

# The Bacterial Communities of Alaskan Mosses and Their Contributions to N<sub>2</sub>-Fixation

Hannah Holland-Moritz (✉ [hannah.hollandmoritz@colorado.edu](mailto:hannah.hollandmoritz@colorado.edu))

University of Colorado, Boulder <https://orcid.org/0000-0002-0854-872X>

Julia E. M. Stuart

Northern Arizona University

Lily R. Lewis

University of Florida

Samantha N. Miller

Northern Arizona University

Michelle C. Mack

Northern Arizona University

Jose Miguel Ponciano

University of Florida

Stuart F. McDaniel

University of Florida

Noah Fierer

University of Colorado Boulder

---

## Research

**Keywords:** Bryophytes, plant microbiome, phyllosphere, nitrogen fixation

**Posted Date:** September 23rd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-80075/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Microbiome on February 23rd, 2021. See the published version at <https://doi.org/10.1186/s40168-021-01001-4>.

RESEARCH

Title Page

# The bacterial communities of Alaskan mosses and their contributions to N<sub>2</sub>-fixation

Hannah E. Holland-Moritz<sup>1\*</sup>, Julia E. M. Stuart<sup>2</sup>, Lily R. Lewis<sup>3</sup>, Samantha N. Miller<sup>2</sup>, Michelle C. Mack<sup>2</sup>, Jose Miguel Ponciano<sup>4</sup>, Stuart F. McDaniel<sup>4</sup> and Noah Fierer<sup>1,5</sup>

## Author Affiliations and Email Addresses

**Hannah E. Holland-Moritz:** Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Boulder CO, USA, hannah.hollandmoritz@colorado.edu

**Julia E. M. Stuart:** Center for Ecosystem Science and Society and the Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA, jms2435@nau.edu

**Lily R. Lewis:** Provost's Office, University of Florida, Gainesville, FL, USA, lilyrlewis@ufl.edu

**Samantha N. Miller:** Center for Ecosystem Science and Society and the Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA, Samantha.Miller@nau.edu

**Michelle C. Mack:** Center for Ecosystem Science and Society and the Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA, Michelle.Mack@nau.edu

**Jose Miguel Ponciano:** Department of Biology, University of Florida, Gainesville, FL, USA, josemi@ufl.edu

**Stuart F. McDaniel:** Department of Biology, University of Florida, Gainesville, FL, USA, stuartmcdaniel@ufl.edu

**Noah Fierer:** Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Boulder CO, USA, Cooperative Institute for Environmental Sciences, University of Colorado, Boulder, Boulder CO, USA, noah.fierer@colorado.edu

\*Corresponding Author: Hannah E. Holland-Moritz

\*Correspondence: hannah.hollandmoritz@colorado.edu

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, USA

Full list of author information is available at the end of the article

## Abstract

**Background:** Mosses in high latitude ecosystems harbor diverse bacterial taxa, including N<sub>2</sub>-fixers which are key contributors to nitrogen dynamics in these systems. Yet, the relative importance of moss host species, and environmental factors, in structuring these microbial communities and their N<sub>2</sub>-fixing potential remains unclear. We studied 26 boreal and tundra moss species across 24 sites in Alaska, USA from 61° to 69° N. We used cultivation-independent approaches to characterize the variation in moss-associated bacterial communities as a function of host species identity and site characteristics. We also measured N<sub>2</sub>-fixation rates via <sup>15</sup>N<sub>2</sub> isotopic enrichment and identified potential N<sub>2</sub>-fixing bacteria using available literature and genomic information.

**Results:** Host species identity and host evolutionary history were both highly predictive of moss microbiome composition, highlighting strong phylogenetic coherence in these microbial communities. Although less important, light availability and temperature also influenced composition of the moss microbiome. Finally, we identified putative N<sub>2</sub>-fixing bacteria specific to some moss hosts, including potential N<sub>2</sub>-fixing bacteria outside well-studied cyanobacterial clades.

**Conclusions:** The strong effect of host identity on moss-associated bacterial communities demonstrates mosses' utility for understanding plant-microbe interactions in non-leguminous systems. Our work also highlights the likely importance of novel bacterial taxa to N<sub>2</sub>-fixation in high-latitude ecosystems.

**Keywords:** Bryophytes; plant microbiome; phyllosphere; nitrogen fixation

## Background

Mosses are ubiquitous in northern ecosystems, often forming the dominant ground cover in boreal forests, peatlands, and Arctic tundra [1]. Mosses play critical ecological roles in these high-latitude environments as they insulate soils, maintain high soil moisture, and mediate carbon (C) and nitrogen (N) cycles by fixing and storing C and N [2]. Moss growth typically represents 20-50% of net primary productivity [2], and mosses are an important source of biologically fixed N in high-latitude ecosystems [3, 4]. Mosses also typically have high N-use efficiencies and low decomposition rates that promote C and N accumulation in many high-latitude ecosystems [5, 6, 2].

Like all plants, mosses associate with microbes and these microbial communities have key roles in multiple ecosystem-level processes. Moss-associated microbes fix N<sub>2</sub> [3, 4, 7], oxidize methane [8, 9], and contribute to the decomposition of organic matter in moss-dominated tundra and peatlands [10]. For these reasons, there has long been an interest in understanding what microbes associate with mosses, their contributions to ecosystem processes, and the factors that structure moss-associated microbial communities (e.g. [11, 4, 12, 13]).

Moss microbiomes are characterized by a core community of bacteria that are typically abundant in many common boreal and tundra mosses [12, 14]. The moss microbiome is often highly host specific, with different moss species harboring distinct bacterial communities [15, 14]. However, host identity is not the sole factor determining moss microbiome composition as there is clearly some degree of variation in the moss microbiome across a given species, variation that is presumed to be related to local environmental conditions [16, 17].

