

RESEARCH ARTICLE

Journal of
Biogeography

WILEY

A global analysis of mosses reveals low phylogenetic endemism and highlights the importance of long-distance dispersal

Katie Kobara Sanbonmatsu  | Daniel Spalink

Department of Ecology and Conservation
Biology, Texas A&M University, College
Station, Texas, USA

Correspondence

Katie Kobara Sanbonmatsu, Department
of Ecology and Conservation Biology,
Texas A&M University, College Station, TX
77843, USA.

Email: katie.sanbonmatsu@tamu.edu

Funding information

Frank W. Gould fellowship; National
Institute of Food and Agriculture (McIntire
Stennis Project 1018692); National
Science Foundation; National Institute of
Food and Agriculture

Handling Editor: Alain Vanderpoorten

Abstract

Aim: Digitization of herbarium specimens and DNA sequencing efforts in the past decade have enabled integrative analyses of patterns of diversity and endemism in a phylogenetic context. Here, we compare the best available floristic databases to a comprehensive specimen database to examine spatial patterns of moss phylogenetic assembly. We test the hypotheses that (1) mosses exhibit phylogenetic regionalization, (2) islands contain significantly high phylogenetic diversity and (3) that moss phylogenetic endemism is low on a global scale.

Location: Global.

Taxon: Mosses.

Methods: We developed a phylogeny of 3654 moss species using 25 markers and compiled a global specimen database from online repositories. We calculated floristic and phylogenetic measures of diversity and endemism and performed randomizations to test for significant deviations from expectations. We use rarefaction and extrapolation to alleviate substantial differences in sampling effort across the globe. We used both phylogenetic and floristic methods to test for spatial regionalization. We compare our specimen-based results to those obtained using a floristic dataset.

Results: Phylogenetic diversity is more robust to missing data than species richness. Mean phylogenetic distance was significantly higher than expected in areas with high species richness, indicating that reported richness in these areas is likely a product of repeated colonization. Phylogenetic endemism is low globally. Phylogenetic regionalizations cluster into a Holarctic/Holantartic temperate region, a pantropical region, and a region composed of Australia, New Zealand and South Africa.

Main Conclusions: Future efforts for collecting, sequencing and databasing moss species should focus on the tropics, particularly Africa and Southeast Asia. We provide further evidence to support several important theories developed in moss biogeography, including the role of long-distance dispersal in shaping floristic patterns, the dominance of anagenesis in driving patterns of island diversity, and the role of climatic instability in driving patterns of assembly in the Holarctic.

KEYWORDS

assembly, biodiversity, biogeography, floristics, global diversity, long-distance dispersal, phylogenetic diversity, phylogenetic endemism, regionalization



1 | INTRODUCTION

Understanding patterns of biodiversity across multiple spatial scales remains a central goal of ecology and evolutionary biology. In vascular plants, considerable progress has been made toward this goal, both in terms of understanding the processes by which lineages diversify and spread across the globe and how they have assembled into communities (Spalink, Kriebel, et al., 2018; Zhang et al., 2021). By contrast, nonvascular plants—such as mosses—have received far less attention. Mosses are comprised of nearly 13,000 species, which form important components of many terrestrial ecosystems across the globe—from tundra to tropical cloud forests (Ah-Peng et al., 2017; Clymo, 1987; Goffinet & Shaw, 2009). The distributions of individual species and clades are complex, ranging from narrowly endemic to nearly cosmopolitan (Geffert et al., 2013). Patterns of assembly are similarly confounding, with hotspots of species diversity occurring in tropical, temperate and arctic ecosystems in both continental and insular regions (Ah-Peng et al., 2012; Câmara et al., 2019; Geffert et al., 2013; Séneca & Söderström, 2008). Given their evolutionary diversity, ecological importance, and functional distinctness from vascular plants, mosses present a unique system for understanding the patterns of assembly on a global scale.

Mosses exhibit several traits that should be expected to defy biogeographical theories developed with vascular plant models. For example, long distance dispersal (LDD) is common in mosses, which have tiny, buoyant spores (Patiño & Vanderpoorten, 2018). A propensity for LDD should be expected to reduce the genetic isolation and speciation of distant populations (Givnish, 2010), but diversification rates in mosses are dynamic, much like other groups of plants (Medina et al., 2018). Further, high rates of LDD can lead to an inverse isolation effect in some cases, wherein more isolated sites (i.e., islands) receive a greater proportion of diaspores from increasingly distant sources and thus exhibit higher diversity than less isolated sites (Barbé et al., 2016; Lönnell et al., 2012; Sundberg, 2005, 2013; Szövényi et al., 2012). However, many moss species exhibit strong genetic structure at both local and regional scales, suggesting that dispersal alone cannot explain distribution patterns, and that establishment success may impose stronger limitations than dispersal (Ledent et al., 2020; Snäll et al., 2004; Vanderpoorten et al., 2019). Further, mosses do not exhibit a clear latitudinal alpha diversity gradient (Geffert et al., 2013; Möls et al., 2013; Shaw et al., 2005), likely due in part to the inverse isolation effect (Patiño & Vanderpoorten, 2018; Sundberg, 2005). Beyond their dispersal ability, mosses also differ from most vascular plants in their ability to tolerate extremely cold temperatures and to withstand drought through opportunistic desiccation and rehydration (Goffinet & Shaw, 2009). Ultimately, mosses rely on water for successful growth and reproduction, but moisture availability can be highly seasonal and come in different forms. These traits should allow mosses to successfully colonize and diversify within habitats that would be otherwise limiting to most vascular plants (Patiño & Vanderpoorten, 2018).

Despite these differences, biogeographic regionalization in vascular and nonvascular plants appears to be largely congruent. For example, similar to vascular plants, liverworts show distinct Laurasian and Gondwanan signatures (Vanderpoorten et al., 2010). Various moss clades also show separation across temperature and precipitation gradients at the continental scale similar to that of vascular plants (Mateo et al., 2016; Van Rooy & Van Wyk, 2013). Though regionalization appears to be congruent, these patterns have likely arisen as a consequence of different processes, as exhibited by overall low levels of endemism in mosses and hotspots of moss diversity occurring in different areas from those of vascular plant diversity (Patiño & Vanderpoorten, 2018).

At a global scale, patterns of diversity in mosses have been studied several times since the first work of this nature was conducted by Herzog (1926) (Buck & Thiers, 1989; Geffert et al., 2013; Hedenäs, 2007; Norhazrina et al., 2017; Schofield, 1992; Schuster, 1983; Shaw et al., 2005). Shaw et al. (2005) investigated global patterns of moss diversity using a dataset of two organellar genes across 554 different taxa and distribution data based on 86 regional checklists. They found that phylogenetic diversity peaks in the Southern Hemisphere, but that all major lineages have extant representatives across all latitudinal partitions. Geffert et al. (2013) utilized an inventory-based diversity mapping approach using a combined dataset composed of both regional and national checklists. Möls et al. (2013) approached the issue in a similar way, obtaining species richness values for specified regions from the literature. They also found that coastline length and annual precipitation ranges were significantly correlated to moss species richness in both tropical and temperate zones, highlighting the relative importance of moisture over temperature or seasonality in moss diversity. Hedenäs (2007) also utilized checklist data to assess patterns of beta diversity in pleurocarpus mosses across the globe, revealing that tropical assemblages have much higher species turnover than extra-tropical ones. In a similar study, Norhazrina et al. (2017) demonstrated a greater differences in species turnover between, than within, tropical regions.

These studies primarily used secondary occurrence data, relying on expert taxonomic opinion and the synthesis of primary occurrence data. Ideally, primary occurrence data should be preferred for such studies because they are more reproducible and increase precision and resolution (Peterson et al., 2011). Thus, specimens housed in biological collections represent a wealth of crucial data that spans across centuries. However, in many cases, these data are subject to a hoard of potential bias (Meineke & Daru, 2021). These include geographic, temporal and taxonomic bias in collecting efforts, as well as bias in sequencing and digitization efforts.

Here, we use massive databases of digitized herbarium specimens and DNA sequences to determine their efficacy in providing a framework for understanding phylogenetic patterns of moss diversity, with direct comparisons to floristic datasets. We explore three fundamental questions in bryology: (1) What are the patterns of phylogenetic diversity across all mosses at a global scale? (2) What are phylogenetic patterns of regionalization in mosses? (3) How do

the best floristic data compare to species distribution data from on-line repositories, and how do metrics derived from these two types of data compare? We test the hypotheses that (1) mosses exhibit phylogenetic regionalization, (2) islands contain significantly high phylogenetic diversity and (3) that moss phylogenetic endemism is low on a global scale.

2 | MATERIALS AND METHODS

2.1 | Taxonomy

We expected sequence data availability to be the limiting taxonomic factor in our goal to develop a comprehensive moss specimen database and global moss phylogeny, so we reduced our search of occurrence records to only those species for which we could obtain adequate sequence data from GenBank (see below). We used Tropicos (tropicos.org) to generate a nomenclatural database to ensure consistency across all datasets. First, species names were parsed from each GenBank entry. We then identified the accepted names for these species, synonyms of the accepted names, synonyms of the names as listed in GenBank and synonyms of the synonyms in the R package *taxize* v0.9.99 (Chamberlain & Szöcs, 2013). If a species had an accepted name, we used that name. If not, we used the name as listed in the original GenBank entry. There were several types of issues that arose during this process. Some names had dual synonymy. Some names as listed by GenBank had multiple accepted names. To resolve these issues, we investigated the relevant literature to clarify the most recent taxonomic treatment for a given species (see Appendix S1). All generic names were standardized to the classification of Goffinet and Buck (2020). We used this final database to ensure nomenclatural consistency in distributional and phylogenetic data. Our nomenclatural database is available on Dryad (doi:10.5061/dryad.905qfttm2).

