphology have a simple genetic basis and have evolved in multiple genetic backgrounds within *S. cribrosum*.

There is some evidence that normal form plants are less adapted for lake margin environments, although additional work is needed to confirm this observation. Incorporation of Jones Lake waveform at Singletary Lake, larger sample sizes, and direct comparisons of growth between the forms are necessary to further investigate the possibility of local adaptation in an experiment of longer duration. Because the adaptation may be to wave action itself rather than any chemical or passive environmental condition, creative transplanting methods would need to be developed to prevent the mortality seen in our experiment.

Multiple origins of the waveform phenotype do not preclude the possibility that genetically heterogeneous plants with this morphology could go on to function as a single, biologically meaningful species. The ecology of the Singletary Lake population suggests a pathway to adaptive differentiation if waveform plants from Jones and Singletary Lake interbreed. Most *S. cribrosum* populations occur in ditches, outlet streams, and *Taxodium* swamps without the wave action that lent the waveform its nickname. This includes Singletary Lake, where normal form plants grow in the outlet stream. Closer to the lake, both types are observed growing side-by-side, giving way to waveform at the lake margin. Only the waveform occurs throughout the open lake, and the common garden results suggest that normal form plants may be maladaptive at the lake margin. This indicates that a distinct ecology may exist for the waveform, in which it can potentially outcompete normal form plants. However, the lack of evidence for sexual reproduction in *S. cribrosum* in North Carolina, despite over 60 years of documentation by herbarium specimens, the lack of genetic differentiation at each waveform site, and the lack of admixture between Singletary Lake waveform and other clusters of *S. cribrosum* (Fig. 5), suggests that little or no sexual reproduction occurs at these sites. Without admixture between the two lakes, it is unlikely the waveform will ever form one, cohesive, evolutionary lineage.

Importance of the Ludowici, GA, population

The phylogenetic reconstruction and haplotype network indicate that the genetic diversity and allelic composition of plants from Ludowici, GA (GA32) are exceptional. There is nothing especially remarkable about the site: plants are found in a disturbed roadside ditch with less than 1 m of standing water. This type of habitat is found throughout all four states sampled for the present study. However, of the 27 unique haploid multilocus genotypes resolved by TCS, ten can be found at GA32. A large percentage (35%) of normal form plants from populations in South Carolina, Florida, North Carolina, and elsewhere in Georgia had exact multilocus haplotypes in common with samples from GA32. From the microsatellite dataset, genetic admixture (STRUCTURE results; Fig. 5) is higher among plants from GA32 than in any other population. Nonparametric clustering (AWCLUST results; Fig. 6) shows two unique haplotype clusters found only at GA32, and ten other clusters at GA32 are shared with at least one other population (Fig. 5A). Standard population genetic statistics show high diversity at GA32, which is not an artefact of the relatively large sample size from this population.

The genetic diversity of GA32 suggests that it could function as a sink, collecting many genotypes from across the range. However, sexual reproduction and therefore long-range dispersal through the production of spores in *S. cribrosum*, is rare. We have visited every known site for *S. cribrosum* (and discovered many previously unknown sites) and have examined every specimen in DUKE, and are able to state confidently that sporophytes have only been collected at two sites, both in Georgia. A few sporophytes were found once in Echols County, Georgia, in 2005 (GA30 in Table 1), although no sporophytes were observed there during visits in subsequent years. Importantly, at the GA32 population near Ludowici, Georgia (Long County), we observed abundant sporophytes in 2005, 2007, and 2009 (specimens vouchered in DUKE).

The occurrence of GA32 multilocus genotypes at distant sites throughout the south-east USA and abundant sporophyte production at GA32 suggest that this site is not a sink but a source of genetic diversity for *S. cribrosum*. Reproduction in other populations appears to be largely asexual, presumably through stem fragmentation. Such vegetative fragments are substantially larger than spores. Sundberg (2005) showed that *Sphagnum* colonization of islands in the Baltic were more often accomplished by species with frequent sporophytes, indicating that spore dispersal is important for the establishment of new populations, whereas asexual propagation is important for their localized spread. This is in agreement with experimental evidence suggesting that spore-producing mosses are critical for the recovery of disturbed peatlands (Campbell, Rochefort & Lavoie, 2003).