More generally, the relative importance of environment versus host species in mediating moss microbiome composition remains unresolved. With our previous work [14], we found environment to have a negligible effect on microbial community structure compared to moss species identity, however the moss samples included in that study were collected from a single four km<sup>2</sup> area. It is likely that over a larger geographic area, spanning broader environmental gradients (i.e. latitudinal climate shifts and forest versus tundra ecosystem types), site conditions could play a more prominent role in structuring microbial communities. Likewise, although moss microbial communities appear to be strongly influenced by host identity, it is unclear whether the differences in microbial community composition can be predicted from host phylogenetic relationships, the role of host-species effects relative to geographic effects, and importantly, whether the composition of the microbial communities found in a given host species is stable across broad environmental gradients.

These questions are not unique to the study of moss microbiomes, but the important role of bacterial N<sub>2</sub>-fixation makes the moss microbiome a useful model system for simultaneously studying plant-microbe associations and how these associations influence N cycling in high-latitude systems. The importance of moss-associated N<sub>2</sub>-fixing bacteria has been recognized for years [18, 19, 3] and numerous studies have focused specifically on the role of cyanobacteria, particularly *Nostoc*, in N<sub>2</sub>-fixation [20, 21, 4, 22]. However, several studies suggest that cyanobacteria may not be the only N<sub>2</sub>-fixing microbes associated with mosses [23, 24], but the relative importance of these non-cyanobacterial N<sub>2</sub>-fixers and their occurrence across different moss species has not been well-documented. Information on the distributions

of those bacteria capable of N<sub>2</sub>-fixation in mosses and their relative contributions to N<sub>2</sub>-fixation rates can be used to better predict how the capacity for N<sub>2</sub>-fixation varies across hosts and environmental gradients.

The environmental factors structuring the microbial communities of common high-latitude mosses remains largely undetermined. However, warming experiments in *Sphagnum* bogs suggest that temperature regime likely plays a role in structuring the moss microbiome [13]. Light availability and moisture have also been shown to be important in the activity of moss-associated cyanobacteria [4, 25, 26], and these factors may also have direct effects on other microbial taxa. Alternatively, light and moisture may indirectly affect abundances of other microbial taxa as a result of higher N availability provided by the light and moisture sensitive N<sub>2</sub>-fixing cyanobacteria. Finally, soil pH may be yet another environmental factor structuring microbial communities in these host-associated systems [27].

To identify the factors structuring moss-associated bacterial communities and their potential contributions to N<sub>2</sub>-fixation, we collected samples of 26 common boreal moss species from 24 sites spanning a latitudinal transect in Alaska, USA (from 61 to 69° N). We used 16S rRNA gene sequencing to compare microbial community composition across host species and compared the relative effects of host versus environmental factors in structuring microbial community composition. We also quantitatively measured N<sub>2</sub>-fixation rates via <sup>15</sup>N<sub>2</sub> isotopic enrichment in paired moss sub-samples (work published in detail in ref [28]). We then identified potential N<sub>2</sub>-fixing bacteria based on the associations between these taxa and measured N<sub>2</sub>-fixation rates, using publicly available genomic data to confirm their putative N<sub>2</sub>-fixation capabilities.

## Methods

### Site environmental data and sample collection

We collected microbial community marker-gene and isotopic N<sub>2</sub>-fixation rate data from a total of 461 individual moss samples spanning 26 different moss species (see Table S1 for a full species list and Table S2 for details on what samples were collected from each site). These samples were collected from 24 sites across a latitudinal transect in Alaska, USA (from 61 to 69° N, Fig. 1). We selected sites that captured a breadth of common boreal and tundra habitat types in each of the three focal areas along the latitudinal transect (for more details about each site, see Table S3). We visited four sites near Toolik Lake, and 10 sites each near Fairbanks and Anchorage, Alaska, USA. We collected the samples over two years (2016 and 2017) with both collection periods occurring in late June. Site locations and the moss species collected at each site are summarized in Fig. 1.

At each site, we established a 30 m transect, with the latitude and longitude of the site (Table S3) corresponding to the mid-point of the transect. The transect was established to capture the breadth of environmental heterogeneity of each site. At each site, we conducted an informal field-identification of the moss species present. Mosses were characterized as either locally common (appearing in 6 or more distinct patches at the site) or locally rare (appearing in 6 or fewer distinct patches). We collected at least 6 replicates of common mosses and as many replicates as were available for rare mosses from each site. Due to the uncertain nature of field identification and the removal of some samples during sequencing quality control (see

below), not every locally common species has exactly 6 replicates. Moss samples were collected randomly along the transect, with at least 5 m between replicates. For each sample, the nearest 5 m mark along the transect was noted and each sample was associated with the environmental data of the nearest measurement to its collection point.

Environmental data collected along the transect included soil pH, soil moisture, canopy cover, canopy composition and canopy density. These environmental factors were chosen as they are typically used in the characterization of boreal and tundra vegetation communities or have been shown to be important in other studies of moss microbial communities. Canopy cover was measured with a spherical densiometer, while canopy composition and density were determined by identifying the species of each tree within a one meter radius of the 30 m transect and recording its basal area. This composition data was then converted into percent softwood and percent hardwood for each site. Soil pH was measured from a slurry created with a 2:1 ratio of water to soil. Soil moisture was measured by the change in wet to dried weight of biomass of a 5 x 5 x 5 cm plug of moss after 48 h in a drying oven at 60<sup>o</sup> C (calculated as (wet weight-dry weight)/dry weight). We used the Worldclim database [29] to extract mean annual temperature and precipitation values corresponding to the latitude and longitude coordinates for each site.

For each sample, we collected moss from a mono-specific patch and stored samples in a cooler for less than two hours until processed. After collecting each sample, we carefully cleaned off brown or decaying material with a gloved hand and sorted 40 ramets of approximately five cm in length into four equal subsamples (10 ramets each) for microbial sequencing, isotopic natural abundance, isotopic enrichment, and a museum voucher and genotyping specimen. When there were fewer than 40 ramets in a mono-specific patch, we placed five ramets each into the microbial and isotopic samples, leaving the remaining ramets for the voucher specimen. This strategy ensured that measurements of N<sub>2</sub>-fixation rate and microbial communities were as comparable as possible.