2.2 | Phylogenetic analyses

We developed a phylogeny of mosses by expanding the taxonomic and molecular sampling in the supermatrix dataset of Rose et al. (2016). All data used by Rose et al. (2016) were retained in our analyses, but we added 524 taxa and 5 genes. These new data were downloaded from GenBank (Benson et al., 2012) using the *rentrez* v1.2.3 R package (Winter, 2017). Sequences for synonymous taxa were either merged or deleted in AliView (Larsson, 2014) and R (see Appendix S2). Individual gene alignments were generated in MAFFT v7.397 (Kato & Standley, 2013) and individually inspected in AliView (Larsson, 2014) before concatenation.

We reconstructed a phylogeny from the concatenated supermatrix using maximum likelihood as implemented in RAxML-HP v.8 on XSEDE (Stamatakis, 2014) on the Cipres Science Gateway V.3.3 (Miller et al., 2010). We partitioned the dataset by gene region, and selected the GTRGAMMA model because the most parameter-rich

model has been shown to lead to similar inferences as the 'best' model (Abadi et al., 2019). We used the RAxML default rapid hill-climbing algorithm, with 867 rapid bootstrapping replicates, with the best ML tree as a topological constraint on the bootstrap replicates.

We used treePL to time-calibrate the phylogeny (Smith & O'Meara, 2012). The calibration dates used were modified from those of Rose et al. (2016) (see Appendix S3; Table S3.1). These included mostly secondary dates from Morris et al. (2018), Newton et al. (2007) and Villarreal and Renner (2014), but we also included three primary dates. Additional details are presented in Appendix S2.

The tree was rooted using 91 liverwort and hornwort taxa. The gene regions represented include two nuclear (*ITS1-5.8s-ITS2*, *26s*), five mitochondrial (*atp1*, *atp6-rps7*, *nad2*, *nad5-nad4*) and 18 plastid (*atpB-rbcL*, *psbB* operon [*psbT-psbN-psbH*], *rpl2*, *trnE*, *trnM*, *atpF-atpH*, *psbE-psbF-psbL-psbJ*, *rpl32*, *trnG*, *rps4-trnS*, *ndhF*, *rbcL*, *rpoC1*, *trnK-psbA*, *psbA*, *rpl16*, *rps4*, *trnL-trnF*) genes.

2.3 | Species distribution data

Distribution data were downloaded from GBIF (<https://doi.org/10.15468/dd.hpjykh>) and iDigBio using the R packages *rgbif* v.3.5.2 (Chamberlain et al., 2021) and *ridigbio* v.0.3.5 (Michonneau & Collins, 2017), respectively. We downloaded data for each species in the phylogeny individually, searching for both accepted and synonymous names and ensuring nomenclatural consistency using our newly generated database. Species without scientific names and authorities were omitted from the datasets to prevent the inclusion of homonyms in the analysis.

We cleaned the distribution data using the R package *CoordinateCleaner* v2.0.18 (Zizka et al., 2019), removing duplicate records, those within 100 meters of a biodiversity institution, within 10 km of country capitals and country centroids, within 100 meters of GBIF headquarters, and with identical lat/long points. We also removed records that occurred outside of reference landmasses, using a customized shapefile that included all minor islands (see Appendix S3 for additional details). The cleaned distribution data are available on Dryad (doi:10.5061/dryad.905qfttm2).

2.4 | Biodiversity indices

We calculated several metrics of biodiversity across a global grid of equal-area cells. We calculated species richness (SR)—the number of unique species labels in each cell, to compare to previously published moss studies. We expect our measures of SR to be low, given that our sampling is restricted to species for which we have phylogenetic data. We subsequently removed cells with fewer than six species. These cells had very low sampling effort and were likely to be under-collected (or under-digitized). We calculated phylogenetic diversity (PD) as Faith's Phylogenetic Diversity (Faith, 1992), and the mean phylogenetic distance (MPD) among co-occurring species. We calculated both weighted endemism (WE) and corrected weighted



endemism (CWE)—and their phylogenetic analogues, weighted phylogenetic endemism (PE) and corrected weighted phylogenetic endemism (CWPE) (Crisp et al., 2001). The corrected metrics are WE and PE adjusted for SR and PD, respectively (Crisp et al., 2001).

Digitized moss herbarium records are abundant across the Northern Hemisphere, particularly North America and Europe, but are underrepresented in the tropics, particularly Southeast Asia and Africa. To help account for this spatial bias, we used size-based rarefaction and extrapolation (Chao1) as implemented in the R packages *iNEXT* v2.0.20 (Chao et al., 2014) and *iNEXTPD2* v1.0.3 (Hsieh & Chao, 2017) for SR and PD, respectively. Rarefaction estimates the number of expected species in a random subsample that is less than or equal to the number of samples in the smallest assemblage (Sanders, 1968). Essentially, rarefaction describes the expected SR of an assemblage if fewer samples had been collected. Extrapolation is a related method that predicts SR based on an estimate of asymptotic richness (Chao, 1984). These two corrections were extended to PD by Chao et al. (2015). Colwell et al. (2012) proposed the integration of these two techniques to standardize sample size and calculate SR without discarding potentially informative data. We tested several different sample sizes to calculate rarefied/extrapolated SR (SRrare) and found that the value which parsed the dataset most evenly was 200. We used the same sample size to calculate rarefied/extrapolated PD (PDrare).

We performed randomizations to test whether observed diversity measures were higher or lower than expected by chance. For PD and MPD, we shuffled the distance matrix labels across all taxa and recalculated PD diversity metrics at each iteration. This approach tests whether observed values are greater or lower than expected if phylogenetic relationships were random. For the endemism metrics, we randomly shuffled species ranges while maintaining species richness in each cell. This approach tests whether endemism within cells is greater or lower than expected than if species range sizes were random, given the observed richness in cells. All randomizations were repeated 1000 times. We calculated WE, CWE and CWPE manually in R. PE was calculated in the R package *phyloregion* v1.0.5 (Daru et al., 2020). MPD and PD were calculated and randomized in the R package *picante* v1.8.2 (Kembel et al., 2010).

Because mosses are under collected, and spatial bias in collecting and digitisation are evident in the GBIF dataset, we also calculated global phylogenetic diversity using the floristic dataset compiled by Geffert et al. (2013). Geffert et al. (2013) note that spatial biases also exist for floristic data; thus, the purpose of our comparison was to determine the extent to which specimen data reflect floristic patterns. Using their species lists and spatial shapefile (delimited by political or regional units), we calculated SR, PD and MPD. Because some species present in their dataset were not in our phylogeny, we removed these species before calculating PD and MPD, and re-calculated SR for comparative purposes. This reduced dataset is herein referred to as the “reduced Geffert” dataset. We then overlaid our specimen dataset onto their shapefile and recalculated those metrics. We used linear regression and calculated Pearson's correlation coefficient to

assess how well the reduced Geffert dataset and the specimen data fit the original results of Geffert et al. (2013).

2.5 | Biogeographic regionalizations

We used both phylogenetic and floristic approaches to identify patterns of regionalization. For the phylogenetic approach, we used the R package *phyloregion* to ascertain areas of phylogenetic similarity and dissimilarity. Phylogenetic beta diversity for each cell was calculated using Simpson's index. The number of clusters (k) was determined through k -means clustering and chosen by silhouette scores, and assemblages were agglomerated according to the Ward2 algorithm (Daru et al., 2020; Salvador & Chan, 2004). We used Infomap Bioregions for the floristic approach (Edler et al., 2017). We specified the minimum and maximum cell sizes to be $1^\circ \times 1^\circ$ and $2^\circ \times 2^\circ$, respectively. We specified a cell capacity of 10–600 species, as our highest observed richness was 591 species in a $2^\circ \times 2^\circ$ cell. The cluster cost was set to 1.02, and the algorithm was run for five trials. These parameters were chosen because a higher cluster cost failed to parse any regions, while a lower cluster cost created numerous bioregions consisting of only single cells (generally with very low data).

3 | RESULTS

3.1 | Phylogenetic analyses

The final phylogeny comprises 3524 species and 129 subspecies/varieties/forms across the moss phylogeny, representing 94.5% of families and 79.2% of genera. Our topology is very similar to the Rose et al. (2016) supermatrix phylogeny. The final concatenated alignment was comprised of 24 loci, with a final alignment length of 76,596 bases and 83% missing sequences. (see AppendixS2, Table S2.1; data housed in Dryad (doi:10.5061/dryad.905qfttm2)).

For the dating analyses, we used values of $\text{opt} = 3$, $\text{optad} = 3$ and $\text{optcvad} = 1$, which were the most frequently suggested values after multiple analyses using the prime command in treePL. Random cross-validation found that the best smoothing parameter was 1×10^{-8} with a chi-squared value of 8169.2. The results of the dating analysis are congruent with time estimates from Newton et al. (2007), Rose et al. (2016) and (Morris et al., 2018).

3.2 | Species distribution data

Out of the 3651 taxa in the phylogeny, 3284 taxa had occurrences in GBIF or iDigBio, constituting 90% of the species in the phylogeny. A total of 2,083,613 records with scientific names and authorities were downloaded from GBIF and iDigBio. After cleaning, the final dataset had 1,257,129 individual data points (see AppendixS3 for additional details).

3.3 | Biodiversity indices

Species richness peaks in the Scandinavian Peninsula, with 615 species (Figure 1a). Other areas of high SR in Europe include the Alps and northern Spain. SR was high in the Andes, peaking in Colombia and Ecuador. In North America, SR was highest in the Pacific Northwest

and northeastern U.S.A. and adjacent Canada. Across Asia, high SR was only found in Japan.

