RESEARCH ARTICLE

AMERICAN JOURNAL OF BOTANY

INVITED PAPER

For the Special Issue: Patterns and Processes of American Amphitropical Plant Disjunctions: New Insights

Resolving the northern hemisphere source region for the long-distance dispersal event that gave rise to the South American endemic dung moss *Tetraplodon fuegianus*[1]

Lily R. Lewis[2,3,8], Elisabeth M. Biersma[4,5], Sarah B. Carey[3], Kent Holsinger[2], Stuart F. McDaniel[3], Ricardo Rozzi[6,7], and Bernard Goffinet[2]

PREMISE OF THE STUDY: American bipolar plant distributions characterize taxa at various taxonomic ranks but are most common in the bryophytes at infra-specific and infrageneric levels. A previous study on the bipolar disjunction in the dung moss genus *Tetraplodon* found that direct long-distance dispersal from North to South in the Miocene–Pleistocene accounted for the origin of the Southern American endemic *Tetraplodon fuegianus*, congruent with other molecular studies on bipolar bryophytes. The previous study, however, remained inconclusive regarding a specific northern hemisphere source region for the transequatorial dispersal event that gave rise to *T. fuegianus*.

METHODS: To estimate spatial genetic structure and phylogeographic relationships within the bipolar lineage of *Tetraplodon*, which includes *T. fuegianus*, we analyzed thousands of restriction-site-associated DNA (RADseq) loci and single nucleotide polymorphisms using Bayesian individual assignment and maximum likelihood and coalescent model based phylogenetic approaches.

KEY RESULTS: Northwestern North America is the most likely source of the recent ancestor to *T. fuegianus*.

CONCLUSIONS: *Tetraplodon fuegianus*, which marks the southernmost populations in the bipolar lineage of *Tetraplodon*, arose following a single long-distance dispersal event involving a *T. mnioides* lineage that is now rare in the northern hemisphere and potentially restricted to the Pacific Northwest of North America. Furthermore, gene flow between sympatric lineages of *Tetraplodon mnioides* in the northern hemisphere is limited, possibly due to high rates of selfing or reproductive isolation.

KEY WORDS amphitropical; Bryopsida; RADseq; Splachnaceae

The geographic ranges of organisms across the tree of life exhibit bipolar disjunctions, spanning at least part of the polar and temperate regions, and potentially with populations in tropical montane localities (Du Rietz, 1940). This pattern occurs in bacteria (Pearce et al., 2007; Sul et al., 2013), animals (Crame, 1993), flowering plants (Popp et al., 2011; Villaverde et al., 2015), liverworts (Kreier et al., 2010), lichens (Fernández-Mendoza and Printzen, 2013), and mosses (Lewis et al., 2014a; Biersma et al., 2017), with the latter two providing the greatest number of infraspecific and infrageneric examples. At least 67 species of mosses (Ochyra, 1992; Ochyra and Buck, 2003; Ochyra et al., 2008; Norris et al., 1999; Schofield, 1974), 26 liverwort species (Schuster, 1983; Streimann, 1998; Norris et al., 1999; Bednarek-Ochyra et al., 2000) and at least 160 species of lichens (Øvstedal and Lewis Smith, 2001) exhibit bipolar distributions.

Three primary processes may result in bipolar disjunctions, namely, vicariance (including tectonic, climatic, or biotic) (Darwin, 1859), stepping-stone migration via high-montane tropical populations (Du Rietz, 1940), and direct long-distance dispersal (LDD) across the tropics (via biotic or abiotic vectors) (Raven, 1963; Wen and Ickert-Bond, 2009). In all studies to date addressing bipolarity

[1] Manuscript received 14 April 2017; revision accepted 4 October 2017.
[2] Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269 USA;
[3] Department of Biology, University of Florida, Gainesville, Florida 32601 USA;
[4] Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK;
[5] British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge CB3 0ET, UK;
[6] Institute of Ecology and Biodiversity, Santiago, Chile; Omora Ethnobotanical Park, Puerto Williams, Antarctic Province, Chile; and
[7] Universidad de Magallanes, Punta Arenas, Magallanes Province, Chile; Department of Philosophy, University of North Texas, Denton, Texas 76201 USA
[8] Author for correspondence (e-mail: LilyRLewis@gmail.com); present address: Department of Biology, University of Florida, Gainesville, FL 32601 USA; ORCID id 0000-0001-6529-9785
https://doi.org/10.3732/ajb.1700144

in bryophytes (Hedenäs, 2009, 2012; Kreier et al., 2010; Piñeiro et al., 2012; Lewis et al., 2014a; Kyrkjeeide et al., 2016b; Biersma et al., 2017), results are incongruent with vicariance and stepping-stone scenarios, suggesting that recent (in the case of phylograms), or specifically Miocene–Pleistocene (based on chronograms) LDD best accounts for bipolar disjunctions between species or populations. A previous molecular study aimed at addressing the topic of bipolar bryophytes concerned the origin of *Tetraplodon fuegianus* Bruch and Schimp., a dung moss endemic to southern South America (Lewis et al., 2014a). Inferences based on four discrete loci resolved *T. fuegianus* as being nested within a bipolar lineage of the widely distributed *T. mnioides* (Hedw.) Bruch & Schimp. complex, with a broad northern hemisphere distribution and additional disjunctions into the highlands of Papua New Guinea, Borneo, and southeastern Brazil. The study suggested that *T. fuegianus* arose following a direct LDD event from the north ~8.63 Ma (95% highest posterior density [HPD] 3.07–10.11 Ma); however, the precise geographic origin of the ancestor of *T. fuegianus* remained unclear. A sample of the Norwegian endemic *T. blyttii* Frisvoll was resolved as the sister group of *T. fuegianus*, but this relationship was weakly supported, and an eastern or western North America, or a Himalayan origin of the ancestor of *T. fuegianus* remained plausible (Lewis et al., 2014a).

Here we report on the analyses of thousands of loci and single nucleotide polymorphisms (SNPs), generated with restriction-site-associated DNA sequencing (RADseq; Baird et al., 2008), to reconstruct the relationships within the bipolar *Tetraplodon* lineage that includes *T. fuegianus*. We aimed to identify the northern hemisphere geographic source of the LDD event that gave rise to *T. fuegianus* using maximum likelihood and coalescence-based phylogenetic tree estimation. We then use Bayesian individual assignment analyses among northern temperate populations of the bipolar lineage to evaluate the potential geographic distribution of the sister group of *T. fuegianus*.