After sorting, we stored the moss samples for microbial analyses at -20<sup>o</sup> C. Isotopic enrichment samples were stored in a cooler until measurements were started, usually within two hours of sample sorting. Isotopic natural abundance samples and voucher specimens were dried at 60<sup>o</sup> C for 48 h within two hours of sorting.

#### N<sub>2</sub>-fixation measurements

To measure the rates of N<sub>2</sub>-fixation for each of the 461 moss samples, we used an established isotopic enrichment protocol [28, 30]. Detailed results from the N<sub>2</sub>-fixation data associated with this study can be found in ref [28]. Briefly, we lightly moistened each moss sample with one spray of ddH<sub>2</sub>O from a spray bottle and placed each moss sample into a 60 ml clear polycarbonate syringe that we filled with 10 ml of air and 10 ml of 15N<sub>2</sub> (98% Atom enriched, Sigma-Aldrich Inc. USA) and sealed with a stopcock (for a total headspace of 20 ml). We incubated the sealed syringes at a natural outdoor location near the processing lab for 24 h to mimic the natural light and temperature conditions of the site from which the sample was collected. After 24 h we removed the samples from the syringes and dried them in a 60 °C for 48 h. Natural abundance samples were immediately placed in a drying oven at

60 °C for 48 h and then shipped to Northern Arizona University for measurement of natural abundance of  $^{15}\text{N}$  and  $^{13}\text{C}$ .

$\text{N}_2$ -fixation rates were calculated by comparing the delta  $^{15}\text{N}$  values between the enriched and natural abundance samples. N concentrations from the enriched samples were then used to calculate the amount of N per dry mass weight. To measure these values, we ground each moss sample to a homogenous powder and analyzed it on a Costech ECS4010 coupled to a Thermo Scientific Delta V Advantage Isotope ratio mass spectrometer at Northern Arizona University. Calculations followed the method outlined in ref [30].

#### Genotyping and moss phylogeny

For each moss species, we accessed previously published target capture DNA sequence data [31] (ref in review, manuscript available to reviewers upon request), using sequences from the liverwort *Ptilidium ciliare*, as an outgroup. Briefly, this method targets about 400 conserved plant exons for sequencing. We were unable to access data from the moss *Ptilium crista-castrensis*, so data from this species was not included in the phylogenetic analyses. We followed the bioinformatic pipeline outlined in ref [31] to generate a data matrix containing a single sequence for each locus for each species. We then concatenated all loci into a single supermatrix and used RAxML 8.2.10 with the GTR CAT model [32] to conduct an ML search (100 nonparametric bootstrapping replicates) to generate a phylogeny. The resulting tree matched the phylogeny presented in ref [31], as well as phylogenies based on other data sources [33].

#### Microbial 16S rRNA gene sequencing

To characterize the microbial communities found in each of the 461 moss samples, we used amplicon-based sequencing of a 253-bp region of the bacterial and archaeal 16S rRNA marker gene. First, we homogenized each sample (5 or 10 ramets per sample) with liquid  $\text{N}_2$  under aseptic conditions. We then extracted DNA from 0.25 g of homogenized tissue using the MoBio Power Soil DNA extraction kit (MoBio Laboratories, Carlsbad, CA). Following extraction, we used the 515f/806r primers to PCR amplify the V4-V5 region of the 16S rRNA gene [34]. Each sample was amplified using primer pairs that included a 12-bp barcode unique to each sample and Illumina sequencing adapters. Each sample was homogenized in a randomized order and was assigned a random location on 96-well plates during the DNA extraction and PCR amplification steps. Additionally, to control for any external contaminants, negative controls were included in both the extraction and PCR amplification steps. After triplicate PCR reactions, we normalized amplicon concentrations using ThermoFisher Scientific SequalPrep Normalization plates (Thermo Fisher Scientific Inc. USA) and pooled the amplicons together. Since samples were collected and processed over two years, they were sequenced in two batches. The first batch contained the 2016 samples, while the second contained all of the 2017 samples, samples that failed to sequence during the first run, and 50 randomly chosen duplicate samples from 2016 to quantify any potential run-to-run variation. We sequenced the pooled amplicons on the Illumina MiSeq platform with the 2 x 150

bp paired-end chemistry at the University of Colorado Next Generation Sequencing Facility. Raw sequence reads and all associated metaadata can be found at the GenBank SRA archive accession number PRJNA622832.

After sequencing, we demultiplexed samples using an in-house custom script [35] and used USEARCH v. 10 [36] to merge paired reads and filter low-quality sequences from the data (`fastq_maxee_rate = 1.000`). We used a single-nucleotide variant approach to create a *de novo* database of phylotypes in our samples. In this phylotype identification strategy, sequencing errors are first removed in a denoising step, and then phylotypes are assigned such that each phylotype is represented by a single unique sequence (i.e. all reads that are assigned to that phylotype are 100% identical). Our approach followed the UNOISE pipeline [37]. Briefly, we dereplicated the filtered and merged reads (“-fastx\_uniques”) and then clustered sequences with the “-unoise3” command in USEARCH. We next filtered this database against the GreenGenes database (version August 2013 [38]) to remove phylotypes with a filter threshold of 75% similarity. We assumed sequences that did not meet this threshold were more likely to be of insufficient quality, chimeric, or a product of non-specific amplification.

Following the creation of the *de novo* database, we mapped the filtered and merged reads to the database to create a phylotype-by-sample table (USEARCH “otutab” command). We obtained between 3528 - 36588 reads per sample, with an average sequencing depth of 16495 reads per sample. We classified the reads against the GreenGenes database (version August 2013) using the RDP Naive Bayesian classifier with a minimum confidence threshold of 0.5 [39]. Chloroplasts and mitochondrial reads made up an average of 13% and 3.1% respectively of the total number of quality-filtered reads per sample and we removed these non-prokaryotic sequences from the table prior to downstream analyses. As in previous studies of moss microbial communities [12, 14], the majority of sequences were bacterial and no archaeal reads were detected in any of our samples.