In the rarefaction/extrapolation analysis, 817 cells were rarefied and 1472 cells were extrapolated. Well-collected areas tended to be rarefied in our analyses, whereas poorly collected areas were extrapolated (Figure 2d). Rarefied/extrapolated species richness

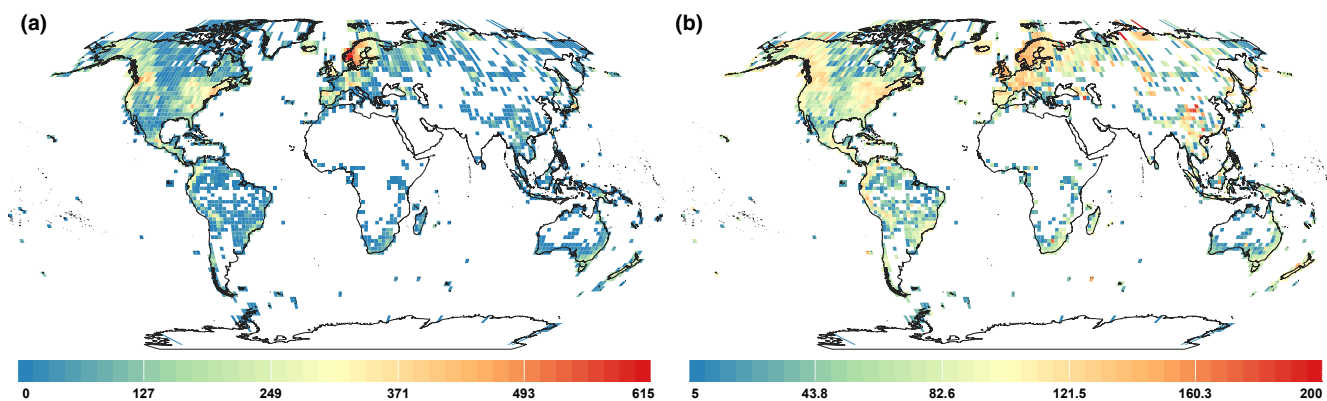


FIGURE 1 Global patterns of moss (a) species richness (SR) and (b) rarefied/extrapolated species richness (SRrare) calculated from taxa included in phylogenetic analyses. (a) SR ranges from 5–615 species, and peaks in the Scandinavian Peninsula. (b) SRrare peaks in northern China and the Taymyr Peninsula

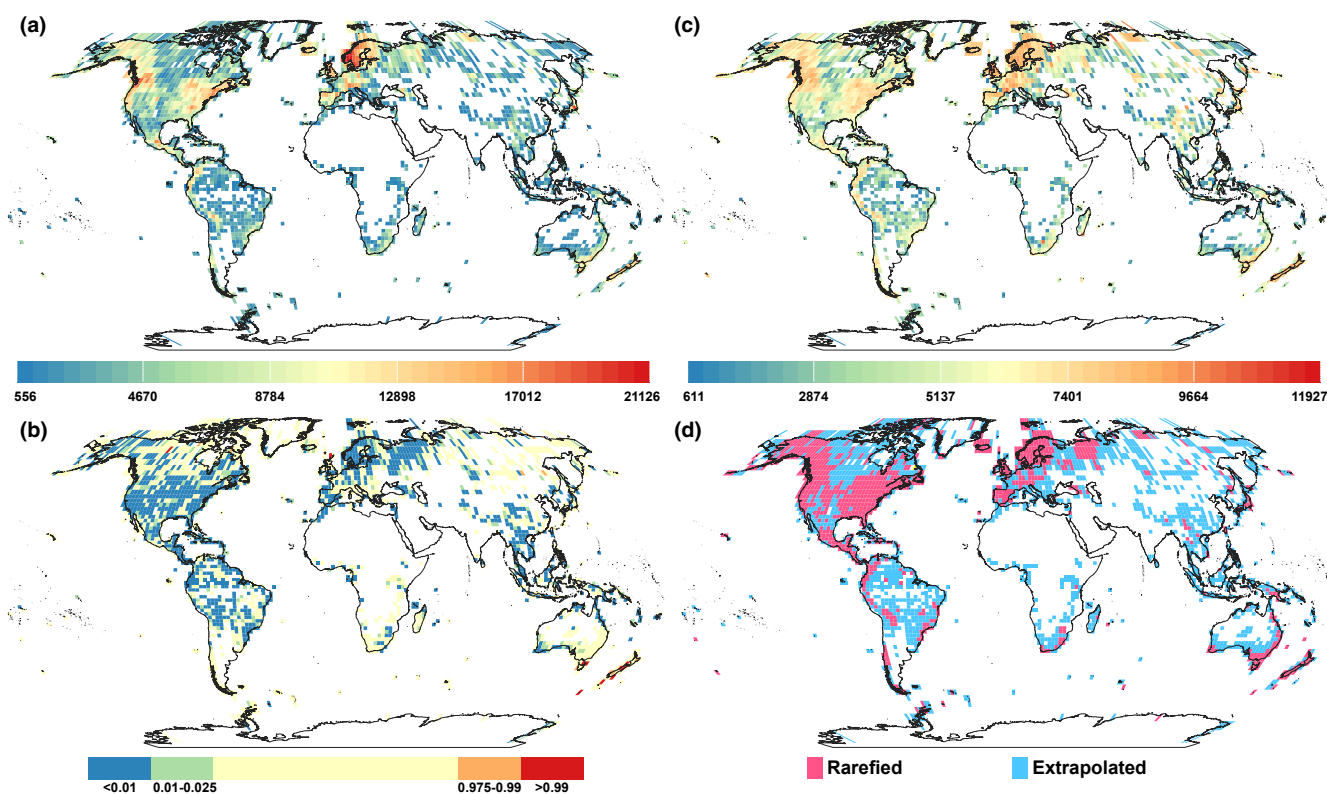


FIGURE 2 Global patterns of moss phylogenetic diversity. (a, b) Depict Faith's phylogenetic diversity (PD) and significant departures of PD from expectations of null models, respectively. (c) Depicts rarefied/extrapolated measures of PD. (d) Displays which cells were rarefied, which were extrapolated, and a single cell containing 200 species (not shown). (a) PD closely mirrors species richness (Figure 4a), peaking in the Scandinavian Peninsula. (b) Most assemblages exhibit significantly less PD given their species richness than expected by chance, or no significant difference than random. The exception to this trend is in Tasmania and New Zealand, which exhibit significantly greater PD than expected. (c) PDrare is overall substantially higher in the tropics and lower in well-sampled areas than raw PD. (d) 817 cells were rarefied, 1472 cells were extrapolated, and one cell had an observed value of 200

(SRrare) peaks in China and in the Taymyr Peninsula (Figure 1b). This is largely in congruence with the original results from Geffert et al. (2013), with the exception of Southeast Asia (Figure S3.1). SRrare is substantially higher in tropical regions than raw SR, particularly in Southeast Asia and nearly all of Africa. Overall, however, SRrare was 91.9% correlated to raw estimates of SR.

Patterns of PD mirror SR, particularly in Europe and North America (Figure 2a). Again, the Scandinavian Peninsula shows extremely high PD values. The Mexican states of Veracruz and Oaxaca, as well as the Andes exhibit markedly high PD. PD is moderately high in southeastern Australia, Tasmania, New Zealand and north-east Papua New Guinea. PD is high in Japan and moderately high in Taiwan. PD is significantly higher than expected by chance only in New Zealand and Tasmania (Figure 2b).

Rarefied/extrapolated phylogenetic diversity (PDrare) was 97.3% correlated to raw estimates of PD, and peaks in the Kola Peninsula of Russia (Figure 2c). PDrare is overall substantially higher in the tropics, South Africa and Madagascar, and lower in well-sampled areas than raw PD.

MPD is highest throughout most of the boreal and arctic north (Figure 3a). In more temperate latitudes, MPD was high in eastern Europe, the east coast of the USA, southeastern Brazil and Uruguay, Colombia, Argentina, Central-Southern China and northern Japan. MPD is significantly higher than expected ($p > 0.99$; Figure 3b) in high latitudes across the globe, and also in the State of Rio de Janeiro (Brazil), Uruguay, central Chile, southern Venezuela and the Falkland Islands. Contrastingly, areas with significantly low MPD ($p < 0.01$) were found mostly in tropical, subtropical and temperate regions of the world.

Non-phylogenetic measures of endemism (WE and CWE; Figure S3.2) were similar to their phylogenetic analogues (PE and CWPE; Figure 4), with a few notable exceptions: PE and WE both peak in Japan, Taiwan, southern China, Southeast Asia, New Zealand, Tasmania, central Mexico, southeastern Brazil and the Andes (Figure 4a and Figure S3.2a). High WE and PE was also found in Australia, Papua New Guinea, the west coast of the U.S.A., Spain,

Austria, the Scandinavian Peninsula, the Western Cape region of South Africa and several islands including New Caledonia, the Azores, Réunion and Mauritius. WE was relatively lower in Tasmania and New Zealand than its phylogenetic counterpart. CWE and CWPE peak in southern and Southeast Asia, in particular in the Tamil Nadu state of South India, New Caledonia and Fiji (Figures 4b and Figure S2b). Other areas of striking endemism are mostly oceanic islands, including the Azores, the Juan Fernández Islands, Île Saint-Paul and the Seychelles.

The randomization results for all measures of endemism are similar (Figures 4c,d and Figure S2c,d). They indicate that, in general, endemism is significantly low in the Northern Hemisphere and significantly high in the Southern Hemisphere. Exceptions to this pattern include China, Japan and areas with Mediterranean climates in North America and Europe, which have significantly high endemism.

3.4 | Comparison between floristic and specimen datasets

We found a 96% correlation in SR between the Geffert dataset and the reduced Geffert dataset (Figure 5a and Figure S3.3a). The specimen data show a 67% correlation in SR to the original Geffert data (2013), with even well-collected and databased continents such as Europe being underrepresented in our dataset (Figure S3.3b). However, in some areas of South America (particularly the northern Andes) and Central America (particularly Mexico) the specimen data appear to match the original Geffert dataset well (Figure 5b and SI Figure S3.1). In others, SR is higher in the specimen data than in the original Geffert data. This trend is apparent in much of North America, particularly British Columbia (Canada) and Alaska (USA), the Scandinavian Peninsula and Spain (Figure 5 and Figure S3.3b).