MATERIALS AND METHODS

Sampling and molecular identification—We collected 81 samples across the geographic range of the bipolar *Tetraplodon* lineage inferred by Lewis et al. (2014a; Appendix S1, see the Supplemental Data with this article). We cover nearly the complete taxonomic range of the lineage, including the ubiquitous *T. mnioides* ($n = 68$), the endemic *T. lamii* of Papua New Guinea ($n = 2$), and the southern South American endemic *T. fuegianus* ($n = 11$). Specimens of Norwegian endemic *T. blyttii* and the southeastern Brazilian endemic *T. itatiaiae* that were suitable for RADseq library preparation were not found. We sampled previous collection sites of the Norwegian *T. blyttii* (Frisvoll, 1978), but we did not find specimens consistent with the morphological description of the species. We did, however, sample abundant populations of *T. mnioides* in localities cited for *T. blyttii*. We confirmed membership in the bipolar clade through maximum likelihood phylogenetic analysis of the chloroplast gene *rps4* using the program Garli v. 2.0 (Zwickl, 2006) following Lewis et al. (2014a; Appendix S2) to account for known incongruence between current morphological species concepts and molecular phylogenetic lineages. We extracted DNA using the sampling approach discussed by Lewis et al. (2016), pooling gametophyte stems and sporophytes from a discrete patch to provide sufficient DNA yields for next-generation sequencing. Plant material was ground with liquid nitrogen, and DNA was extracted using Nucleospin Plant II Midi kits (Macherey-Nagel, Bethlehem, Pennsylvania, USA) following the manufacturer's guidelines.

RADseq library preparation—We prepared RADseq libraries according to the protocol described by Etter and Johnson (2012) and Etter et al. (2011). Fresh vegetative tissue of *T. fuegianus* was ground to initiate cultures (Shaw, 1986) used for flow cytometry genome size estimates with CyStain PI Absolute P DNA Staining Kit for Plant Genome Size (Sysmex Partec, Görlitz, Germany) following the manufacturer's guidelines. We used the estimated genome size and GC content (GC content based on the model moss *Physcomitrella patens*; Rensing et al., 2008) to optimize the choice of restriction enzyme.

DNA was digested with SbfI and ligated to (P1) barcoded modified Solexa adapters (2006 Illumina, San Diego, California, USA; Etter et al., 2011). We designed 8-bp barcodes with three differences between all barcodes. Barcoded samples were pooled and sheared to an average size of 400 bp using an M220 Focused-ultrasonicator (Covaris, Woburn, Massachusetts, USA) following the manufacturer's guidelines. DNA fragments between 400 and 600 bp were selected using Agencourt AMPure XP beads (Beckman Coulter, Brea, California, USA) at a 0.8:1.0 beads to library ratio and ligated to a second (P2) modified Solexa adapter. DNA fragments with both P1 and P2 adapters were PCR amplified using the primers specified by Etter et al. (2011). Complete P1 barcoded and P2 adapter and PCR oligo sequences are listed in Appendix S3. Libraries were sequenced on an Illumina MiSeq with 600 cycle v.3 chemistry (2015 Illumina).

Bioinformatics processing—We conducted bioinformatic processing in the pipeline PyRAD v. 2.17 (Eaton, 2014). Reads were demultiplexed allowing for a maximum of two mismatches in each 8-bp barcode, and restriction enzyme cut sites and adapter sequences were removed. Reads with base calls having Phred quality scores of <20 were changed to N (undetermined), and reads were discarded if they contained more than a total of 10-bp with a Phred quality score of <20. Cleaned reads were dereplicated with number of replicate read occurrences recorded and clustered within samples at an 88% similarity threshold. Clusters with a depth greater than the mean depth of all within-sample clusters plus two standard deviations of that mean depth or a depth >500 reads were excluded as putatively assembled paralogs. Sequencing error rate estimation was done with expected heterozygosity set to zero (haploid) using the maximum likelihood equation of Lynch (2008) implemented in PyRAD, and consensus sequences were generated within each sample for each cluster based on error rate estimations. We excluded loci if they contained more than five undetermined bases (N) or contained more than one allele, meaning that we retained only loci that were fixed for a single allele within a sample. We clustered consensus sequences across samples at an 88% similarity threshold and aligned the resultant loci to generate the data sets detailed in Tables 1 and 2. Detailed rationale for selection of minimum sequencing depth, clustering similarity threshold, and heterozygosity are included in Appendix S4.

We generated multiple data sets in the final step of the PyRAD pipeline based on absolute minimum number of samples. Loci for which a minimum number of samples had data, i.e., minimum sample (MS) data sets, were produced for minimum sample thresholds of 60, 40, and 20 (Table 1). Minimum sample data sets vary in the number of loci included, as well as in the percentages of missing data (Tables 1 and 2), and allow for the assessment of congruence among data sets.

TABLE 1. Multiple supermatrices were generated using the PyRAD pipeline to assess congruence among data sets varying in number of loci and percentages of missing data. A locus was included in a supermatrix if it had sequence data for a minimum number of samples. Supermatrices were generated for minimum sample thresholds of 60, 40, and 20. Two sample partitions were used, samples from all localities (n = 81) and a subset of samples collected across the northern temperate zone (n = 65), to assess congruence among sample partitions.

Minimum samples / locus	No. loci	Base pairs	No. PI	% PI	Mean PI / locus	% Missing data
All-localities (n = 81)						
60	1407	407,604	8942	2.194	6	18.460
40	3148	912,453	20,319	2.227	6	29.224
20	4077	1,189,336	27,247	2.291	7	37.075
Northern temperate (n = 65)						
60	73	21,089	98	0.465	1	6.430
20	3880	1,126,370	9437	0.838	2	35.470

Note: PI, parsimony informative sites.