The GA32 population is additionally important because it is one of only two sites where *S. cribrosum* and its sister-species *S. macrophyllum* occur in intimate sympatry. It is the only site where the two species co-occur and produce sporophytes. All plants were sterile in the other sympatric population, located in the Francis Marion National Forest in South Carolina. Preliminary investigations have not revealed evidence of either hybrid sporophytes or recombinant gametophytes at GA32 (M. G. Johnson, P. Zhou & A. J. Shaw, unpubl. data).

Systematic implications

Sphagnum cribrosum and its sister species *S. macrophyllum* are both part of the monophyletic subgenus *Subsecunda* (Shaw, Cox & Boles, 2004). Growth form variation is notably high in the subg. *Subsecunda*. Phylogenetic analyses place *S. cribrosum* and *S. macrophyllum* in a clade within the subgenus *Subsecunda* with two other species exhibiting atypical morphologies: *Sphagnum pylaesii* Brid. and *Sphagnum cyclophyllum* Sull & Lesq. *Sphagnum pylaesii* has few or

no branch fascicles and inconspicuous capitula at best; *S. cyclophyllum* lacks fasciculate branching completely but is largely unbranched (unlike the waveform), nor does it have any hint of a capitulum. Elsewhere in subgenus *Subsecunda*, several species including *Sphagnum subsecundum* Nees and *Sphagnum lescurii* Sull. sometimes grow as unbranching, simplex forms, although these appear to be nongenetic modifications that can be observed side-by-side with normal forms (Crum, 1992; McQueen & Andrus, 2007; Anderson *et al*., 2009). An unbranched form of *Sphagnum denticulatum* Brid. (= *Sphagnum auriculatum* Schimp.) was found to be genetically indistinguishable from the normal branched form but, nevertheless, was described as a new variety, *S. denticulatum* var. *monocladum* J-P. Frahm and Sabovl (Frahm & Sabovljevic, 2006). Unpublished data (M. Ricca and A. J. Shaw) indicate that field-collected plants with normal (capitulate) morphology can become simplex under greenhouse conditions. This within-species phenotypic variation and modifications of the typical *Sphagnum* bauplan among species in the subgenus *Subsecunda* suggest the possibility that branching pattern in *Sphagnum* is controlled by transcription factors rather than exon substitutions.

In flowering plants, morphological divergence typically reserved for species-level distinctions can sometimes be found within species. A white/yellow form of the typically purple *Trillium cuneatum* Rafinesque occurs in the southern Appalachian Mountains and, throughout the past 100 years, it has been recognized as a separate species or subspecies, or synonimized with the more common purple-flowered form (Case, 1997). Hopkins & Rausher (2011) and Zufall & Rausher (2004) have genetically characterized flower colour variants in *Phlox drummondii* Hook. and *Ipomea quamoclit* L., respectively. They demonstrated a simple genetic basis for the variants and were able to identify the underlying biochemical pathways and mutations underlying them. In those species, independently arising flower colour mutations can potentially be recombined into a common genetic background because the species reproduce sexually.

Growth form variants in *S. cribrosum* could reflect modifications in the MADS-box family of transcription factors that influence growth in angiosperms and are known to have homologues in *S. subsecundum* (Zobel *et al*., 2010), which is in the same subgenus as *S. cribrosum*. Alternatively, branching pattern could be under the control of epialleles, which, in flowering plants, have been shown to he heritable and affect both floral structure and responses to external environments (Kalisz & Purugganan, 2004; Bossdorf, Richards & Pigliucci, 2008).