Phylogenetic measures of alpha diversity tend to be more robust to the missing data than SR. PD calculated from the reduced Geffert dataset shows a 70% correlation to the specimen data (Figures S3.4a

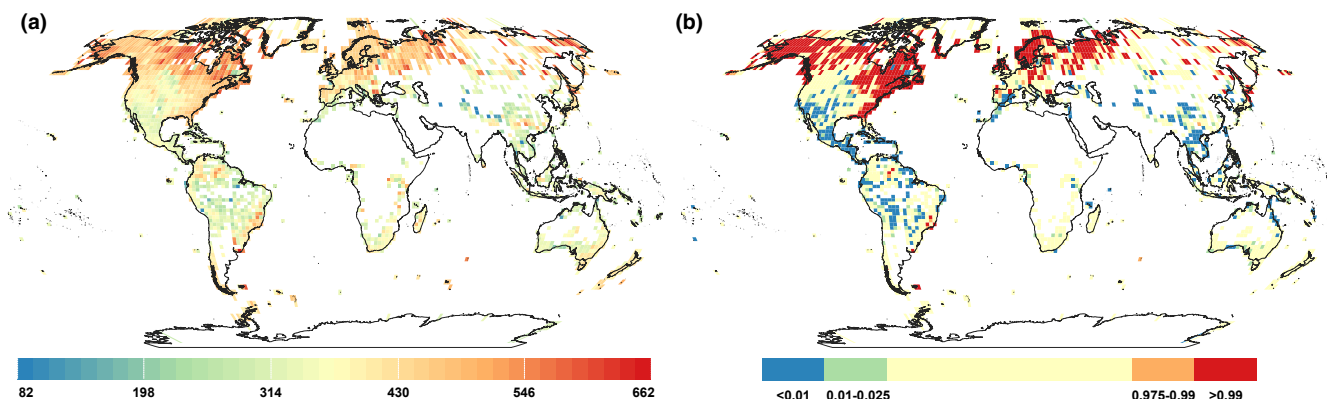


FIGURE 3 Global patterns of moss (a) mean phylogenetic distance among co-occurring taxa (MPD) and (b) significant departures of MPD from expectations of null models. (a) MPD peaks in the high latitudes in the Northern Hemisphere and is generally low-moderate elsewhere. (b) Many assemblages in the boreal-Arctic zone contain species that are significantly more distantly related to each other than expected from null models, whereas those in the tropical and subtropical zones contain species that are more closely related than expected

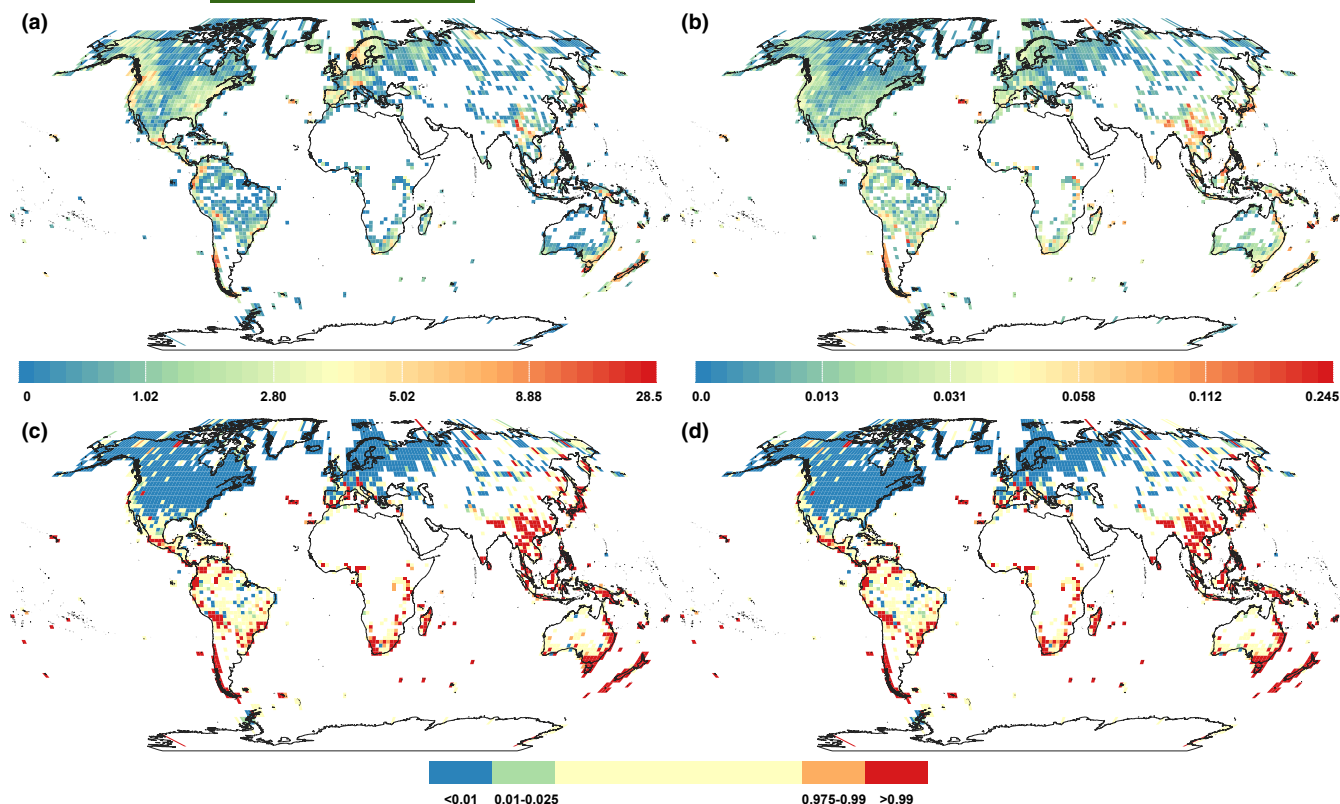


FIGURE 4 Global patterns of phylogenetically determined estimates of moss endemism. (a, b) depict richness-based phylogenetic endemism (PE) and per-species phylogenetic endemism (CWPE), respectively. (c-d) depict significant departures of these metrics from expectations of null models. (a) PE peaks in much of East and Southeast Asia, Oceania and parts of Central and South America, particularly the Andes. (b) CWPE shows much higher values in much of Asia and several oceanic islands than its taxonomic analogue. (c, d) Across the Southern Hemisphere, both richness-based endemism and per-species endemism are markedly higher than expected by chance, or no significant difference than a random expectation. Contrastingly, both measures of endemism are lower than expected according to null models in the cold northern climates of Europe and North America

and S3.5). Differences between these two datasets are apparent in North America, with British Columbia, Alaska and Mexico having higher PD according to the specimen data than in the floristic data. Contrastingly, most of the Middle East, as well as Central Asia and much of China and Mongolia have lower PD in the specimen data than in the floristic data. A similar pattern arises in Central and Northern Africa.

MPD calculated from the specimen data shows deficiencies in the same regions compared to the floristic data (Figure S3.6). The correlation in MPD between the floristic data and the specimen data is 76%, indicating MPD tends to be much more robust to bias between the two datasets. MPD calculated from the specimen data is noticeably higher than MPD calculated from the floristic data in Alaska (U.S.A.) and Canada. Interestingly, differences between the two datasets in terms of MPD were more evenly spread out across continents (Figure S3.4b).

3.5 | Biogeographic regionalizations

The optimal number of phyloregions recovered was $k = 6$ (Figure 6a). The six regions are a Holarctic/Holantarctic region (region 1); a

temperate region (region 2); a region composed of the eastern U.S.A and Mexico, southern China and the western coast of Peru and Ecuador (region 3); a region including Central America, the Caribbean and the South American tropics (region 4); a region comprising South Africa, Australia and New Zealand (region 5) and a region comprised of Southeast Asia and Oceania (region 6). Based on non-metric multidimensional scaling (NMDS) ordination, 3 regions of major similarity are clear: Holarctic/Holantarctic temperate region (regions 1–3), a pantropical region (regions 4 and 6) and a region composed of Australia, New Zealand and South Africa (region 5) (Figure S3.7). Region 5 is the most phylogenetically distinct in terms of Simpson's Index among all six regions.

The floristic approach parsed the data into 15 bioregions (Table S3.2; Figure 6b). The major bioregions identified included a Holarctic/Antarctic region (bioregion 3); an American tropical region with a few cells in Western Africa (bioregion 7); a region composed of central/southern Chile, Patagonia and the Falkland islands (bioregion 4); an East and South African region that includes Madagascar and small portions of subtropical Eastern Australia (bioregion 9); similarly, a region that includes parts of South Africa and central Australia (bioregion 10); a region encompassing New Zealand, Tasmania, the Eastern coast and the Southwestern tip of Australia (bioregion 1); a distinct