We generated MS data sets for a data partition including samples from all localities (n = 81) and a subset of samples collected across the northern temperate zone (n = 65) to assess congruence among sample partitions. Loci were concatenated to form supermatrices, and MS 20 and MS 60 were mined for single nucleotide polymorphisms (SNPs), with one SNP randomly selected from each locus in the associated supermatrix. Preliminary analyses revealed that MS 40 and MS 60 supermatrices supported congruent phylogenetic results, so SNP supermatrices were only produced for MS 60 and MS 20 supermatrices. MS 20 supermatrices (i.e., the most inclusive in terms of loci) for both sample partitions (n = 81 and n = 65) were used in a blastn search (default settings) against a custom database comprising the complete chloroplast (KU095851) and mitochondrial (KT373818) genomes and nuclear ribosomal repeat (KU095852) of *T. fuegianus* (Lewis et al., 2016) to identify cytoplasmically inherited RAD loci.

Phylogenetic analyses—We performed maximum likelihood phylogenetic analyses with the program RAxML v. 8.1.3 (Stamatakis, 2006; Stamatakis et al., 2008) for all MS supermatrices. We used the GTRCAT model approximation (Stamatakis, 2006) to complete single, full maximum likelihood tree searches with 100 bootstrap replicates using the rapid bootstrapping algorithm (Stamatakis et al., 2008).

Misleading phylogenetic results due to incomplete lineage sorting may not be detectable from the estimation of a single optimal topology based on concatenated supermatrices, and thus the program SVDquartets (Chifman and Kubatko, 2014) was employed alongside RAxML. SVDquartets, as implemented in the program PAUP* v. 4.0a142 (Swofford, 2001), estimates species-level phylogenetic relationships by inferring relationships between quartets of samples under the coalescent model directly from multilocus sequence data using an algebraic statistical approach (Chifman and Kubatko, 2014). Lineage trees were inferred for n = 81 (all localities) supermatrices by evaluating all possible quartets with 100 bootstrap replicates under the multispecies coalescent tree model without *a priori* sample partition information. Species trees were inferred for MS 60 and 20 supermatrices (n = 81) evaluating all possible quartets under the multispecies coalescent tree model with 100 bootstrap replicates. Samples were constrained under two different *a priori* sample partitions: (1) based on collection locality and (2) based on RAxML inferred lineages.

TABLE 2. Multiple single nucleotide polymorphism (SNP) supermatrices were generated using the PyRAD pipeline to assess congruence among data sets varying in number of SNPs and percentages of missing data. SNPs were mined from minimum sample 20 and 60 supermatrices. A locus was included in a supermatrix if it had sequence data for a minimum number of samples (either 20 or 60), and then one SNP was randomly selected from each locus. Two sample partitions were used, samples from all localities (n = 81) and a subset of samples collected across the northern temperate zone (n = 65), to assess congruence among sample partitions.

Sample partition	MS / locus	No. SNPs	% Missing data
All localities (n = 81)	60	1397	19.731
	20	4020	37.871
Northern temperate (n = 65)	60	63	10.452
	20	3508	36.947

Notes: MS, minimum sample; SNPs, single nucleotide polymorphisms.

Spatial genetic structure—We used the program STRUCTURE v. 2.3.4 (Pritchard et al., 2000; Falush et al., 2003; Hubisz et al., 2009) for individual assignment analyses for MS 60 and MS 20 SNP data sets for the all-localities sampling (n = 81) and the northern temperate sampling (n = 65). We ran analyses under the admixture model with alpha inferred, with analyses not informed by sample location data, and with allele frequencies treated as independent with lambda set to 1.0. Preliminary trials suggested that the range of reasonable K values (i.e., number of genetic groups) for all data sets was K = 2 through K = 7 for each sample partition. Each data set was run with a burn-in period of 10,000 MCMC reps, followed by 1,000,000 MCMC reps for 10 independent runs at each K value. We compiled the results from each run using the Clumpak server (Cluster Markov Packager Across K; Kopelman et al., 2015), which calls on CLUMPP (Jakobsson and Rosenberg, 2007) and DISTRUCT (Rosenberg, 2004) and calculates optimal K values according to peaks in ΔK values as described by Evanno et al. (2005) and the highest mean Prob(K) value as described by Pritchard et al. (2000).

Pairwise F_{ST} (Weir and Cockerham, 1984) was estimated with sample group membership determined by collection locality (Table 3). Samples collected in Washington, United States and Nunavut, Canada, were excluded based on their divergence from other samples and limited sampling (i.e., n = 1) for these localities. Concatenated fasta alignments were converted to genotype data in R (version 3.3.0, R Foundation for Statistical Computing, Vienna, Austria) using the package *adegent* (version 2.0.1; Jombart, 2008; Jombart and Ahmed, 2011) and Weir and Cockerham's F_{ST} was estimated in the package HIERFSTAT (version 0.04-22; Goudet, 2005) using the command pairwise.WCfst() with diploid = FALSE.

RESULTS

Features of MS supermatrices used in RAxML, SVDquartets, and F_{ST} analyses are listed in Table 1, and SNP matrices analyzed in

TABLE 3. Pairwise F_{ST} (Weir and Cockerham) calculated with sample group membership determined by collection locality for eastern and western North American, South American, and western European samples in the bipolar *Tetraplodon* lineage. Abbreviations and sample sizes for each group are western Europe (W Eur; n = 19), eastern North America (E NAm; n = 14), western North America (W NAm; n = 33), and southern South America (S SAm; n = 11). Samples included in the S SAm group represent the southern South American endemic species *T. fuegianus*. All other samples are morphologically circumscribed as *T. mnioides*.

Locality	W NAm	E NAm	W Eur
W NAm	—		
E NAm	0.1106	—	
W Eur	0.1957	0.1429	—
S SAm	0.4782	0.5683	0.5267

STRUCTURE are described in Table 2. Results of sample barcoding, genome size estimation, sequencing, and bioinformatics processing are presented in Appendix S4.

Maximum likelihood phylogenetic analyses—RAxML analyses of MS 60, 40, and 20 supermatrices for the all-localities sampling (n = 81) yielded the same clades but slightly different topologies (Fig. 1; Appendix S5). MS 60 and 40 topologies differ only in branch lengths and bootstrap support values (Appendix S5). All supermatrices maximally support samples from Papua New Guinea (PNG; n = 2) and Nepal (n = 2) as composing distinct monophyletic lineages. RAxML trees inferred from the MS 60, 40, and 20 supermatrices were rooted with PNG samples based on the previous genus-wide study (Lewis et al., 2014a). Each rooted topology infers samples from Nepal and Nunavut, Canada as a grade sister to all other samples.