Similar cases of highly divergent infraspecific morphotypes are known in other bryophytes. The island endemic *Leptodon corsicus* has morphological traits associated with other moss genera, *Homalia* Brid. and *Neckera*, Hedw., yet the endemic taxon is phylogenetically embedded within the continental *L. smithii* (Hedw.) F. Weber & D. Mohr (Sotiaux *et al*., 2009). It is argued that there has been insufficient time for slowly-evolving DNA sequence characters to catch up to the morphological differentiation observed in *L. corsicus*, which is hypothesized to be recent and of simple genetic basis (Sotiaux *et al*., 2009). The aberrant *Thamnobryum angustifolium* shares sequence identity with a common sympatric species (*Thamnobryum alopercurum* (Hedw.) Nieuwl. ex Gangulee), and its distinct morphology is convergent with two other recognized rheophyte species in the genus (Olsson *et al*., 2009). Despite this, *T. angustifolium* remains a taxonomically accepted and is considered critically endangered (IUCN Red List, 2011; Olsson *et al*., 2011). The predominant reason for describing the morphologically distinct *P. mutatum* as a new species, despite sequence identity with *P. riparioides* (Hedw.) Dixon (Stech & Frahm, 1999), was the presence of fully developed sporophytes, which indicate that sexual reproduction might maintain the lineage.

None of these infraspecific variants were tested in a common garden or reciprocal transplant to demonstrate the genetic basis of the variant, as described in the present study. Whatever the underlying mechanism, it is clear that the waveform morphology is not transient, as is the simplex forms of other species in *Sphagnum* subg. *Subsecunda*. However, the lack of sexual reproduction renders the waveform unlikely to form a single evolutionary lineage worthy of taxonomic rank. Fortunately, because the form has considerable biomass at both lakes, in North Carolina State Parks, it is likely this unique form will remain preserved.

ACKNOWLEDGEMENTS

The authors thank S. Boles for support generating the sequences and genotyping, as well as M. Ricca, N. Devos, H. Stenøien, and four anonymous reviewers for their comments on earlier drafts. Singletary Lake State Park also graciously provided access for sampling and setting up common garden experiments. This work was supported by NSF grant DEB-0710178 to A. J. Shaw.

REFERENCES

Agapow P, Burt A. 2001. Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* **1:** 101–102.

Anderson LE, Shaw AJ, Shaw B. 2009. *Peat mosses of the Southeastern United States*, Buck W, ed., Vol. 102. New York, NY: Memoirs of the New York Botanical Garden Press.

Becker A, Theissen G. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular Phylogenetics and Evolution* **29:** 464–489.

Becker RA, Wilks AR. 1993. Maps in S. *AT&T Bell Laboratories Statistics Research Report*, 93.

Bossdorf O, Richards CL, Pigliucci M. 2008. Epigenetics for ecologists. *Ecology Letters* **11:** 106–115.

Campbell D, Rochefort L, Lavoie C. 2003. Determining the immigration potential of plants colonizing disturbed environments: the case of milled peatlands in Quebec. *Journal of Applied Ecology* **40:** 78–91.

Case FW Jr. 1997. Trillium. In: Flora of North America Editorial Committee, ed. 1993+ *Flora of North America north of Mexico*, Vol. 26. 12+ vols. New York, NY.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9:** 1657–1659.

Crum HA. 1992. Miscellaneous notes on the genus *Sphagnum*. *The Bryologist* **93:** 274–279.

Crum HA, Anderson LE. 1981. *Mosses of Eastern North America*. New York, NY: Columbia University Press.

Cubas P, Vincent C, Coen E. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401:** 157–161.

Frahm J-P, Sabovljevic M. 2006. *Sphagnum denticulatum* var. *monocladum*, a new taxon from the Azores, with notes on *Sphagnum monocladum* Warnst. *Tropical Bryology* **27:** 113–119.

Gao X, Starmer JD. 2008. AWclust: point-and-click software for non-parametric population structure analysis. *BMC Bioinformatics* **9:** 77.

Hintz M, Bartholomes C, Nutt P, Ziermann J, Hameister S, Neuffer B, Theissen G. 2006. Catching a 'hopeful monster': shepherd's purse (*Capsella bursa-pastoris*) as a model system to study the evolution of flower development. *Journal of Experimental Botany* **57:** 3531–3542.