After creating the phylotype abundance table, we performed several cleaning and quality-control steps prior to downstream analyses. First, to control for differences in read depth across samples, we randomly selected 3000 reads per sample. This cut-off was chosen based on the read depth of the non-control sample (i.e. a sample which was neither an extraction blank, nor a PCR no-template-control) with fewest reads after mitochondrial and chloroplast removal. Next, we compared our samples to blanks and filtered samples that were statistically indistinguishable from the blanks; samples that fell within a 97% ellipse of the blanks in multivariate NMDS space were considered “indistinguishable.” Finally, we assessed the contribution of run-to-run variation by analyzing the effect of run in explaining variation in the microbial community of samples included on both runs. The effect of run was very small when compared to any other significant factor (100-fold smaller than any other significant factor at 0.02% variation explained, as determined by PERMANOVA tests, see below). From this we concluded that while there were differences attributable to run, the effect was negligible in the context of our questions. We retained only the duplicates that had been run on the second sequencing run for downstream analysis as that run had slightly better raw sequence quality.

### Statistical analyses

All statistical analyses were conducted in R [40] and data visualized using the `ggplot2` and `igraph` packages [41, 42]. We calculated pairwise community distances using the Bray-Curtis distance metric. We used PERMANOVA tests (R package `vegan` [43]) to assess the degree to which variation in microbial community composition could be attributed to site and host species identity. We next tested whether the moss-associated microbial communities were structured by host phylogeny. In other words, we wanted to know if the differences between communities found in different host species are random with regards to the phylogenetic relatedness of the hosts. We used Mantel tests (R package `vegan` [43]) to assess the degree of phylogenetic correlation between the host phylogeny and the composition of the microbial communities. Mantel tests have been shown to be effective measures of phylogenetic correlation in both real and simulated data and robust to community parameters such as the number of individuals in a community, the beta-diversity metric used, and the presence of microbe-microbe competition [44]. Mantel tests are also less susceptible to false-positive results when compared to phylogenetic congruence methods such as the Robinson-Foulds metric [44].

To isolate the effects of site and moss species in our assessment of factors structuring microbial community differences, we subset the data. For the subsets by site, we divided the data by site and performed the analyses on the eight most moss-rich sites (i.e. the eight sites with the largest number of moss species). For the subsets by moss species, we focused on samples from those moss species which were most abundant and ubiquitous (Fig. 1). A moss species was defined as "abundant and ubiquitous" if it was found at all three Alaskan locations (i.e. Toolik, Fairbanks, and Anchorage) and had more than 10 samples collected in total. The seven moss species meeting these criteria were *Sphagnum russowii*, *Aulacomnium palustre*, *Hylocomnium splendens*, *Pleurozium schreberi*, *Ptilium crista-castrensis*, *Dicranum polysetum*, and *Polytrichum strictum*. This subset of moss species was also used for the analyses of putative N<sub>2</sub>-fixers (see below).

For each of the seven focal moss species, we used BIO-ENV analysis (R package `vegan` [43]) to select the subset of environmental variables strongly correlated with bacterial community composition. Following this variable selection, we used multiple regression of distance matrices (MRM, R package `ecodist` [45]) to determine which variables were significantly correlated (Spearman rank correlations) with the observed differences in microbial community composition.

### Identification of putative N<sub>2</sub>-fixers

To identify phylotypes that were well-correlated with measured N<sub>2</sub>-fixation rates, we ran Spearman correlations between N<sub>2</sub>-fixation rate (log-transformed) and the relative abundance of individual phylotypes. For each of the seven most abundant host species, we first filtered phylotypes for abundance and prevalence. Only those phylotypes with a mean relative abundance equal to or greater than 0.1% across all samples and those phylotypes that were present in at least three samples for a given host species were included in these analyses. We then identified phylotypes that were significantly positively correlated with N<sub>2</sub>-fixation rate as measured by FDR-corrected, Spearman correlations.

After identifying phylotypes that were positively correlated with N<sub>2</sub>-fixation, we attempted to independently verify whether these selected phylotypes were indeed likely capable of N<sub>2</sub>-fixation. A phylotype was identified as likely to possess the ability to fix N<sub>2</sub> if it was (1) already known from the literature to fix N<sub>2</sub> (e.g. members of Nostocaceae), (2) possessed a 16S rRNA gene more than 97% similar to an isolate able to grow on N-free media, or (3) if a publicly-accessible metagenome-assembled genome (MAG) from a similar environment with at least a 97% similar 16S rRNA gene sequence also contained the *nifH* gene.

To identify closely-related isolates capable of growth on N-free media, we used the RDP SeqMatch tool (Cole et al., 2014). We identified MAGs from similar environments using publicly-accessible MAGs from the Integrated Microbial Genomes and Microbiomes database (IMG [46]). We selected medium and high quality MAGs possessing a 16S rRNA gene that were categorized as coming from environments that were similar to those studied here. The environment categories (ontologies, as per MIMAG [47]) were as follows: (1) Host-associated, Plants, Peat Moss (14 MAGs); (2) Environmental, Terrestrial, Peat (174 MAGs); (3) Environmental, Terrestrial, Soil, Wetlands, Permafrost (59 MAGs); and (4) Environmental, Terrestrial, Unclassified, Permafrost (293 MAGs). MAGs were accessed on August 30th, 2019. For each putative N<sub>2</sub>-fixer, we identified any MAGs with a 16S rRNA gene at least 97% similar across our 253-bp PCR amplified gene region using the USEARCH global command [36]. For a full table of independently verified N<sub>2</sub>-fixers and the method of verification, see Table S5.