All analyses support the monophyly of Chilean *T. fuegianus* (n = 11) and suggest closer affinities to North American samples, specifically to those of northwestern North America (Figs. 1, 2), than to those of northwestern Europe, as previously hypothesized (Lewis et al., 2014a). A sample from the Olympic Peninsula of Washington State, United States (WA; n = 1) was inferred as the sister to samples from Chile with high support in RAxML analyses of MS 60, 40, and 20 supermatrices (BS 93, 98, and 100, respectively).

Two widespread northern temperate clades were inferred in RAxML analyses of all supermatrices (Fig. 1), each of which is consistently subdivided into distinct and well-supported lineages. Northern Temperate clade 1 (NT 1) is subdivided into two geographically widespread groups and Northern Temperate clade 2 (NT 2) into three regionally structured groups, i.e., Norway + Sweden (western Europe), Labrador (eastern North America), and Alaska (western North America). NT 1 and 2 form a monophyletic group sister to samples from Chile + WA in the MS 20 topology, but a paraphyletic group, with samples from Chile + WA nested between the two northern temperate clades in the MS 60 and 40 topologies (Appendix S5). Analysis of the MS 20 and MS 60 northern temperate subset (n = 65) supermatrices also recovered the same northern temperate clades, but with ambiguous placement of WA among the northern temperate samples (Appendix S6).

Coalescent phylogenetic analyses—The SVDquartets lineage tree exhaustive search assessed 1,663,740 quartets. For the MS20 supermatrix 13.31% (221,291) of quartets were incompatible and 86.69% (1,441,408) were compatible, with 1041 discarded as uninformative or indecisive. For the MS60 supermatrix 12.29% (204,207) were incompatible, 87.71% (1,457,862) were compatible, and 1608 were discarded as uninformative or indecisive.

Lineage tree topologies were entirely congruent across the MS 20, 40, and 60 supermatrices (MS 20 lineage tree is shown in Appendix S7). As expected, support values increased with the size of the supermatrix. The inferred topology (Appendix S7) most closely matched the topology inferred from RAxML analysis of the MS20 supermatrix (Fig. 1), including the monophyly of NT clades 1 and 2, and three subclades of NT 2. Unlike the RAxML inferences, SVDquartets analyses resolved NT 1 as a grade of samples from across the northern temperate sampling areas, including the sample from Nunavut, rather than as two subclades, with Nunavut forming a distinct lineage. All SVDquartets lineage tree analyses (Appendix S7) and species tree analyses (Appendix S8; with the exception of the MS 60 species tree where *a priori* sample assignment was based on collection locality) resolved a sister relationship between samples from Chile and the sample from WA.

Individual assignment—STRUCTURE analyses were run for both sample subsets (n = 81 and n = 65) using MS 60 and 20 SNP data sets for K values 2 through 7. Results were consistent across MS 60 and MS 20 data sets, and only MS 20 results at optimal K values are reported for each sample subset. Major and minor alternate modes were inferred in all STRUCTURE analyses. The most commonly inferred modes, or major modes, in 10 total runs for each K value are reported in Fig. 2, and minor modes are reported in online Appendix S9. The optimal K value for the all-localities sample subset (n = 81) was K = 3 under the criteria of Evanno et al. (2005) and K = 4 under the criteria of Pritchard et al. (2000). At K = 4, samples from Chile and NT 1 are clustered separately. We regard the results for K = 4, as supported by the highest mean Prob(K) value (Pritchard et al., 2000), or K = 5, as more informative than K = 3 for the n = 81 sample set, congruent with RAxML and SVDquartets results, which consistently resolved samples from Chile and NT 1 as distinct lineages.

Analysis of the northern temperate sample subset (n = 65) provided better resolution of the population structure among the northern temperate samples, with clear resolution of both NT 1 and NT 2 as well as haplotype groups within each of these clades as inferred in phylogenetic analyses (Fig. 1 and Appendix S7). The optimal K value for the northern temperate sampling MS 20 data set was K = 5 according to both criteria. At K = 5, samples from Norway and Sweden are resolved as a distinct cluster suggesting that K = 4 is too stringent. At K = 5 and K = 6 samples from Labrador cluster as an admixed group.

DISCUSSION

The sister lineage to* Tetraplodon fuegianus *recovered in western North America—The southern South American endemic *Tetraplodon fuegianus* marks the southernmost disjunction in a broadly distributed bipolar lineage (Lewis et al., 2014a). A lack of phylogenetic structure in the bipolar lineage based on four discrete loci previously precluded efforts to determine the source of the direct LDD event that gave rise to *T. fuegianus*. Here, based on a large-scale RADseq data set, we resolve much greater structure within the bipolar lineage. The monophyletic *T. fuegianus* lineage is resolved as sister to a sample from Washington, USA, which appears to represent a rare lineage of *T. mnioides* (Figs. 1, 2). Based on this result, we hypothesize that northwestern North America was the geographic