Hopkins R, Rausher MD. 2011. Identification of two genes causing reinforcement in the Texas wildflower *Phlox drummondii*. *Nature* **469:** 411–414.

IUCN. 2011. *IUCN Red List of Threatened Species*. Version 2011.2. http://www.iucnredlist.org. Accessed on 10 August 2011.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *BMC Bioinformatics* **23:** 1801–1806.

Johnson MG, Shaw B, Zhou P, Shaw AJ. 2012. Data from: Genetic analysis of the peatmoss *Sphagnum cribrosum* (Sphagnaceae) indicates independent origins of an extreme infra-specific morphology shift. *Dryad digital repository*. doi: 10.5061/dryad.860jq72n.

Kalisz S, Purugganan MD. 2004. Epialleles via DNA methylation: consequences for plant evolution. *Trends in Ecology & Evolution* **19:** 309–314.

McQueen CB, Andrus RE. 2007. Sphagnaceae. In: Flora of North America Editorial Committee, eds. 1993+ *Flora of North America north of Mexico*. New York, NY.

Muller J, Muller KF, Neinhuis C, Quandt D. 2007. PhyDE – phylogenetic data editor. http://www.phyde.de

Olsson S, Enroth J, Buchbender V, Hedenas L, Huttunen S, Quandt D. 2011. *Neckera* and *Thamnobryum* (Neckeraceae, Bryopsida): paraphyletic assemblages. *Taxon* **60:** 36–50.

Olsson S, Rumsey F, Grundmann M, Russell S, Enroth J, Quandt D. 2009. The origin of the British and Macaronesian endemic *Thamnobryum* species (Neckeraceae). *Journal of Bryology* **31:** 1–10.

Peakall R, Smouse P. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6:** 288–295.

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

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

R Development Core Team. 2011. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rosenberg N. 2004. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4:** 137–138.

Shamel A, Pomeroy C. 1936. Bud mutations in horticultural crops. *Journal of Heredity* **27:** 487–494.

Shaw AJ, Cao T, Wang L, Flatberg KI, Flatberg B. 2008. Genetic variation in three Chinese peat mosses (*Sphagnum*) based on microsatellite markers, with primer information and analysis of ascertainment bias. *The Bryologist* **111:** 271–281.

Shaw AJ, Cox C, Boles SB. 2004. Phylogenetic relationships among *Sphagnum* sections: hemitheca, isocladus, and subsecunda. *The Bryologist* **107:** 189–196.

Shaw AJ, Cox CJ, Boles SB. 2003. Polarity of peatmoss (*Sphagnum*) evolution: who says bryophytes have no roots? *American Journal of Botany* **90:** 1777–1787.

Sotiaux A, Enroth J, Olsson S, Quandt D, Vanderpoorton A. 2009. When morphology and molecules tell us different stories: a case-in-point with *Leptodon corsicus*, a new and unique endemic moss from Corsica. *Journal of Bryology* **31:** 186–196.

Stech M, Frahm JP. 1999. The status and systematic position of *Platyhypnidium mutatum* and the Donrichardaceae. *Journal of Bryology* **21:** 189–193.

Sundberg S. 2005. Larger capsules enhance short-range spore dispersal in *Sphagnum*, but what happens further away? *Oikos* **108:** 115–124.

Swofford DL. 2003. *PAUP*. Phylogenetic analysis using parsimony (*and other methods)*, Version 4.0a109. Sunderland, MA: Sinauer Associates.

Theissen G. 2009. Saltational evolution: hopeful monsters are here to stay. *Theory in Biosciences* **128:** 43–51.

Wilkinson M. 1996. Majority-rule reduced consensus trees and their use in bootstrapping. *Molecular Biology and Evolution* **13:** 437–444.