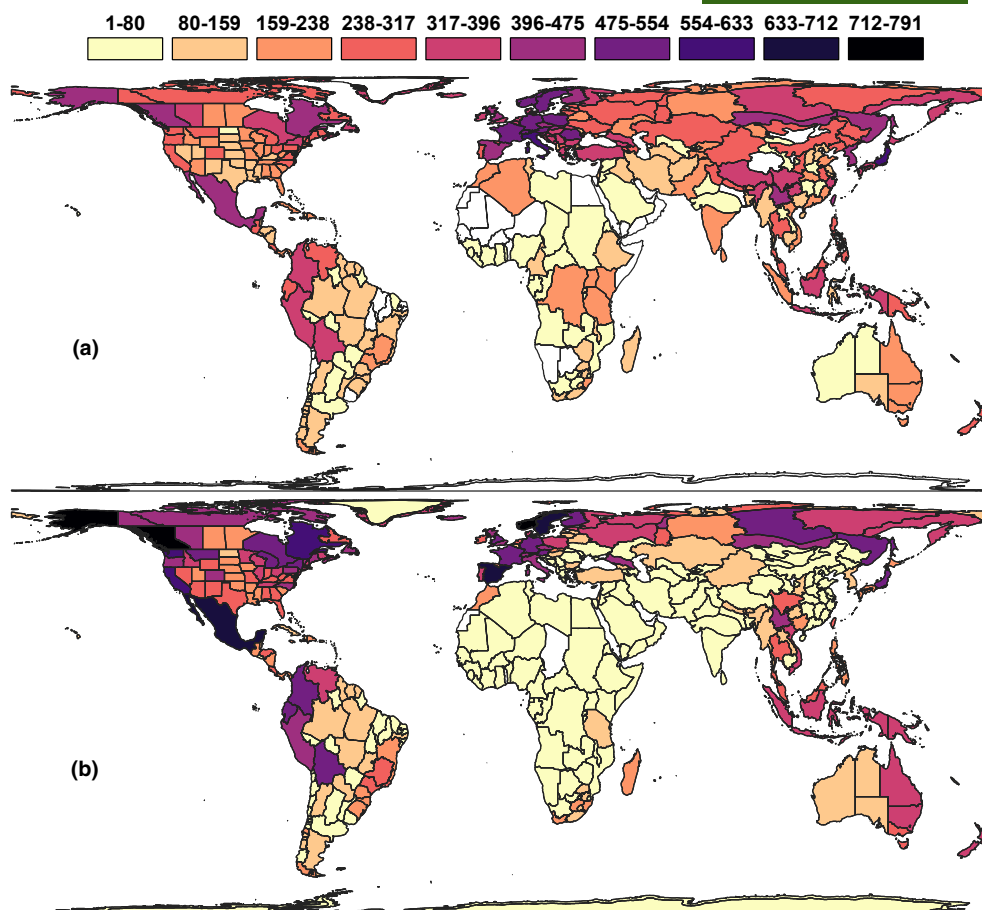


FIGURE 5 Comparisons of moss species richness between the (a) the reduced Geffert et al. (2013) dataset, and (b) the GBIF/iDigBio data. (a) There is a strong correlation (96%) between SR calculated originally by Geffert et al. (2013), and SR calculated using only those species that are included in our phylogeny. (b) The specimen data show a 67% correlation in SR to the original Geffert data (2013), with even well-collected and databased continents like Europe being underrepresented by herbarium data. The African and Asian tropics are particularly underrepresented in the specimen data

region composed of Southeast Asia, Southern India, Sri Lanka, Melanesia, Micronesia, Macaronesia and North Australia (bioregion 2); a Sino-Japanese region (bioregion 12) and a Himalayan region (bioregion 14). Two small regions composed of island archipelagos and characterized by a very distinct species endemic to these areas were found in Okinawa (bioregion 15) and the Canaries and Madeira.

4 | DISCUSSION

Our results provide evidence to support several important theories developed in moss biogeography, including the role of long-distance dispersal in shaping large-scale floristic patterns, the dominance of anagenesis in driving island diversity and the role of climatic instability in the assembly of the Holarctic. Through a comparison of the best available floristic and specimen-based datasets, we also demonstrate that phylogenetic metrics overcome some of the limitations of species-based estimates of richness, though substantial gaps remain in the global collection, digitization and sequencing of bryophyte specimens.

Comparisons between metrics derived from specimen data to those derived from floristic data are largely congruent. The main exception to this trend is in tropical regions across the globe, exemplifying a spatial bias in moss specimen databases. Several forms of bias can be introduced when incorporating specimen occurrence databases with phylogenetic data to examine global patterns of diversity. First, only a limited number of species have available phylogenetic data, which can result in both spatial and taxonomic biases as the result of differential sequencing efforts. Nevertheless, we found a strong correlation (96%) between SR calculated originally by Geffert et al. (2013), and SR calculated using only those species that are included in our phylogeny (Figure 5a and Figure S3.3a). These results suggest that the phylogenetic sampling does not introduce a strong spatial or taxonomic bias to observed patterns of diversity.

A more pernicious problem is that herbarium-based meta-analyses are subject to spatial bias in collecting, digitization and geo-referencing efforts (Beck et al., 2012; Boakes et al., 2010; Meineke & Daru, 2021; Spalink, Kriebel, et al., 2018; Spalink, Pender, et al., 2018; Yang et al., 2013), not to mention potential regional or taxonomic errors in species identification. Bias is manifest in our data, where

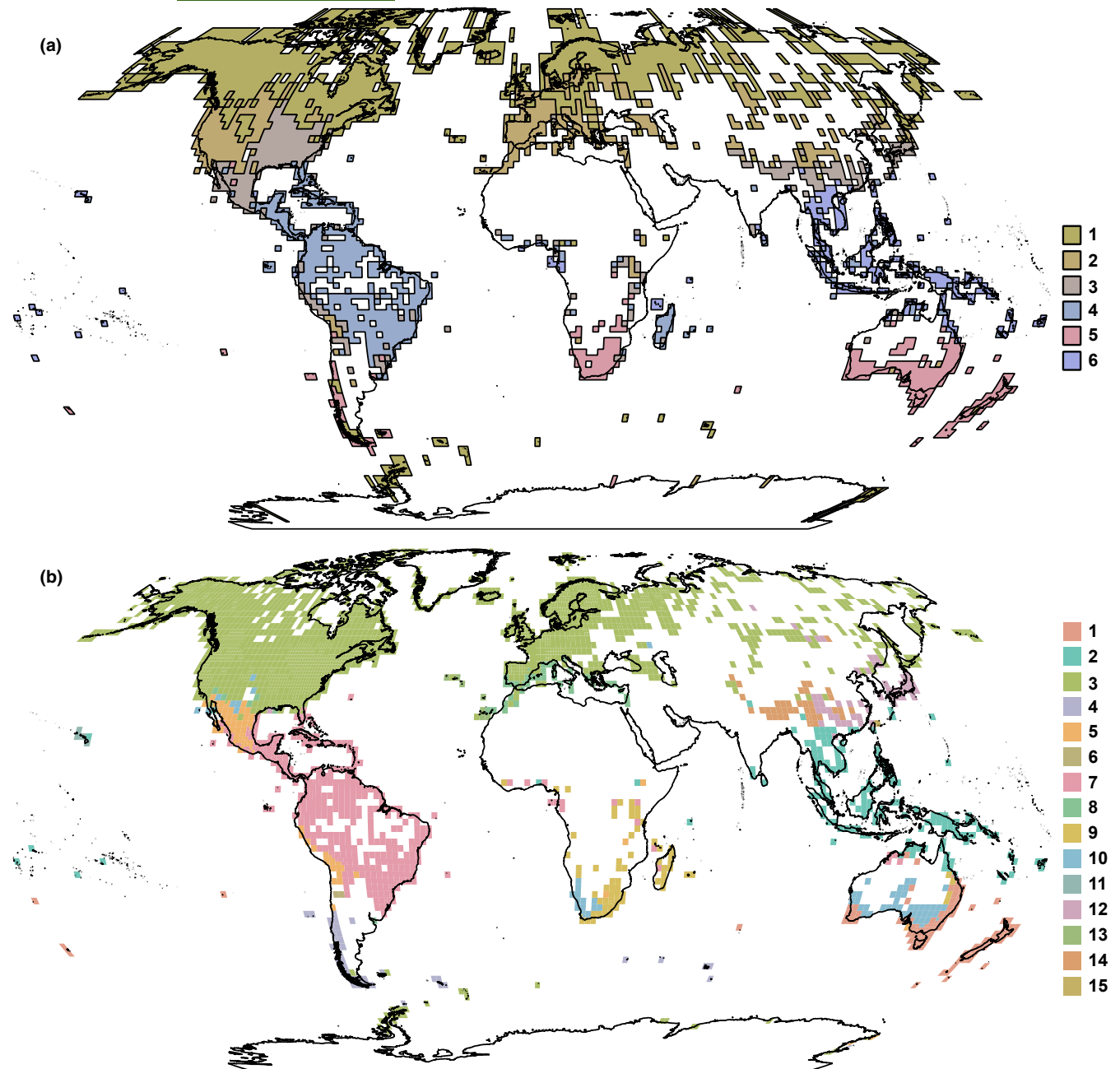


FIGURE 6 Moss biogeographic regionalizations. (a) depicts phyloregions based on optimized clustering of cells according to Simpson's phylogenetic index. Based on the NMDS ordination (Figure S3.7), 3 regions of major similarity are clear: A Holarctic/Holantartic temperate region (regions 1–3), a pantropical region (regions 4 and 6), and a region composed of Australia, New Zealand and South Africa (region 5). The colours amount to the difference in beta diversity (turnover) between different regions (i.e., similarity). (b) depicts floristically determined moss bioregions. Major floristic regions include a Holarctic/Holantartic region, an Amazonian region, a Southeast Asian region and a Sub-Saharan African region

sampling effort varies substantially across the globe (Figure 2d). In general, most collecting, digitization and sequencing efforts in mosses have occurred in North America, Europe, Japan, Australia and New Zealand. As a result, sampling effort is low in areas known to harbour many moss species. Indeed, the Amazon Basin, Eastern Africa, Madagascar, Southeast Asia, Macaronesia, Micronesia and Polynesia are all under sampled in our analyses (Geffert et al., 2013; Möls et al., 2013; Figures 5 and 2d). In the comparison between the

original data from Geffert et al. (2013) and the specimen data, we found only a 66% correlation in SR (Figure 5 and Figure S3.3b).

A key question is whether phylogenetic metrics can overcome some of this bias. While we examine a limited number of species in our analyses, these species are representative of nearly the full evolutionary diversity of mosses. Thus, measures of phylogenetic diversity should be expected to better reflect patterns across the globe (Mishler et al., 2014; Thornhill et al., 2016, 2017), and we found



that this is true to an extent. Comparisons of PD from the reduced Geffert and specimen datasets show a 70% correlation (Figures S3.4a and S3.5). MPD performs slightly better at 76% (Figures S3.4b and S3.6). Thus, while a benefit of phylogenetic estimates of diversity is that they can overcome, to some degree, errors resulting from missing data in floristically based calculations, we remain limited in the conclusions, we can draw about under sampled regions using currently available data.