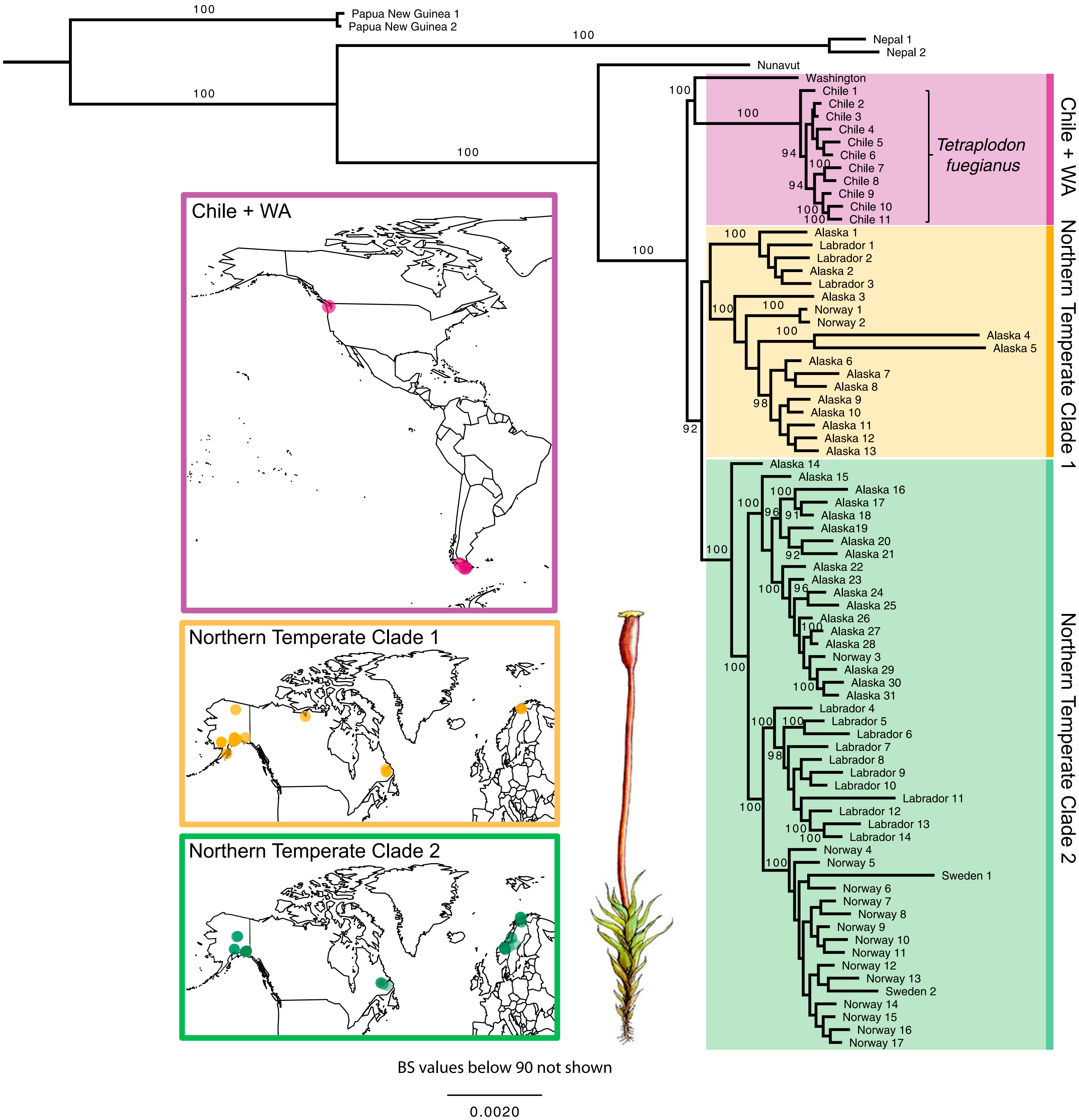

FIGURE 1 Phylogenetic relationships within the bipolar lineage of the dung moss *Tetraplodon* based on the analysis of 4077 RADseq loci for 81 samples in RAxML (including loci present in ≥20 samples). The three major clades, as found in the SVDquartets coalescent model and STRUCTURE Bayesian individual assignment analyses, are highlighted, and their geographic distributions are shown in the maps. Samples from Chile represent the southern South American endemic species *T. fuegianus*. Samples from Papua New Guinea represent the endemic species *T. lamii*. All other samples are morphologically circumscribed as *T. mnioides*. Bootstrap values are shown at the relevant node.

source of the ancestor to *T. fuegianus*, which is in line with the link between the floras of the American North and South Pacific coasts discussed for angiosperm groups based on floristic data (Raven, 1963) and molecular data (Popp et al., 2011).

Widespread northern temperate lineages—All analyses except those constrained by geographic locality (Appendix S8) resolve two widespread and partially sympatric northern temperate lineages. NT 1 is composed primarily of Alaskan samples, as well as samples