Zhou P, Menzel F, Shaw AJ. 2007. Systematics and population genetics of *Sphagnum macrophyllum* and

S. cribrosum (Sphagnaceae). *Systematic Botany* **32:** 493–503.

Zobel O, Faigl W, Saedler H, Münster T. 2010. MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants. *Molecular Biology and Evolution* **27:** 1201–1211.

Zufall RA, Rausher MD. 2004. Genetic changes associated with floral adaptation restrict future evolutionary potential. *Nature* **428:** 847–850.

Zwickl DJ. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD Dissertation, University of Texas at Austin.

APPENDIX

Sample voucher information: ID; sample type (normal form or waveform of *Sphagnum cribrosum* or *Sphagnum macrophyllum*); population; state; county; collector; collection number; GenBank IDs for RapdA, RapdB, and RapdF. All vouchered specimens are stored in DUKE herbarium.

15_1_cribrosum_GA32, normal form, GA32, USA: Georgia, Long Co., Zhou *et al*., PZ15_1, JQ028900, JQ028960, JQ029020;**15_3_cribrosum_GA32,** normal form, GA32, USA: Georgia, Long Co., Zhou *et al*., PZ15_3, JQ028901, JQ028961, JQ029021;**42_2_cribrosum_GA32,** normal form, GA32, USA: Georgia, Long Co., Zhou *et al*., PZ42_2, JQ028902, JQ028962, JQ029022;**47_1_cribrosum_GA32,** normal form, GA32, USA: Georgia, Long Co., Zhou *et al*., PZ47_1, JQ028903, JQ028963, JQ029023; **47_2_macrophyllum_GA32,** macrophyllum, GA32, USA: Georgia, Long Co., Zhou *et al*., PZ47_2, JQ028876, JQ028936, JQ028996;**58_3_macrophyllum_GA32,** macrophyllum, GA32, USA: Georgia, Long Co., Zhou *et al*., PZ58_3, JQ028877, JQ028937, JQ028997;**59_1_cribrosum_GA32,** normal form, GA32, USA: Georgia, Long Co., Zhou *et al*., PZ59_1, JQ028915, JQ028975, JQ029035;**JL1_cribrosumW_JonesLake,** waveform, NCJones, USA: North Carolina, Bladen Co., Ping Zhou, W1, JQ028934, JQ028994, JQ029054;**MJ1_cribrosum_NCHorse,** normal form, NCHorse, USA: North Carolina, Bladen Co., M. Johnson, MJ1, JQ028923, JQ028983, JQ029043; **MJ2_cribrosum_NCHorse,** normal form, NCHorse, USA: North Carolina, Bladen Co., M. Johnson, MJ2, JQ028924, JQ028984, JQ029044;**MJ4_cribrosum_NCHorse,** normal form, NCHorse, USA: North Carolina, Bladen Co., M. Johnson, MJ4, JQ028925, JQ028985, JQ029045;**MJ9_cribrosum_NCHorse,** normal form, NCHorse, USA: North Carolina, Bladen Co., M. Johnson, MJ9, JQ028913, JQ028973, JQ029033;**NC11_cribrosum_SingLake,** normal form, NCSingletary, USA: North Carolina, Bladen Co., Ping Zhou, NC11, EF158596, EF158634, EF158678;**NC16_cribrosum_SingLake,** normal form, NCSingletary, USA: North Carolina, Bladen Co., Ping Zhou, NC16, EF158598, EF158636, EF158680;**NC19_cribrosum_SingLake,** normal form, NCSingletary, USA: North Carolina, Bladen Co., Ping Zhou, NC19, EF158600, EF158638, EF158682;**NC7_cribrosumW_SingLake,** waveform, NCSingletary, USA: North Carolina, Bladen Co., Ping Zhou, NC7, EF158611, EF158651, EF158696;**Z1000_cribrosum_SC42,** normal form, SC42, USA: South Carolina, Horry Co., Ping Zhou, Z1000, JQ028885, JQ028945, JQ029005;**Z102_cribrosum_NCSingLake,** normal form, NCSingletary, USA: North Carolina, Bladen Co., Ping Zhou, Z102, JQ028922, JQ028982, JQ029042; **Z104_cribrosum_NCSingLake,** normal form, NCSingletary, USA: North Carolina, Bladen Co., Ping Zhou, Z104, JQ028932, JQ028992, JQ029052; **Z116_cribrosumW_NCSingLake,** waveform, NCSingletary, USA: North Carolina, Bladen Co., Ping Zhou, Z116, JQ028935, JQ028995, JQ029055;**Z372_cribrosum_FL8,** normal form, FL8, USA: Florida, Liberty Co., Ping Zhou, Z372, JQ028886, JQ028946, JQ029006;**Z390_cribrosum_FL8,** normal form, FL8, USA: Florida, Liberty Co., Ping Zhou, Z390, JQ028887, JQ028947, JQ029007; **Z531_cribrosum_FL20,** normal form, FL20, USA: Florida, Wakulla Co., Ping Zhou, Z531, JQ028914, JQ028974, JQ029034;**Z532_cribrosum_FL20,** normal form, FL20, USA: Florida, Wakulla Co., Ping Zhou, Z532, JQ028888, JQ028948, JQ029008;**Z541_cribrosum_FL21,** normal form, FL21, USA: Florida, Wakulla Co., Ping Zhou, Z541, JQ028889, JQ028949, JQ029009;**Z55_cribrosum_FLFL,** normal form, NCNC, USA: FL, Polk Co., Doug Goldman, 2762, JQ028933, JQ028993, JQ029053;**Z556_cribrosum_FL21,** normal form, FL21, USA: Florida, Wakulla Co., Ping Zhou, Z556, JQ028926, JQ028986, JQ029046; **Z56_cribrosum_NCNC,** normal form, NCNC, USA: NC, Scotland Co., J.Shaw, 9944a, JQ028912, JQ028972, JQ029032;**Z57_cribrosum_NCNC,** normal form, NCNC, USA: NC, Carteret Co., Allen Risk, 8152, JQ028927, JQ028987, JQ029047;**Z579_cribrosum_GA22,** normal form, GA22, USA: Georgia, Echols Co., Ping Zhou, Z579, JQ028890, JQ028950, JQ029010;**Z58_cribrosum_NCNC,** normal form, NCNC, USA: NC, Carteret Co., Allen Risk, 8153, JQ028921, JQ028981, JQ029041; **Z582_cribrosum_GA22,** normal form, GA22, USA: Georgia, Echols Co., Ping Zhou, Z582, JQ028907, JQ028967, JQ029027;**Z60_cribrosum_NCNC,** normal form, NCNC, USA: NC, Bladen Co., L. Anderson, 25718, JQ028891, JQ028951, JQ029011; **Z62_cribrosum_NCNC,** normal form, NCNC, USA: NC, Bladen Co., M. Johnson, 10, JQ028928,