## Results

### Host structuring of bacterial community composition

The moss-associated bacterial communities had an average richness of 812 unique phylotypes per sample and were dominated by eight bacterial phyla. The average relative abundances of these phyla across all samples were: *Proteobacteria* (49.6%), *Acidobacteria* (12.3%), *Actinobacteria* (8.4%), *Bacteroidetes* (7.5%), *Verrucomicrobia* (7.3%), *Candidatus Eremiobacterota* (WPS-2) (5.1%), *Planctomycetes* (4.2%), and *Cyanobacteria* (3.5%) (see Fig. 2 and Table S6 for details of family abundances across host species). Although the most abundant families and phylotypes found to be associated with the 26 moss species were fairly consistent across the samples, the relative abundances of these taxa varied depending on the host species in question. Differences in bacterial community composition were well-explained by host species identity and site (17.2 and 19.2% of variation, respectively), with a significant portion of the variation explained by the interaction between site and host species (21.7%) (Table 1).

A strong effect of host identity does not necessarily imply strong phylogenetic structuring; however, the differences in bacterial communities across the 26 host species were consistent with phylogenetic structuring as they were well-correlated with the phylogenetic distances between moss hosts (Mantel test,  $\rho = 0.53$ ,  $p = 0.001$ , Table S4). Much of this observed phylogenetic signal in moss bacterial community composition was driven by the unique communities associated with the *Sphagnum* mosses which themselves had a strong and significant phylogenetic signal when the analysis was restricted only to the seven *Sphagnum* moss species (Mantel test,  $\rho = 0.65$ ,  $p = 0.005$ , Table S4).

Because moss species distributions are intrinsically linked to environmental conditions and not every species was present at every site, we used a subset of the samples to determine if the effects of host identity were still significant after controlling for site effects. To do this we selected sites to test individually for the explanatory effect of host species, focusing only on those sites where greater than four moss species co-occurred with at least three individual samples collected per species from that site (for a total of eight sites that met the criteria)(Table S7). With site held constant, the variation explained by host species identity was significant at all eight sites and ranged from 29% (site 16, a white spruce upland) to 55% (Site 5, a bog-like site dominated by black spruce) (Table S7). Together, these results indicate that there is a strong signal of both host identity and host phylogeny on moss bacterial community composition and this signal is independent of differences in site-level environmental conditions.

#### Site-based drivers of bacterial community composition

In addition to the strong host taxonomic and phylogenetic signal in the composition of the moss bacterial communities, we also observed variation in bacterial community composition within a given host species across sites. In analyses of the seven most abundant and ubiquitous moss species we found significant variation in bacterial communities attributable to site when host species was held constant (PERMANOVA,  $R^2=0.37$  to  $0.62$ ,  $p<0.005$  for all seven moss species, Table S8). To better understand the environmental factors that might help explain this variation attributable to the site differences, we determined which environmental variables were correlated with bacterial community composition (via a BIO-ENV analysis on each of the seven focal species) and then assessed the significance and strength of these variables (via MRM analyses) (Fig. S1). We identified between three and six environmental variables as important for structuring bacterial communities in each host species with total correlation values ranging from 0.36 (*Ptilium crista-castrensis*) to 0.60 (*Polytrichum strictum* and *Hylocomium splendens*). The most common predictors were site mean annual temperature, canopy cover, and percent hardwood (for the sites in this study, the hardwoods were all deciduous trees). These results suggest that temperature, light availability (as determined by canopy cover and tree type), and other canopy-based effects (such as litter depth) are important in structuring bacterial communities within individual moss species. However, when we tested these variables for significance with MRM, most predictors had small coefficient values ( $< 0.9$ ) and the subset of significant predictors varied greatly between host species. Thus, although some of the variation in the composition of the bacterial communities is associated with light availability and temperature, the importance of these effects was generally small and highly variable depending on the moss species in question.

#### Identifying potential N<sub>2</sub>-fixers found in boreal mosses

Measured rates of N<sub>2</sub>-fixation ranged from 0 to 21.45  $\mu\text{g N g moss dry wt.}^{-1} \text{ day}^{-1}$  (across the 301 samples collected from the seven focal moss species that were abundant and ubiquitous in our sample set (Fig. 3). To identify those bacterial taxa that were associated with higher N<sub>2</sub>-fixation rates across the seven focal species, we ran

Spearman correlations between the relative abundances of taxa and log-transformed N<sub>2</sub>-fixation rates. We identified 16 bacterial families containing 22 phylotypes that were positively correlated with measured N<sub>2</sub>-fixation rates (Fig. 4). Correlations ranged from 0.34 (phylotype 33 - *Sinobacteraceae*) to 0.75 (phylotype 94 - *Sphingomonas echinoides*). The bacterial families with relative abundances positively correlated with measured N<sub>2</sub>-fixation rates included both taxa known to fix N<sub>2</sub> in mosses (i.e. taxa from the cyanobacterial family *Nostocaceae*) and those which have not yet been identified as capable of N<sub>2</sub>-fixation (i.e. taxa from *Acidobacteraceae*). Interestingly, the phylotypes found to be positively associated with measured N<sub>2</sub>-fixation rates were not necessarily consistent between moss hosts. To better visualize these associations, we created an association network of the correlations (Fig. 5). While most phylotypes were significantly correlated with N<sub>2</sub>-fixation rates across multiple moss hosts, *S. russowii* hosted a unique set of taxa. Furthermore, although most mosses were associated with taxa from a range of bacterial lineages, all of the phylotypes significantly associated with *S. russowii* were from the phylum *Proteobacteria*.