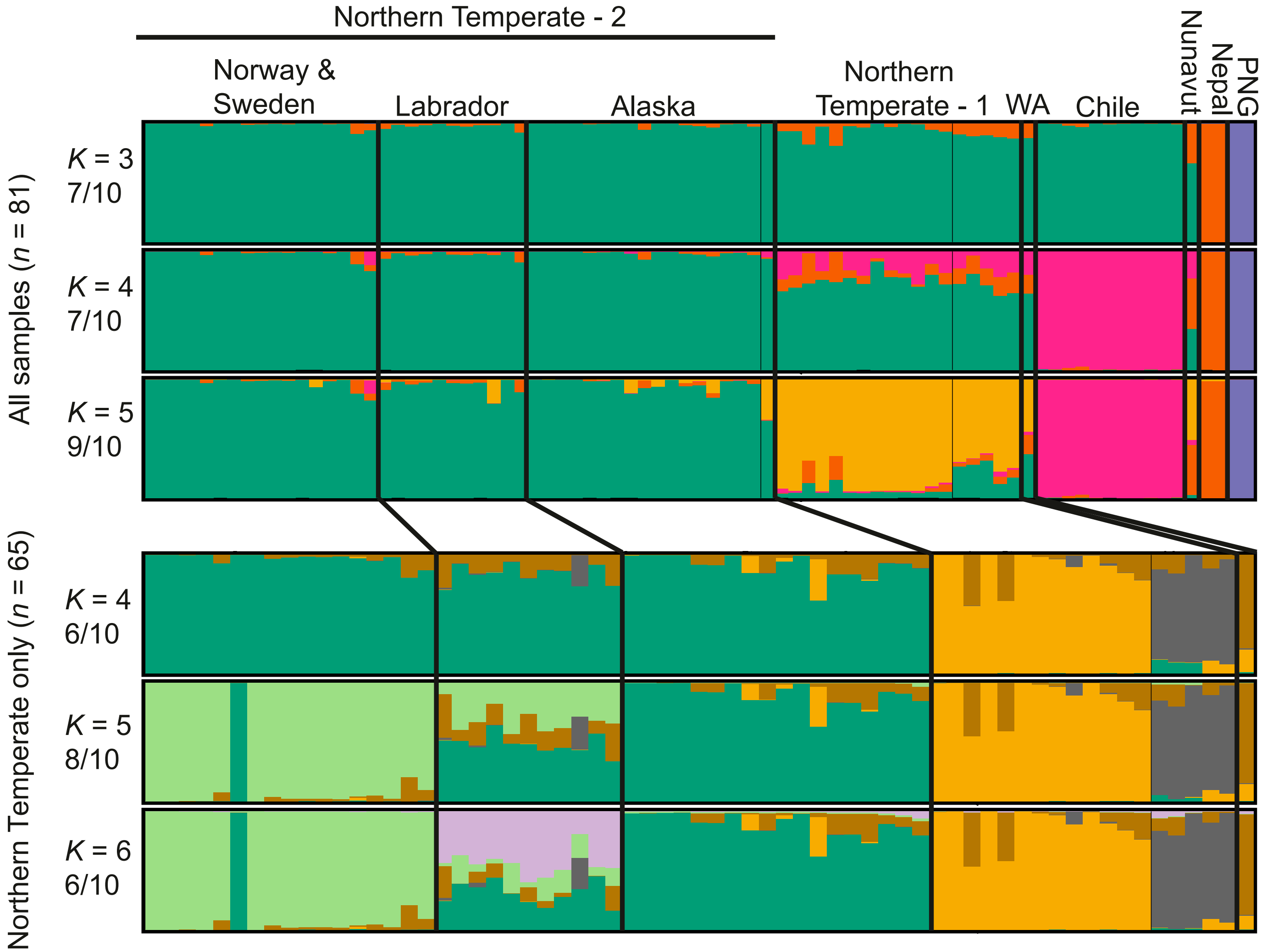

FIGURE 2 Spatial genetic structure as estimated with STRUCTURE Bayesian individual assignment analyses within the bipolar lineage of the dung moss *Tetraplodon*. All analyses inferred major and minor modes, and only major modes are shown here with the fraction of runs supporting each mode indicated below the respective *K* value. The top three plots depict results from analyses of all 81 samples based on 4020 single nucleotide polymorphisms (SNPs) including loci present for ≥20 samples; the bottom three plots show results from analyses including 65 northern temperate samples based on 3508 SNPs including loci present for ≥20 of samples. Samples from Chile represent the southern South American endemic species *T. fueginus*. Samples from Papua New Guinea (PNG) represent the endemic species *T. lamii*. All other samples are morphologically circumscribed as *T. mnioides*, including the sample from Washington (WA).

from Norway and Labrador, and NT 2 is composed of three geographically structured groups corresponding to Alaska, Norway + Sweden, and Labrador. Given that most of these sites deglaciated in the last 20 kyr, this distribution is the hallmark of recent migration. The presence of genetically distinct, widespread, and sympatric northern temperate clades mirrors the pattern previously recovered within the genus as a whole (Lewis et al., 2014a) and is consistent with patterns found in other phylogeographic studies on bryophytes (e.g., Hedenäs, 2012; Piñeiro et al., 2012). For example, recent analysis of genetic data for eight bryophyte species disjunct across the North Atlantic Ocean (i.e., amphi-Atlantic disjuncts) recovered evidence supporting frequent dispersal and limited geographic structuring of genetic diversity among amphi-Atlantic disjunct bryophyte populations. F_{ST} values for amphi-Atlantic disjunct populations were similar to those estimated here for western European and western or eastern North American *Tetraplodon* populations (Table 3), congruent with the suggestion that North American and European populations may be best considered as large meta-populations for bryophytes (Désamoré et al., 2016).

Given the widespread nature of northern temperate *Tetraplodon* clades, it is possible that the ancestor of *T. fuegianus* was also part of a widespread haplotype group, the descendants of which we sampled only in Washington and southern Chile. Given the rarity of this lineage, we cannot be certain that extant representatives are not found in other localities, but went undetected despite our sampling intensity. Alternatively, the WA sample may represent a haplotype restricted to areas previously south of the Laurentide ice sheet. Haplotypes distributions in many *Sphagnum* species show a distinct genetic discontinuity between Beringian populations and those in areas previously south of the Laurentide ice sheet (Kyrkjeeide et al., 2016a) despite the high dispersal potential of *Sphagnum* spores (Sundberg, 2013). Based on the sampling at hand, our data indicate that a single LDD event gave rise to *T. fuegianus*, and that event involved a lineage that is now rare and possibly restricted to the Pacific Northwest of North America.

In spite of the evidence for intercontinental dispersal, the main lineages in the bipolar *Tetraplodon* lineage remain surprisingly isolated, much like cryptic species (Shaw, 2001; Heinrichs et al., 2009). Representatives of these lineages were often found in sites quite close to one another, and yet the lineages remain distinct. Gene flow between the lineages caused by mating would generate what looks like homoplasy, resulting in a star phylogeny with long tips and limited topological structure, a pattern seen in highly outbred mosses, like *Ceratodon purpureus* (Hedw.) Brid. (McDaniel and Shaw, 2005). The topological structure resolved here suggest that the hermaphroditic (autoicous) *T. mnioides*, may be highly inbred, consistent with patterns found in other hermaphroditic mosses (Eppley et al., 2007; Perroud et al., 2011). High rates of inbreeding in *T. mnioides*, however, would be surprising since single patches of *T. fuegianus* may contain multiple haplotypes (Lewis et al., 2016), thus providing ample opportunity for outcrossing between individuals in a patch. Alternatively, the different lineages may be at least partially reproductively isolated. Nevertheless, no morphological traits have been identified that distinguish the different lineages, and apart from its distribution, *T. fuegianus* is morphologically indistinguishable from *T. mnioides*.

Finding phylogeographic signal despite incongruent locus histories—The difference in bootstrap (BS) support values between RAxML and SVDquartets analyses inferring congruent topologies, as well as resolution of minor and major modes for all STRUCTURE analyses (Fig. 2; Appendix S9), suggests incongruent gene histories. SVDquartets analyses using the coalescent model resulted in lower BS support values due to the ambiguity associated with incongruent gene histories (Chifman and Kubatko, 2014). The presence of incongruent gene histories is evident in the results of SVDquartet analyses, which yielded 13.31% and 12.29% incompatible quartets for MS 20 and MS 60 supermatrices, respectively. Simulation studies have shown, however, that even in cases of incomplete lineage sorting, phylogenies may be reliably reconstructed given sufficient locus and taxon sampling (Maddison and Knowles, 2006). Phylogenetic inferences based on RADseq loci for oak trees (Hipp et al., 2014) and the Lake Victoria cichlids (Wagner et al., 2013) provide notable examples supporting the power of extensive genomewide locus sampling in recovering accurate phylogenies in groups despite incomplete lineage sorting.