JQ028988, JQ029048;**Z631_cribrosum_GA25,** normal form, GA25, USA: Georgia, Clinch Co., Ping Zhou, Z631, JQ028892, JQ028952, JQ029012; **Z651_cribrosum_GA25,** normal form, GA25, USA: Georgia, Clinch Co., Ping Zhou, Z651, JQ028893, JQ028953, JQ029013;**Z653_cribrosum_GA26,** normal form, GA26, USA: Georgia, Clinch Co., Ping Zhou, Z653, JQ028894, JQ028954, JQ029014; **Z664_cribrosum_GA26,** normal form, GA26, USA: Georgia, Clinch Co., Ping Zhou, Z664, JQ028908, JQ028968, JQ029028;**Z688_cribrosum_GA27,** normal form, GA27, USA: Georgia, Ware Co., Ping Zhou, Z688, JQ028916, JQ028976, JQ029036;**Z698_cribrosum_GA27,** normal form, GA27, USA: Georgia, Ware Co., Ping Zhou, Z698, JQ028904, JQ028964, JQ029024;**Z743_cribrosum_GA30,** normal form, GA30, USA: Georgia, Brantley Co., Ping Zhou, Z743, JQ028895, JQ028955, JQ029015;**Z782_cribrosum_GA30,** normal form, GA30, USA: Georgia, Brantley Co., Ping Zhou, Z782, JQ028896, JQ028956, JQ029016;**Z784_cribrosum_GA31,** normal form, GA31, USA: Georgia, Brantley Co., Ping Zhou, Z784, JQ028897, JQ028957, JQ029017;**Z797_macrophyllum_GA32,** macrophyllum, GA32, USA: Georgia, Long Co., Ping Zhou, Z797, JQ028878, JQ028938, JQ028998;**Z800_macrophyllum_GA32,** macrophyllum, GA32, USA: Georgia, Long Co., Ping Zhou, Z800, JQ028879, JQ028939, JQ028999;**Z802_cribrosum_GA32,** normal form, GA32, USA: Georgia, Long Co., Ping Zhou, Z802, JQ028917, JQ028977, JQ029037;**Z803_macrophyllum_GA32,** macrophyllum, GA32, USA: Georgia, Long Co., Ping Zhou, Z803, JQ028880, JQ028940, JQ029000;**Z819_macrophyllum_GA32,** macrophyllum, GA32, USA: Georgia, Long Co., Ping Zhou, Z819, JQ028881, JQ028941, JQ029001;**Z820_macrophyllum_GA32,** macrophyllum, GA32, USA: Georgia, Long Co., Ping Zhou, Z820, JQ028882, JQ028942, JQ029002;**Z822_cribrosum_GA32,** normal form, GA32, USA: Georgia, Long Co., Ping Zhou, Z822, JQ028905, JQ028965, JQ029025;**Z826_cribrosum_GA32,** normal form, GA32, USA: Georgia, Long Co., Ping Zhou, Z826, JQ028906, JQ028966, JQ029026;**Z827_cribrosum_GA32,** normal form, GA32, USA: Georgia, Long Co., Ping Zhou, Z827, JQ028930, JQ028990, JQ029050;**Z828_cribrosum_GA32,** normal form, GA32, USA: Georgia, Long Co., Ping Zhou, Z828, JQ028931, JQ028991, JQ029051;**Z829_cribrosum_GA32,** normal form, GA32, USA: Georgia, Long Co., Ping Zhou, Z829, JQ028918, JQ028978, JQ029038; **Z830c3_cribrosum_GA32,** normal form, GA32, USA: Georgia, Long Co., Ping Zhou, Z830c3, JQ028910, JQ028970, JQ029030;**Z830m1_macrophyllum_GA32,** macrophyllum, GA32, USA: Georgia, Long Co., Ping Zhou, Z830m1, JQ028883, JQ028943, JQ029003