The correlations between the relative abundances of individual bacterial taxa and measured N<sub>2</sub>-fixation rates (shown in Fig. 3), do not necessarily indicate that those bacterial taxa are actually capable of N<sub>2</sub>-fixation. Thus, we used several approaches to try to confirm the N<sub>2</sub>-fixation capabilities of the bacterial taxa that were correlated with measured N<sub>2</sub>-fixation rates. A phylotype was identified as likely to possess the ability to fix N<sub>2</sub> if it was already known from the literature to fix N<sub>2</sub> (e.g. members of *Nostocaceae*), was closely related to a previously-isolated bacterial strain able to grow on N-free media, or if a highly-similar metagenome-assembled genome (MAG) obtained from a similar environment also contained the *nifH* gene. Using these methods, six of the 22 phylotypes were confirmed as likely to possess the ability to fix N<sub>2</sub> (Fig. 5 and Table S5). Notably, only two of the six confirmed phylotypes belong to the family *Nostocaceae*, an N<sub>2</sub>-fixing clade known to associate with mosses. The other four phylotypes come from the family *Methylocystaceae* and the genera *Bradyrhizobium*, *Methylibium*, and *Acidisoma*, expanding the list of bacterial taxa which are likely capable of N<sub>2</sub>-fixation in boreal mosses. For the other 16 phylotypes, there was not enough evidence to determine their N<sub>2</sub>-fixation ability (although some of these phylotypes came from clades where N<sub>2</sub>-fixation is common such as *Burkholderiales*, *Frankiaceae*, or *Acetobacteraceae*). Finally, several phylotypes, such as the three from the phylum *Candidatus Eremiobacterota* (formerly WPS-2) are unlikely to be N<sub>2</sub>-fixers as no genomes from members of this phylum have been shown to contain the *nifH* gene. In total, we can only confirm that 6 of the 22 bacterial phylotypes identified through our correlation-based analyses (Fig. 5) are likely capable of N<sub>2</sub>-fixation.

## Discussion

### The role of host phylogeny in shaping moss bacterial communities

A primary goal of this study was to determine the relative importance of host identity versus site in structuring moss microbial communities. Here we show that moss-associated bacterial communities were strongly structured by host identity and host phylogeny across the 26 moss species surveyed. These results are consistent with previous studies that focused on smaller subsets of moss species which

have also shown that distinct moss host species harbor distinct bacterial communities [15, 14]. Our study extends these results by encompassing a broader diversity of moss species sampled across a large geographic area. We also show that a significant portion of the observed variation in bacterial community composition can be attributed to the phylogenetic relationships between moss species and that this phylogenetic structure is largely attributable to the *Sphagnum* mosses harboring distinct bacterial communities. This phylogenetic signal in the microbiome may be the result of co-diversification between a host and its microbial inhabitants, but it is more likely a product of an independent phylogenetically structured trait (or traits) selecting for certain microbes from the environment (“ecological filtering” [44, 48]). Importantly, the phylogenetic signal we observed is stronger than simulations suggest should be expected for an ecological filtering model [44]. This suggests that either there has been co-diversification of the bacterial communities with their moss hosts (unlikely given the divergence of *Sphagnum* from other mosses is close to 400 mya, [49]) or that the moss traits that are responsible for the ecological filtering of bacterial communities exhibit a phylogenetic signal that is stronger than is typically expected (i.e. Blomberg K  $\ll$  1 [50]).

This strong phylogenetic signal in the structure of the moss microbiome stands in stark contrast to most vascular plants which typically exhibit no signal or very weak phylogenetic signals between host species (for a summary of available literature – See Table S9). However, the majority of studies investigating phylogenetic signals in plant microbiomes focus on rhizosphere communities. The patterns for aboveground bacterial communities in plants are less well understood. Although previous studies have found that plant species have distinct phyllosphere communities (reviewed in refs [51, 52]), we know of only three studies that have explicitly tested for a phylogenetic signal in these communities. Across these studies, all of which focused on vascular plants (and trees, in particular), the strength of phylogenetic signal ranges from moderate to strong [53, 54, 55]. The strength of the phylogenetic signal in moss microbial communities is stronger than any of these previous phyllosphere studies and particularly strong when taken in the context of the plant microbiome literature as a whole.

The observed phylogenetic signal in microbiome composition is strongly driven by *Sphagnum* mosses harboring bacterial communities distinct from those found in other mosses. The strong phylogenetic signal in the *Sphagnum* mosses’ bacterial communities could be due to ecological filtering by the internal compartments (i.e. hyaline cells) that are unique to *Sphagnum* anatomy. Hyaline cells hold moisture and harbor microbes [12] and attributes of these cells vary among *Sphagnum* species [56]. Internal compartments are often linked with stronger phylogenetic signals in other host microbiomes [44, 48]. *Sphagnum* species also acidify their local environment [56] and induce changes in canopy cover and moisture availability [2] - factors that could effectively select for the unique microbial communities associated with *Sphagnum* mosses.

#### Diverse potential N<sub>2</sub>-fixing taxa in moss bacterial communities

Mosses in northern latitude ecosystems harbor a diverse pool of potential N<sub>2</sub>-fixers, which play a key role in regulating N availability. We were able to identify 22 bacterial phylotypes from 14 bacterial families that had relative abundances positively

correlated with measured  $N_2$ -fixation rates (Fig. 4), pointing to strong associations between a diverse group of bacterial taxa and measured  $N_2$ -fixation rates. Moss species host distinct communities of potential  $N_2$ -fixers with a different set of phylotypes correlating more strongly with each host species (Fig. 5). Members of the *Sphagnum* genus (particularly *S. russowii*) generally hosted a unique set of bacterial taxa that was not associated with  $N_2$ -fixation rates in any of the other moss species sampled (Fig. 5). *Sphagnum* species not only host unique, phylogenetically structured microbial communities (Fig. 2, Table S4), they also host a unique community of potential  $N_2$ -fixers and this unique community may partly explain the high  $N_2$ -fixation rates observed across the *Sphagnum* species (Fig. 3).

Our results suggest that  $N_2$ -fixation may be carried out by a number of bacterial taxa beyond those within the phylum *Cyanobacteria* (Fig. 5), which are often considered the predominant  $N_2$ -fixers in mosses [4, 22]. While abundance and  $N_2$ -fixation are not necessarily linked and even low abundance  $N_2$ -fixers can contribute to high  $N_2$ -fixation rates [7], our hypothesis that there are more non-cyanobacterial  $N_2$ -fixers in boreal mosses than previously recognized is in line with previous studies in *Sphagnum* mosses [23, 24]. Our results support these findings and broaden them to include other moss lineages since the phylotypes most strongly associated with fixation rates on *Sphagnum russowii* were non-cyanobacterial and each of our seven focal moss species also hosted multiple putative  $N_2$ -fixers that were non-cyanobacterial.