4.1 | Island bryogeography

We expected to see significantly high PD and MPD on islands, as these should harbour a larger number of unique lineages due to the inverse isolation effect in mosses (Szövényi et al., 2012), yet, we only found significantly high PD in New Zealand and Tasmania. This pattern is somewhat puzzling, given the affinities between the bryofloras of Southeast Australia and New Zealand (Miller, 1982; Renner et al., 2017). Indeed, the floristic affinities between these two floras are corroborated by our data, which show they belong to the same bioregion and phyloregion (Figure 6). Dispersal between eastern Australia and New Zealand is largely unidirectional as a consequence of easterly moving weather systems (Sanmartín et al., 2007). Thus, the high PD in New Zealand could be expected because of repeated colonization of Australian lineages, followed by isolation and in-situ diversification.

While MPD is not higher *than expected* on most oceanic islands (except for the Falkland Islands), MPD is notably high on the French Southern and Antarctic Lands, Île Saint-Paul and the Chatham Islands (Figure 3). Several island systems with moderate values of MPD, such as the Hawai'ian archipelago, the Azores, Mauritius and Réunion, and much of Southeast Asia, also show moderate to high levels of CWE and PCWE (Figures 4b and Figure S3.2b) but low PD (Figures 2a,b) and SR (Figures 1a,b). Taken together, these results corroborate the dominance of anagenesis in driving patterns of island endemism in bryophytes (Patiño et al., 2014). Here, islands are colonized by lineages that span the moss phylogeny, leading to highly unique phylogenetic assemblages. Subsequently, these lineages do not undergo radiations in-situ, but evolve anagenetically into evolutionarily distinct, and thus, range-restricted species. Anagenesis may be driven by the maintenance of gene flow among islands through the inverse-isolation effect (Szövényi et al., 2012). Patiño et al. (2014) favour an alternative hypothesis; that a mixture of intrinsic and external factors shape this pattern, notably a preference for subtropical environments and long-term stability of these types of habitats. Subsequent research on the environmental correlates of diversity in mosses is needed to fully explain these patterns.

4.2 | Patterns of endemism

It is challenging to estimate endemism in lineages that are under collected. Missing data may underestimate species ranges and thereby

inflate rates of endemism. Conversely, narrowly distributed species tend to be rare and are less likely to be observed where they do occur, resulting in an underestimation of endemism. Further, it is thus far impossible to implement a standardization technique for endemism to make equitable comparisons across assemblages. Even with adequate sampling, noise resulting from taxonomic shortcomings could still bias the results.

Nevertheless, our specimen-based estimates of endemism (Figure 4a and Figure S3.2a) are largely congruent with the patterns of endemism reported by Vanderpoorten and Hallingbäck (2009) and Patiño and Vanderpoorten (2018), with some differences. Unlike these studies, we found high endemism in central Chile, southeastern Brazil, Veracruz and Oaxaca (Mexico), Japan and Southeast Asia, Tasmania and the southeastern tip of New South Wales and Victoria. Our results in Australia and Tasmania are largely consistent with the findings of Stevenson et al. (2013). Carter et al. (2016) identified major hotspots for moss endemism in northern California/Pacific Northwest and the southern Appalachians. Our results confirmed these patterns, and our analysis of MPD corroborates their hypothesis that in situ diversification may have played an important role in driving patterns of endemism in the North American west (Figure 3a).

Patterns of endemism are similar in liverworts and mosses (Patiño & Vanderpoorten, 2018; Stevenson et al., 2013; Vanderpoorten & Hallingbäck, 2009). However, liverworts show higher rates of endemism on small oceanic islands, indicating that barriers to dispersal are a more important driver of endemism in liverworts. It is difficult to compare mosses and liverworts to hornworts because of the dearth of studies that exist on patterns of endemism in the latter, but one study in Australia found that hotspots of endemism for hornworts was similar to that of other bryophytes (Stevenson et al., 2013). Contrastingly, India is a hotspot of endemism for the hornwort genera *Notothylas* and *Folioceros* (Villarreal et al., 2014), but is not a hotspot of endemism in liverworts. Richness-based endemism in India for mosses is low, while per-species richness in India is high (Figures 4a,b and Figures S3.2a,b). Buck and Thiers (1989) state that India boasts the richest bryoflora in Asia. Because we have no data for most of India, we believe that the low levels of richness-based endemism paired with high per-species endemism in this region is likely an artefact. Better sampling, databasing and sequencing efforts are needed across Southeast Asia, especially India, to ascertain whether India is an endemism hotspot for mosses.

4.3 | The mossy north

The boreal-arctic Northern Hemisphere is characterized by significantly high MPD, suggesting the species in these regions tend to be distantly related and a product of repeated colonization (Figure 3b). This result, combined with the finding that endemism is significantly low in these regions ($p < 0.01$; Figures 4c,d), indicates that these assemblages are composed of widespread

species that readily track suitable habitat following repeated glaciations. This finding is corroborated by the fact that species tend to be relatively widespread across the Northern Hemisphere (Frahm & Vitt, 1993). Several studies have reported rapid post-glacial recolonization of moss species, and there are numerous routes by which moss species could have recolonized northern areas after the LGM (Ellis & Tallis, 2000; Jonsgard & Birks, 1995; Kyrkjeeide et al., 2014). These patterns are particularly evident in the Scandinavian Peninsula, which has been previously shown to have the highest number of *Sphagnum* species in Europe (Séneca & Söderström, 2008). This richness is certainly related to the abundance of suitable wetland habitats, but possibly also to the long history of bryological research in the region.

4.4 | Biogeographic regionalizations

Phylogenetic reconstructions indicate that liverworts split into well-supported Laurasian and Gondwanan clades, and also appear to segregate at Wallace's line (Vanderpoorten et al., 2010). Further, the same analysis found another clade that includes Sub-Saharan Africa and South America. Together, these results indicate that vicariance has played an important role in shaping liverwort biogeographic structure at a global scale. Our results indicate that mosses, unlike liverworts, do not segregate into Gondwanan and Laurasian clades, but instead tend to segregate into tropical and temperate groupings, with the Holarctic and Holantarctic being dominated by similar phylogenetic assemblages (Phyloregions 1 & 2, Figure 6a and Figure S3.7; Bioregion 3, Figure 6b). This finding is demonstrated by the results of the NMDS clustering analysis, wherein phyloregions 1 and 2 are ordinated close together (Figure S3.7).

Our results indicate that long-distance dispersal of diaspores obscures signals of vicariance and ecological filtering plays a major role in shaping moss distributions. Floristically, moss diversity is well-parsed into many regions with unique assemblages (Figure 6b). This pattern is particularly evident in the tropics. Hedenäs (2007) found that, at least among the pleurocarpus mosses, the Asian, African and American tropics segregate into well-differentiated groups based on the analysis of floristic beta diversity. Previous analyses have also revealed greater differences in beta diversity between, than within, tropical regions (Norhazrina et al., 2017), as well as higher beta diversity within the tropics than outside them (Hedenäs, 2007). However, we report here that these regions are phylogenetically similar, suggesting that the high beta diversity in these regions is the result of species turnover, not lineage turnover. This result would be expected if these lineages have dispersed across the tropics many times (Figure 6a and Figure S3.7).

4.5 | Future research

We have demonstrated that there is significant bias in collecting, sequencing and databasing efforts for mosses. Like previous studies,

we were unable to find any significant relationship between latitude and species richness (Geffert et al., 2013; Möls et al., 2013; Shaw et al., 2005). This lack of an alpha diversity gradient has been questioned, particularly after Wang et al. (2017) showed a strong latitudinal diversity gradient in liverworts and hornworts. Further, it has already been demonstrated that, in pleurocarpus mosses, beta diversity is strongly latitudinally stratified, with the tropics, subtropics and Holantarctic regions having much higher beta diversity than the boreal region (Hedenäs, 2007). The main impetus for the criticism of the "mosses are everywhere" hypothesis stems from extreme taxonomic and distributional shortcomings for tropical mosses (Patiño & Vanderpoorten, 2018; Wang et al., 2017). Our results further corroborate these shortcomings. We stress that our results, while corroborating those of previous studies, remain incomplete. Until a global checklist of moss names and distributions comparable to the one available for liverworts and hornworts (Soderstrom et al., 2016) is constructed, the biogeography, diversity and evolutionary ecology of tropical mosses—and hence, the latitudinal diversity gradient—will remain obscure.

ACKNOWLEDGEMENTS

We thank R. P. Dorn, D. Kruse, L. Morley and K. M. Simpson for generative discussions that improved this manuscript. We thank J. L. Geffert for generously providing his datasets. This work was supported in part by a Frank W. Gould fellowship to KS and by the USDA National Institute of Food and Agriculture (McIntire Stennis Project 1018692) and NSF #1902064 to DS. No permits were needed to carry out this work.

CONFLICT OF INTEREST

The authors have no conflict of interest.