CONCLUSIONS

Here we have shown that the bipolar disjunct *T. fuegianus* arose following a single long-distance dispersal event involving a *T. mnioides* lineage that is now rare in the northern hemisphere and potentially restricted to the Pacific Northwest of North America. *Tetraplodon mnioides* is a highly vagile species in the northern hemisphere, but curiously, gene flow among sympatric lineages is limited, possibly due to high rates of selfing or reproductive isolation. Much of the previous work on bipolar taxa have resolved the northern hemisphere as the source for bipolar range expansions, but have been unable to identify specific geographic regions of origin. This trend may be due to the difficulty of resolving recent and rapid radiations with data sets of several standard discrete loci. However, in cases of highly vagile organisms, additional challenges may arise if ancestors were part of a widespread haplotype group.

ACKNOWLEDGEMENTS

We thank the U. S. National Science Foundation for Doctoral Dissertation Improvement Grant DEB-1311405 and Graduate Research Fellowship to L.R.L, the Chilean Institute for Ecology and Biodiversity and Omora Ethnobotanical Park (for grants ICM P05-002 and Basal-CONICYT PFB-23), the U. K. Natural Environment Research Council (NERC) for studentship (ref NE/K50094X/1) to E.M.B, and NERC core funding to the British Antarctic Survey Biodiversity, Evolution and Adaptation Team, and Ms Trudy Gerlach for financial support. We are grateful to Nasim Rahmatpour, who performed flow cytometry and assisted with interpretation of results. We thank Paul O. Lewis and Jill Wegrzyn at the UCONN Bioinformatics Facility for help with data processing and analysis, Joerg Graf and Mike Nelson at the UCONN MiSeq Facility and Bo Reese at the UCONN Center for Genome Innovation for help with data generation. We also thank William R. Buck, Juan Larraín, Matt von Konrat, Dietmar Quandt, Leopoldo G. Sancho, Kristian Hassel, Tommy Prestø, Lars Hedenäs, Blanka Aguero, and Jon Shaw for help with fieldwork. Liu Yang and Rafael Medina helped with various aspects of the project, and the McDaniel Laboratory provided feedback on drafts. Finally, we thank the anonymous reviewers, whose thoughtful comments significantly strengthened this manuscript.

DATA ACCESSIBILITY

Chloroplast *rps4* DNA sequences are available in NCBI GenBank. GenBank accession numbers for all *rsp*4 sequences are listed in Appendix S1. Trimmed and demultiplexed RAD sequences, aligned supermatrices (NEXUS files), and SNP matrices (STR files) are available in the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.2cs4v).

LITERATURE CITED

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3: e3376.

Bednarek-Ochyra, H., J. Vana, R. Ochyra, and R. I. Lewis Smith. 2000. The liverwort flora of Antarctica. W. Szafer Institute of Botany, Polish Academy of Sciences, Krakow, Poland.

Biersma, E. M., J. A. Jackson, J. Hyvönen, S. Koskinen, K. Linse, H. Griffiths, and P. Convey. 2017. Global biogeographic patterns in bipolar moss species. *Royal Society Open Science* 4: 170147.

Chifman, J., and L. Kubatko. 2014. Quartet inference from SNP data under the coalescent Model. *Bioinformatics (Oxford, England)* 30: 3317–3324.

Crame, J. A. 1993. Bipolar molluscs and their evolutionary implications. *Journal of Biogeography* 20: 145–161.

Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. Murray, London, UK.

Désamoré, A., J. Patiño, P. Mardulyn, S. F. Mcdaniel, F. Zanatta, B. Laenen, and A. Vanderpoorten. 2016. High migration rates shape the postglacial history of amphi-Atlantic bryophytes. *Molecular Ecology* 25: 5568–5584.

Du Rietz, G. E. 1940. Problems of bipolar plant distribution. *Acta Phytogeographica Suecica* 13: 215–282.

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

Eppley, S., P. Taylor, and L. Jesson. 2007. Self-fertilization in mosses: A comparison of heterozygote deficiency between species with combined versus separate sexes. *Heredity* 98: 38–44.

Etter, P. D., and E. Johnson. 2012. RAD paired-end sequencing for local de novo assembly and SNP discovery in non-model organisms. *In* F. Pompanon and A. Bonin [eds.], Data production and analysis in population genomics: Methods and protocols, methods in molecular biology, vol. 888, 135–151. Springer Science+Business Media, New York, New York, USA.

Etter, P. D., J. L. Preston, S. Bassham, W. A. Cresko, and E. A. Johnson. 2011. Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PLoS One* 6: e18561.

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

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.

Fernández-Mendoza, F., and C. Printzen. 2013. Pleistocene expansion of the bipolar lichen *Cetraria aculeata* into the Southern Hemisphere. *Molecular Ecology* 22: 1961–1983.

Frisvoll, A. A. 1978. The genus *Tetraplodon* in Norway. A taxonomic revision. *Lindbergia* 4: 225–246.

Goudet, J. 2005. HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. *Molecular Ecology Resources* 5: 184–186.

Hedenäs, L. 2009. Haplotype variation of relevance to global and European phylogeography in *Sarmentypnum exannulatum* (Bryophyta: Calliergonaceae). *Journal of Bryology* 31: 145–158.

Hedenäs, L. 2012. Global phylogeography in *Sanionia uncinata* (Amblystegiaceae: Bryophyta). *Botanical Journal of the Linnean Society* 168: 19–42.

Heinrichs, J., J. Hentschel, K. Feldberg, A. Bombosch, and H. Schneider. 2009. Phylogenetic biogeography and taxonomy of disjunctly distributed bryophytes. *Journal of Systematics and Evolution* 47: 497–508.

Hipp, A. L., D. A. R. Eaton, J. Cavender-Bares, E. Fitzek, R. Nipper, and P. S. Manos. 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. *PLoS One* 9: e93975.

Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9: 1322–1332.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics (Oxford, England)* 23: 1801–1806.

Jombart, T. 2008. *adegenet*: A R package for the multivariate analysis of genetic markers. *Bioinformatics (Oxford, England)* 24: 1403–1405.