The potential for mosses to harbor a broad diversity of  $N_2$ -fixing bacteria is likely an important ecological feature. While it is possible that distinct  $N_2$ -fixers may be in direct competition [57], it is also possible that they could have distinct niches within the moss host. For example, oxygenic photosynthesizers, such as *Cyanobacteria*, may be limited by light conditions, while anoxygenic photosynthesizing  $N_2$ -fixers, such as members of the genus *Acidisoma* [58], may be able to fix  $N_2$  in light conditions that are sub-optimal for *Cyanobacteria* [59]. A promising direction for future work is to determine how  $N_2$ -fixation rates in mosses are the product of interactions (direct or indirect) between the diverse array of  $N_2$ -fixers that can be found in a given host (e.g. ref [60]).

#### The role of environment in shaping moss bacterial communities

In addition to distinct species and lineages of mosses having unique bacterial communities, the bacteria associated with individual moss species were variable in composition and some of this intra-species variation in bacterial community composition could be predicted from the measured site characteristics, namely light availability and temperature. Light availability is particularly interesting as it is a commonly cited factor governing  $N_2$ -fixation rates [11, 61, 62]. Multiple studies have attributed light-induced increases in  $N_2$ -fixation to the presence of photosynthetic  $N_2$ -fixing cyanobacteria. However, other phototrophic organisms (including anoxygenic phototrophs from *Alphaproteobacteria* or the candidate phylum *Eremiobacterota* (WPS-2)) are also prominent members of moss bacterial communities [14], and these taxa may also be contributing to the bacterial community responses to light availability, highlighting the potentially important contributions of phototrophs in the moss microbiome. This finding is important as temperature and light availability are some

of the more variable environmental conditions in high-latitude systems and both factors are likely to shift with climate change due to elevated temperatures and as denser-canopied deciduous birch (*Betula*) species encroach on formerly open-canopied spruce-dominated areas with increases in fire prevalence [63, 64, 65].

Since a large portion of variation is attributable to site, but the environmental predictors we measured were fairly weak predictors, it is likely that “site” may be better represented by a suite of factors not measured in this study. These factors might include abiotic factors such as stand age, parent bedrock material, and even trace metal availability [4, 66] or local biotic factors such as host moss genotype, changes in light availability through shading by understory vascular plant communities (not captured in our measurements of canopy cover), top-down controls by protists and microfauna [67, 68], or community turnover due to microbial colonization from the surrounding local micro-environment [69]. Regardless, our results suggest that host species identity, rather than measured or unmeasured environmental factors, is a more consistent indicator of bacterial community structure and that moss species distributions may be more relevant than site characteristics for understanding the contributions of the moss microbiome to ecosystem processes, including N<sub>2</sub>-fixation.

## Conclusions

The results from this study indicate that microbiomes of boreal mosses are phylogenetically structured and that moss species identity, not site environmental conditions, is the best predictor of microbial community composition. Conversely, moss microbial communities are not easily predictable from site characteristics, although our results suggest that light and temperature can have significant, though subtle, effects on the composition of the moss microbiome. We also found that the identities of the potential N<sub>2</sub>-fixing bacteria were host species-specific suggesting that N<sub>2</sub>-fixation may be best predicted at the host-level. Finally, many of the potential N<sub>2</sub>-fixing microbes we identified were not cyanobacteria. These non-cyanobacterial bacteria are worthy of future study to better understand the processes controlling fixation rates in boreal mosses. More generally, our work demonstrates the utility of mosses for studying plant microbiomes and broadens our understanding of how moss-microbe interactions contribute to N dynamics in high-latitude ecosystems.

### Declarations

#### Acknowledgments

We are grateful for the field and laboratory assistance provided by Dakshina Marlier, Jessica Henley, Lady Grant, and Maya Montoya-Pimolwatana. We deeply appreciate Manuel Delgado-Baquerizo's assistance in the retrieval of climate data from the Worldclim database. We also acknowledge infrastructural support provided by the University of Colorado Next Generation Sequencing Facility, the University of Alaska, Fairbanks, and Toolik Field Station.

#### Funding

Financial support for this work was provided by the U.S. National Science Foundation Dimensions of Biodiversity program (DEB 1542609). The funding body had no direct role in the design, collection, analysis, interpretation of the data, or writing of the manuscript.

#### Author Contributions

This project was conceived of by N.F., M.C.M., S.F.M., and J.M.P. The manuscript was written by H.H.M. and N.F. with contributions from all co-authors. All authors helped collect the field data and samples. J.E.M.S., M.C.M., and S. M. performed the isotopic analyses. H.H.M. and N.F. performed the microbial analyses. L.R.L. and S.F.M. performed the moss genotyping and constructed the host phylogenetic tree. H.H.M. and J.M.P. conducted the statistical analyses.

## Availability of data and material

The datasets generated and analyzed during the current study are available in the GenBank Sequence Read Archive (SRA), accession number: PRJNA622832 <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA622832>. Under this identifier the raw, unprocessed reads and all associated environmental data used in this study are available for each sample. The permanent identifiers for each genome used from IMG as evidence for nitrogen-fixing potential can be found in Table S5. IMG genomes can be accessed at: <https://img.jgi.doe.gov/>

## Competing Interests

The authors declare that they have no competing interests.

## Ethics approval and consent to participate

Not applicable

## Consent for publication

Not applicable

## Author details

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, USA. <sup>2</sup>Center for Ecosystem Science and Society and the Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA. <sup>3</sup>Provost Office, University of Florida, Gainesville, FL, USA. <sup>4</sup>Department of Biology, University of Florida, Gainesville, FL, USA. <sup>5</sup>Cooperative Institute for Environmental Sciences, University of Colorado, Boulder, USA.