DATA AVAILABILITY STATEMENT

All data used in this study are available for download from Dryad (doi:10.5061/dryad.905qfttm2). DOI's for the datasets downloaded from GBIF are available as a derived dataset (<https://doi.org/10.15468/dd.hpjykh>). All R scripts are available on GitHub: [ksanb/GlobalMossDiversity](https://github.com/ksanb/GlobalMossDiversity)

ORCID

Katie Kobara Sanbonmatsu  <https://orcid.org/0000-0003-0680-2592>

REFERENCES

- Abadi, S., Azouri, D., Pupko, T., & Mayrose, I. (2019). Model selection may not be a mandatory step for phylogeny reconstruction. *Nature Communications*, 10(1), 934. <https://doi.org/10.1038/s41467-019-08822-w>
- Ah-Peng, C., Cardoso, A. W., Flores, O., West, A., Wilding, N., Strasberg, D., & Hedderson, T. A. J. (2017). The role of epiphytic bryophytes in interception, storage, and the regulated release of atmospheric moisture in a tropical montane cloud forest. *Journal of Hydrology*, 548, 665–673. <https://doi.org/10.1016/j.jhydrol.2017.03.043>
- Ah-Peng, C., Wilding, N., Kluge, J., Descamps-Julien, B., Bardat, J., Chuah-Petiot, M., ... Hedderson, T. A. J. (2012). Bryophyte diversity and range size distribution along two altitudinal gradients: Continent



- vs. Island. *Acta Oecologica*, 42, 58–65. <https://doi.org/10.1016/j.actao.2012.04.010>
- Barbé, M., Fenton, N. J., Bergeron, Y., & Veski, P. (2016). So close and yet so far away: Long-distance dispersal events govern bryophyte metacommunity reassembly. *Journal of Ecology*, 104(6), 1707–1719. <https://doi.org/10.1111/1365-2745.12637>
- Beck, J., Ballesteros-Mejia, L., Buchmann, C. M., Dengler, J., Fritz, S. A., Gruber, B., Hof, C., Jansen, F., Knapp, S., Kreft, H., Schneider, A.-K., Winter, M., Dormann, C. F. (2012). What's on the horizon for macroecology? *Ecography*, 35(8), 673–683. <http://dx.doi.org/10.1111/j.1600-0587.2012.07364.x>
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, 41(D1), D36–D42.
- Boakes, E. H., McGowan, P. J., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K., & Mace, G. M. (2010). Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biology*, 8(6), e1000385. <https://doi.org/10.1371/journal.pbio.1000385>
- Buck, W. R., & Thiers, B. M. (1989). Review of bryological studies in the tropics. In D. G. Campbell & H. D. Hammond (Eds.), *Floristic inventory of tropical countries* (pp. 485–493). New York Botanical Garden.
- Câmara, P., Soares, A. E., Henriques, D. K., Peralta, D. F., Bordin, J., Carvalho-Silva, M., & Stech, M. (2019). New insights into the species diversity of *Bartramia* Hedw. (Bryophyta) in Antarctica. *Antarctic Science*, 31(4), 208–215.
- Carter, B. E., Shaw, B., & Shaw, A. J. (2016). Endemism in the moss flora of North America. *American Journal of Botany*, 103(4), 769–779. <https://doi.org/10.3732/ajb.1500484>
- Chamberlain, S., Barve, V., McGlinn, D., Oldoni, D., Desmet, P., Geffert, L., & Ram, K. (2021). Rgbif: Interface to the global biodiversity information facility API. R package version. 3.5.2. <https://CRAN.R-project.org/package=rgbif>
- Chamberlain, S., & Szöcs, E. (2013). taxize: taxonomic search and retrieval in R. *F1000Research*, 2.
- Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, 11(4), 265–270.
- Chao, A., Chiu, C. H., Hsieh, T. C., Davis, T., Nipperess, D. A., Faith, D. P., & O'Hara, R. B. (2015). Rarefaction and extrapolation of phylogenetic diversity. *Methods in Ecology and Evolution*, 6(4), 380–388. <https://doi.org/10.1111/2041-210x.12247>
- Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, 84(1), 45–67.
- Clymo, R. S. (1987). The ecology of peatlands. *Science Progress*, 71(4), 593–614.
- Colwell, R. K., Chao, A., Gotelli, N. J., Lin, S. Y., Mao, C. X., Chazdon, R. L., & Longino, J. T. (2012). Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology*, 5(1), 3–21. <https://doi.org/10.1093/jpe/rtr044>
- Crisp, M. D., Laffan, S., Linder, H. P., & Monro, A. (2001). Endemism in the Australian Flora. *Journal of Biogeography*, 28(2), 183–198.
- Daru, B. H., Karunarathne, P., Schliep, K., & Silvestro, D. (2020). Phylregion: R package for biogeographical regionalization and macroecology. *Methods in Ecology and Evolution*, 11(11), 1483–1491. <https://doi.org/10.1111/2041-210x.13478>
- Edler, D., Guedes, T., Zizka, A., Rosvall, M., & Antonelli, A. (2017). Infomap bioregions: Interactive mapping of biogeographical regions from species distributions. *Systematic Biology*, 66(2), 197–204. <https://doi.org/10.1093/sysbio/syw087>
- Ellis, C. J., & Tallis, J. H. (2000). Climatic control of blanket mire development at Kentra Moss North-West Scotland. *Journal of Ecology*, 88(5), 869–889.
- Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61(1), 1–10.
- Frahm, J. P., & Vitt, D. H. (1993). Comparisons between the moss floras of North America and Europe. *Nova Hedwigia*, 56(3–4), 307–333.
- Geffert, J. L., Frahm, J.-P., Barthlott, W., & Mutke, J. (2013). Global moss diversity: Spatial and taxonomic patterns of species richness. *Journal of Bryology*, 35(1), 1–11. <https://doi.org/10.1179/1743282012y.00000000038>
- Givnish, T. J. (2010). Ecology of plant speciation. *Taxon*, 59(5), 1326–1366.
- Goffinet, B., & Buck, W. R. (2020). Classification of the Bryophyta. Online version available at <http://bryology.uconn.edu/classification/>. Checked on July 2020.
- Goffinet, B., & Shaw, A. J. (2009). *Bryophyte biology*. Cambridge University Press.
- Hedenäs, L. (2007). Global diversity patterns among Pleurocarpous mosses. *The Bryologist*, 110(2), 319–331.
- Herzog, T. (1926). *Geographie der Moose*. Gustav Fischer.
- Hsieh, T. C., & Chao, A. (2017). Rarefaction and extrapolation: Making fair comparison of abundance-sensitive phylogenetic diversity among multiple assemblages. *Systematic Biology*, 66(1), 100–111. <https://doi.org/10.1093/sysbio/syw073>
- Jonsgard, B., & Birks, H. H. (1995). Late-glacial mosses and environmental reconstructions at Kråkenes Western Norway. *Lindbergia*, 20(2), 64–82.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780.
- Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26, (11), 1463–1464. <http://dx.doi.org/10.1093/bioinformatics/btq166>
- Kyrkjeeide, M. O., Stenøien, H. K., Flatberg, K. I., & Hassel, K. (2014). Glacial refugia and post-glacial colonization patterns in European bryophytes. *Lindbergia*, 2, 47–59. <https://doi.org/10.25227/linbg.01046>
- Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22), 3276–3278.
- Ledent, A., Gauthier, J., Pereira, M., Overton, R., Laenen, B., Mardulyn, P., Gradstein, S. R., Haan, M., Ballings, P., Van der Beeten, I., Zartman, C., Vanderpoorten, A. (2020). What do tropical cryptogams reveal? Strong genetic structure in Amazonian bryophytes. *New Phytologist*, 228(2), 640–650. <http://dx.doi.org/10.1111/nph.16720>
- Lönnell, N., Hylander, K., Jonsson, B. G., & Sundberg, S. (2012). The fate of the missing spores – Patterns of realized dispersal beyond the closest vicinity of a sporulating Moss. *PLoS ONE*, 7(7), e41987. <https://doi.org/10.1371/journal.pone.0041987>
- Mateo, R. G., Broennimann, O., Normand, S., Petitpierre, B., Araújo, M. B., Svenning, J.-C., Baselga, A., Fernández-González, F., Gómez-Rubio, V., Muñoz, J., Suarez, G. M., Luoto, M., Guisan, A., Vanderpoorten, A. (2016). The mossy north: an inverse latitudinal diversity gradient in European bryophytes. *Scientific Reports*, 6(1). <http://dx.doi.org/10.1038/srep25546>
- Medina, R., Johnson, M., Liu, Y., Wilding, N., Hedderson, T. A., Wickett, N., & Goffinet, B. (2018). Evolutionary dynamism in bryophytes: Phylogenomic inferences confirm rapid radiation in the moss family Funariaceae. *Molecular Phylogenetics and Evolution*, 120, 240–247. <https://doi.org/10.1016/j.ympev.2017.12.002>
- Meineke, E. K., & Daru, B. H. (2021). Bias assessments to expand research harnessing biological collections. *Trends in Ecology & Evolution*, 36(12), 1071–1082. <https://doi.org/10.1016/j.tree.2021.08.003>
- Michonneau, F., & Collins, M. (2017). Ridigbio: Interface to the iDigBio data API. R package version 0.3.5. <https://CRAN.R-project.org/package=ridigbio>
- Miller, H. A. (1982). Bryophyte evolution and geography. *Biological Journal of the Linnean Society*, 18, 145–196.