Z831_cribrosum_GA32, normal form, GA32, USA: Georgia, Long Co., Ping Zhou, Z831, JQ028911, JQ028971, JQ029031

Z831m_macrophyllum_GA32, macrophyllum, GA32, USA: Georgia, Long Co., Ping Zhou, Z831m, JQ028884, JQ028944, JQ029004

Z834_cribrosum_GA32, normal form, GA32, USA: Georgia, Long Co., Ping Zhou, Z834, JQ028919, JQ028979, JQ029039

Z844_cribrosum_GA33, normal form, GA33, USA: Georgia, Wayne Co., Ping Zhou, Z844, JQ028898, JQ028958, JQ029018

Z915_cribrosum_SC37, normal form, SC37, USA: South Carolina, Colleton Co., Ping Zhou, Z915, JQ028909, JQ028969, JQ029029

Z919_cribrosum_SC37, normal form, SC37, USA: South Carolina, Colleton Co., Ping Zhou, Z919, JQ028929, JQ028989, JQ029049

Z963_cribrosum_SC40, normal form, SC40, USA: South Carolina, Charleston Co., Ping Zhou, Z963, JQ028920, JQ028980, JQ029040

Z988_cribrosum_SC42, normal form, SC42, USA: South Carolina, Horry Co., Ping Zhou, Z988, JQ028899, JQ028959, JQ029019