Jombart, T., and I. Ahmed. 2011. *adegenet* 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics (Oxford, England)* 27: 3070–3071.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. CLUMPAK: Program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources* 15: 1179–1191.

Kreier, H. P., K. Feldberg, F. Mahr, A. Bombosch, A. R. Schmidt, R. L. Zhu, M. von Konrat, et al. 2010. Phylogeny of the leafy liverwort *Ptilidium*: Cryptic speciation and shared haplotypes between the Northern and Southern Hemispheres. *Molecular Phylogenetics and Evolution* 57: 1260–1267.

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

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

Lewis, L. R., Y. Liu, R. Rozzi, and B. Goffinet. 2016. Infraspecific variation within and across complete organellar genomes and nuclear ribosomal repeats in a moss. *Molecular Phylogenetics and Evolution* 96: 195–199.

Lewis, L. R., R. Rozzi, and B. Goffinet. 2014a. Direct long-distance dispersal shapes a New World amphitropical disjunction in the dispersal-limited dung moss *Tetraplodon* (Bryopsida: Splachnaceae). *Journal of Biogeography* 41: 2385–2395.

Lynch, M. 2008. Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. *Molecular Biology and Evolution* 25: 2409–2419.

Maddison, W. P., and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55: 21–30.

McDaniel, S. F., and A. J. Shaw. 2005. Selective sweeps and intercontinental migration in the cosmopolitan moss *Ceratodon purpureus* (Hedw.) Brid. *Molecular Ecology* 14: 1121–1132.

Norris, D. H., T. Koponen, and S. Piippo. 1999. Bryophyte flora of the Huon Peninsula, Papua New Guinea. LXVI. Meesiaceae (Musci), with lists of boreal to temperate disjunct, bipolar, and widely spread species in New Guinea. *Annales Botanici Fennici* 36: 257–263.

Ochyra, R. 1992. *Amblyodon dealbatus* (Musci, Meesiaceae)—A bipolar disjunct. *Fragmenta Floristica et Geobotanica* 37: 251–259.

Ochyra, R., and W. R. Buck. 2003. *Arctoa fulvella*, new to Tierra del Fuego, with notes on trans-American bipolar bryogeography. *The Bryologist* 106: 532–538.

Ochyra, R., R. I. Lewis Smith, and H. Bednarek-Ochyra. 2008. The Illustrated Moss Flora of Antarctica. Cambridge University Press, New York, New York, USA.

Øvstedal, D. O., and R. I. Lewis Smith. 2001. Lichens of Antarctica and South Georgia: A guide to their identification and ecology. Studies in polar research. Cambridge University Press, Cambridge, UK.

Pearce, D. A., C. S. Cockell, E. S. Lindström, and L. J. Tranvik. 2007. First evidence for a bipolar distribution of dominant freshwater lake bacterioplankton. *Antarctic Science* 19: 245–252.

Perroud, F.-P., D. J. Cove, R. S. Quatrano, and S. F. McDaniel. 2011. An experimental method to facilitate the identification of hybrid sporophytes in the moss *Physcomitrella patens* using fluorescent tagged lines. *The New Phytologist* 191: 301–306.

Piñeiro, R., M. Popp, K. Hassel, D. Listl, K. B. Westergaard, K. I. Flatberg, H. K. Stenøien, and C. Brochmann. 2012. Circumarctic dispersal and long-distance colonization of South America: The moss genus *Cinclidium*. *Journal of Biogeography* 39: 2041–2051.

Popp, M., V. Mirré, and C. Brochmann. 2011. A single mid-Pleistocene long-distance dispersal by a bird can explain the extreme bipolar disjunction in crowberries (*Empetrum*). *Proceedings of the National Academy of Sciences, USA* 108: 6520–6525.

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

Raven, P. H. 1963. Amphitropical relationships in the floras of North and South America. *The Quarterly Review of Biology* 38: 151–177.

Rensing, S. A., D. Lang, A. D. Zimmer, A. Terry, A. Salamov, H. Shapiro, T. Nishiyama, et al. 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319: 64–69.

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

Schofield, W. B. 1974. Bipolar disjunctive mosses in the Southern Hemisphere, with particular reference to New Zealand. *Journal of the Hattori Botanical Laboratory* 38: 13–32.

Schuster, R. M. 1983. Phytogeography of the bryophyta. *In* R. M. Schuster [ed.], New manual of bryology, 463–626. Hattori Botanical Laboratory, Nichinan, Japan.

Shaw, A. J. 1986. A new approach to the experimental propagation of bryophytes. *Taxon* 35: 671–675.

Shaw, A. J. 2001. Biogeographic patterns and cryptic speciation in bryophytes. *Journal of Biogeography* 28: 253–261.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)* 22: 2688–2690.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* 57: 758–771.

Streimann, H. 1998. Bryological comparisons between Australia and Northern Europe. *Folia Cryptogamica Estonica* 32: 97–105.

Sul, W. J., T. A. Oliver, H. W. Ducklow, L. A. Amaral-zettler, and M. L. Sogin. 2013. Marine bacteria exhibit a bipolar distribution. *Proceedings of the National Academy of Sciences, USA* 110: 2342–2347.

Sundberg, S. 2013. Spore rain in relation to regional sources and beyond. *Ecography* 36: 364–373.

Swofford, D. L. 2001. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0a142. Sinauer, Sunderland, Massachusetts, USA.

Villaverde, T., M. Escudero, M. Luceño, and S. Martín-Bravo. 2015. Long-distance dispersal during the middle-late Pleistocene explains the bipolar disjunction of *Carex maritima* (Cyperaceae). *Journal of Biogeography* 42: 1820–1831.

Wagner, C. E., I. Keller, S. Wittwer, O. M. Selz, S. Mwaiko, L. Greuter, A. Sivasundar, and O. Seehausen. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* 22: 787–798.

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

Wen, J., and S. M. Ickert-Bond. 2009. Evolution of the Madrean–Tethyan disjunctions and the North and South American amphitropical disjunctions in plants. *Journal of Systematics and Evolution* 47: 331–348.

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence data sets under the maximum likelihood criterion. University of Texas at Austin, Austin, Texas, USA. https://www.nescent.org/wg_garli/Main_Page.