- Miller, M., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES science gateway for inference of large phylogenetic trees. Paper presented at the 2010 gateway computing environments workshop (GCE).
- Mishler, B. D., Knerr, N., Gonzalez-Orozco, C. E., Thornhill, A. H., Laffan, S. W., & Miller, J. T. (2014). Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian acacia. *Nature Communications*, 5, 4473. <https://doi.org/10.1038/ncomms5473>
- Möls, T., Vellak, K., Vellak, A., & Ingerpuu, N. (2013). Global gradients in moss and vascular plant diversity. *Biodiversity and Conservation*, 22(6–7), 1537–1551. <https://doi.org/10.1007/s10531-013-0492-6>
- Morris, J. L., Puttick M. N., Clark J. W., Edwards D., Kenrick P., Pressel S., Wellman C. H., Yang Z., Schneider H., Donoghue P. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences*, 115(10), E2274–E2283. <http://dx.doi.org/10.1073/pnas.1719588115>
- Newton, A. E., Wikström, N., Bell, N., Forrest, L. L., & Ignatov, M. S. (2007). Dating the diversification of Pleurocarpous mosses: Systematics and evolution. In A. E. Newton & R. S. Tangney (Eds.), *Pleurocarpous mosses: Systematics and evolution* (pp. 337–366). CRC Press.
- Norhazrina N., Wang J., Hagborg A., Geffert J. L., Mutke J., Gradstein S. R., Basella A., Vanderpoorten A., Patiño J. (2016). Tropical bryophyte floras: a homogeneous assemblage of highly mobile species? Insights from their spatial patterns of beta diversity. *Botanical Journal of the Linnean Society*. <http://dx.doi.org/10.1111/boj.12495>
- Patiño, J., Carine, M., Fernandez-Palacios, J. M., Otto, R., Schaefer, H., & Vanderpoorten, A. (2014). The anagenetic world of spore-producing land plants. *New Phytologist*, 201(1), 305–311. <https://doi.org/10.1111/nph.12480>
- Patiño, J., & Vanderpoorten, A. (2018). Bryophyte biogeography. *Critical Reviews in Plant Sciences*, 37(2–3), 175–209. <https://doi.org/10.1080/07352689.2018.1482444>
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). *Ecological niches and geographic distributions (MPB-49)*. Princeton University Press.
- Renner, M. A. M., Heslewood, M. M., Patzak, S. D. F., Schafer-Verwimp, A., & Heinrichs, J. (2017). By how much do we underestimate species diversity of liverworts using morphological evidence? An example from Australasian *Plagiochila* (Plagiochilaceae: Jungermanniopsida). *Molecular Phylogenetics and Evolution*, 107, 576–593. <https://doi.org/10.1016/j.ympev.2016.12.018>
- Rose, J. P., Kriebel, R., & Sytsma, K. J. (2016). Shape analysis of moss (Bryophyta) sporophytes: Insights into land plant evolution. *American Journal of Botany*, 103(4), 652–662. <https://doi.org/10.3732/ajb.1500394>
- Salvador, S., & Chan, P. (2004). Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. Paper presented at the 16th IEEE international conference on tools with artificial intelligence.
- Sanders, H. L. (1968). Marine benthic diversity: A comparative study. *The American Naturalist*, 102(925), 243–282.
- Sanmartín, I., Wanntorp, L., & Winkworth, R. C. (2007). West wind drift revisited: Testing for directional dispersal in the southern hemisphere using event-based tree fitting. *Journal of Biogeography*, 34(3), 398–416.
- Schofield, W. B. (1992). Bryophyte distribution patterns. In J. W. Bates & A. M. Farmer (Eds.), *Bryophytes and lichens in a changing environment* (pp. 101–130). Clarendon Press.
- Schuster, R. (1983). Phytogeography of the bryophyte. *New Manual of Bryology*, 1, 463–626.
- Séneca, A., & Söderström, L. (2008). Species richness and distribution ranges of European sphagnum. *Folia Cryptogamica Estonica*, 44, 125–130.
- Shaw, A. J., Cox, C. J., & Goffinet, B. (2005). Global patterns of moss diversity: Taxonomic and molecular inferences. *Taxon*, 54(2), 337–352. <https://doi.org/10.2307/25065362>
- Smith, S. A., & O'Meara, B. C. (2012). treePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, 28(20), 2689–2690. <https://doi.org/10.1093/bioinformatics/bts492>
- Snäll, T., Fogelqvist, J., Ribeiro, P. J., Jr., & Lascoux, M. (2004). Spatial genetic structure in two congeneric epiphytes with different dispersal strategies analysed by three different methods. *Molecular Ecology*, 13(8), 2109–2119. <https://doi.org/10.1111/j.1365-294X.2004.02217.x>
- Söderström, L., Hagborg, A., von Konrat, M., Bartholomew-Began, S., Bell, D., Briscoe, L., Brown, E., Cargill, D. C., da Costa, D. P., Crandall-Stotler, B. J., Cooper, E., Dauphin, G., Engel, J., Feldberg, K., Glenn, D., Gradstein, S. R., He, X., Heinrichs, J., Hentschel, J., ... Zhu, R. (2016). World checklist of hornworts and liverworts. *PhytoKeys*, 59, 1–828. <http://dx.doi.org/10.3897/phytokeys.59.6261>
- Spalink, D., Kriebel, R., Li, P., Pace, M. C., Drew, B. T., Zaborsky, J. G., Rose, J., Drummond, C. P., Feist, M. A., Alverson, W. S., Waller, D. M., Cameron, K. M., Givnish, T. J., Sytsma, K. J. (2018). Spatial phylogenetics reveals evolutionary constraints on the assembly of a large regional flora. *American Journal of Botany*, 105(11), 1938–1950. <http://dx.doi.org/10.1002/ajb.2.1191>
- Spalink, D., Pender, J., Escudero, M., Hipp, A. L., Roalson, E. H., Starr, J. R., Waterway, M. J., Bohs, L., Sytsma, K. J. (2018). The spatial structure of phylogenetic and functional diversity in the United States and Canada: An example using the sedge family (Cyperaceae). *Journal of Systematics and Evolution*, 56(5), 449–465. <http://dx.doi.org/10.1111/jse.12423>
- Stamatakis, A. (2014). RAXML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313.
- Stevenson, L. A., González-Orozco, C. E., Knerr, N., Cargill, D. C., & Miller, J. T. (2013). Species richness and endemism of Australian bryophytes. *Journal of Bryology*, 34(2), 101–107. <https://doi.org/10.1179/1743282012y.0000000004>
- Sundberg, S. (2005). Larger capsules enhance short-range spore dispersal in sphagnum, but what happens further away? *Oikos*, 108(1), 115–124.
- Sundberg, S. (2013). Spore rain in relation to regional sources and beyond. *Ecography*, 36(3), 364–373. <https://doi.org/10.1111/j.1600-0587.2012.07664.x>
- Szővényi, P., Sundberg, S., & Shaw, A. J. (2012). Long-distance dispersal and genetic structure of natural populations: An assessment of the inverse isolation hypothesis in peat mosses. *Molecular Ecology*, 21(22), 5461–5472. <https://doi.org/10.1111/mec.12055>
- Thornhill, A. H., Baldwin, B. G., Freyman, W. A., Nosratinia, S., Kling, M. M., Morueta-Holme, N., Madsen, T. P., Ackerly, D. D., & Mishler, B. D. (2017). Spatial phylogenetics of the native California flora. *BMC Biology*, 15(1), 96. <https://doi.org/10.1186/s12915-017-0435-x>
- Thornhill, A. H., Mishler, B. D., Knerr, N. J., González-Orozco, C. E., Costion, C. M., Crayn, D. M., Laffan, S. W., & Miller, J. T. (2016). Continental-scale spatial phylogenetics of Australian angiosperms provides insights into ecology, evolution and conservation. *Journal of Biogeography*, 43(11), 2085–2098. <https://doi.org/10.1111/jbi.12797>
- Van Rooy, J., & Van Wyk, A. E. (2013). The bryofloristic regions of southern Africa. *Journal of Bryology*, 32(2), 80–91. <https://doi.org/10.1179/037366810x12578498136039>
- Vanderpoorten, A., Gradstein, S. R., Carine, M. A., & Devos, N. (2010). The ghosts of Gondwana and Laurasia in modern liverwort distributions. *Biological Reviews of the Cambridge Philosophical Society*, 85(3), 471–487. <https://doi.org/10.1111/j.1469-185X.2009.00111.x>
- Vanderpoorten, A., & Hallingbäck, T. (2009). Conservation biology. In B. Goffinet & J. Shaw (Eds.), *Bryophyte biology* (pp. 487–533). Cambridge University Press.
- Vanderpoorten, A., Patiño, J., Désambré, A., Laenen, B., Górski, P., Papp, B., Holá, E., Korpelainen, H., Hardy, O., & Rees, M.



- (2019). To what extent are bryophytes efficient dispersers? *Journal of Ecology*, 107(5), 2149–2154. <https://doi.org/10.1111/1365-2745.13161>
- Villarreal, J. C., Cargill, D. C., Hagborg, A., Soderstrom, L., & Renzaglia, K. S. (2014). A synthesis of hornwort diversity: Patterns, causes and future work. *Phytotaxa*, 9(1), 150–164. <https://doi.org/10.11646/phytotaxa.9.1.8>
- Villarreal, J. C., & Renner, S. S. (2014). A review of molecular-clock calibrations and substitution rates in liverworts, mosses, and hornworts, and a timeframe for a taxonomically cleaned-up genus *Nothoceros*. *Molecular Phylogenetics and Evolution*, 78, 25–35. <https://doi.org/10.1016/j.ympev.2014.04.014>
- Wang, J., Vanderpoorten, A., Hagborg, A., Goffinet, B., Laenen, B., & Patiño, J. (2017). Evidence for a latitudinal diversity gradient in liverworts and hornworts. *Journal of Biogeography*, 44(3), 487–488. <https://doi.org/10.1111/jbi.12909>
- Winter, D. J. (2017). rentrez: An R package for the NCBI eUtils API. 9, 520–526.
- Yang, W., Ma, K., Kreft, H., & Daniel Kissling, W. (2013). Geographical sampling bias in a large distributional database and its effects on species richness-environment models. *Journal of Biogeography*, 40(8), 1415–1426. <https://doi.org/10.1111/jbi.12108>
- Zhang, Y., Qian, L., Spalink, D., Sun, L., Chen, J., & Sun, H. (2021). Spatial phylogenetics of two topographic extremes of the Hengduan Mountains in southwestern China and its implications for biodiversity conservation. *Plant Diversity*, 43(3), 181–191. <https://doi.org/10.1016/j.pld.2020.09.001>
- Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., Antonelli, A., & Quental, T. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, 10(5), 744–751. <https://doi.org/10.1111/2041-210x.13152>

BIOSKETCH

Katie K. Sanbonmatsu is broadly interested in the evolution, diversification and historical biogeography of bryophytes. This work represents a component of her PhD thesis in the Department of Ecology and Conservation Biology at Texas A&M University.

Daniel Spalink is an Assistant Professor of Plant Systematics in the Department of Ecology and Conservation Biology at Texas A&M University and Director of the S.M. Tracy Herbarium (TAES). Spalink's principal interests include linking biogeography, morphology and ecology with phylogenetics to illuminate processes of diversification in plants, particularly within monocots.

Author contributions: D.S. and K.K.S. conceived the project. Data were generated by K.K.S. and D.S. All analyses were conducted by K.K.S., who also wrote the manuscript with contributions from D.S.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Sanbonmatsu, K. K., & Spalink, D. (2022). A global analysis of mosses reveals low phylogenetic endemism and highlights the importance of long-distance dispersal. *Journal of Biogeography*, 49, 654–667. <https://doi.org/10.1111/jbi.14333>