## References

- Turetsky, M.R., Mack, M.C., Hollingsworth, T.N., Harden, J.W.: The role of mosses in ecosystem succession and function in Alaska's boreal forest. *Canadian Journal of Forest Research* **40**(7), 1237–1264 (2010). doi:10.1139/X10-072
- Turetsky, M.R., Bond-Lamberty, B., Euskirchen, E., Talbot, J., Frohling, S., McGuire, A.D., Tuittila, E.S.: The resilience and functional role of moss in boreal and arctic ecosystems. *New Phytologist* **196**, 49–67 (2012). doi:10.1111/j.1469-8137.2012.04254.x
- DeLuca, T.H., Zackrisson, O., Nilsson, M.-C., Sellstedt, A.: Quantifying nitrogen-fixation in feather moss carpets of boreal forests. *Nature* **419**(6910), 917–920 (2002)
- Rousk, K., Jones, D.L., Deluca, T.H.: Moss-cyanobacteria associations as biogenic sources of nitrogen in boreal forest ecosystems. *Frontiers in Microbiology* **4**(June), 1–10 (2013). doi:10.3389/fmicb.2013.00150
- Hobbie, S.E., Schimel, J.P., Trumbore, S.E., Randerson, J.R.: Controls over carbon storage and turnover in high-latitude soils. *Global Change Biology* **6**(SUPPLEMENT 1), 196–210 (2000). doi:10.1046/j.1365-2486.2000.06021.x
- Lindo, Z., Gonzalez, A.: The bryosphere: An integral and influential component of the Earth's biosphere. *Ecosystems* **13**(4), 612–627 (2010). doi:10.1007/s10021-010-9336-3
- Warshan, D., Bay, G., Nahar, N., Wardle, D.A., Nilsson, M.-C., Rasmussen, U.: Seasonal variation in *nifH* abundance and expression of cyanobacterial communities associated with boreal feather mosses. *The ISME Journal*, 1–11 (2016). doi:10.1038/ismej.2016.17
- Kip, N., Winden, J.F.V., Pan, Y., Bodrossy, L., Reichart, G.-j., Smolders, A.J.P., Jetten, M.S.M., Damsté, J.S.S., Camp, H.J.M.O.D.: Global prevalence of methane oxidation by symbiotic bacteria in peat-moss ecosystems. *Nature Geoscience* **3**(9), 617–621 (2010). doi:10.1038/ngeo939
- Larmola, T., Leppänen, S.M., Tuittila, E.-S., Aarva, M., Merilä, P., Fritze, H., Tirola, M.: Methanotrophy induces nitrogen fixation during peatland development. *Proceedings of the National Academy of Sciences of the United States of America* **111**(2), 734–9 (2014). doi:10.1073/pnas.1314284111
- Turetsky, M.: The Role of Bryophytes in Carbon and Nitrogen Cycling. *The Bryologist* **106**(3), 395–409 (2003). doi:10.1639/05
- Gundale, M.J., Nilsson, M., Bansal, S., Jäderlund, A.: The interactive effects of temperature and light on biological nitrogen fixation in boreal forests. *New Phytologist* **194**(2), 453–463 (2012). doi:10.1111/j.1469-8137.2012.04071.x
- Kostka, J.E., Weston, D.J., Glass, J.B., Lilleskov, E.A., Shaw, A.J., Turetsky, M.R.: The *Sphagnum* microbiome: new insights from an ancient plant lineage. *New Phytologist*, 1–8 (2016). doi:10.1111/nph.13993
- Carrell, A.A., Kolton, M., Glass, J.B., Pelletier, D.A., Warren, M.J., Kostka, J.E., Iversen, C.M., Hanson, P.J., Weston, D.J.: Experimental warming alters the community composition, diversity, and N<sub>2</sub> fixation activity of peat moss (*Sphagnum fallax*) microbiomes. *Global Change Biology* (November 2018), 2993–3004 (2019). doi:10.1111/gcb.14715
- Holland-Moritz, H., Stuart, J., Lewis, L.R., Miller, S., Mack, M.C., McDaniel, S.F., Fierer, N.: Novel bacterial lineages associated with boreal moss species. *Environmental Microbiology* **20**(7), 2625–2638 (2018). doi:10.1111/1462-2920.14288
- Bragina, A., Berg, C., Berg, G.: The core microbiome bonds the Alpine bog vegetation to a transkingdom metacommunity. *Molecular Ecology* **24**(18), 4795–4807 (2015). doi:10.1111/mec.13342
- Opelt, K., Chobot, V., Hadacek, F., Schönmann, S., Eberl, L., Berg, G.: Investigations of the structure and function of bacterial communities associated with *Sphagnum* mosses. *Environmental Microbiology* **9**(11), 2795–2809 (2007). doi:10.1111/j.1462-2920.2007.01391.x
- Cutler, N.A., Arróniz-crespo, M., Street, L.E., Jones, D.L., Chaput, D.L., Deluca, T.H.: Long-term recovery of microbial communities in the boreal bryosphere following fire disturbance. *Microbial Ecology*, 75–90 (2017). doi:10.1007/s00248-016-0832-7
- Alexander, V., Billington, M., Schell, D.M.: Nitrogen fixation in arctic and alpine tundra. In: Tieszen, L.L. (ed.) *Vegetation and Production Ecology of an Alaskan Arctic Tundra*, pp. 539–558. Springer, New York (1978)

19. Basilier, K., Granhall, U., Stenstrom, T.-A.: Nitrogen Fixation in Wet Minerotrophic Moss Communities of a Subarctic Mire **31**(2), 236–246 (1978)
20. Bay, G., Nahar, N., Oubre, M., Whitehouse, M.J., Wardle, D.A., Zackrisson, O., Nilsson, M.C., Rasmussen, U.: Boreal feather mosses secrete chemical signals to gain nitrogen. *New Phytologist* **200**(1), 54–60 (2013). doi:10.1111/nph.12403
21. Berg, A., Danielsson, Å., Svensson, B.H.: Transfer of fixed-N from N<sub>2</sub>-fixing cyanobacteria associated with the moss *Sphagnum riparium* results in enhanced growth of the moss. *Plant and Soil* **362**(1-2), 271–278 (2013). doi:10.1007/s11104-012-1278-4
22. Warshan, D., Espinoza, J.L., Stuart, R.K., Richter, R.A., Kim, S.Y., Shapiro, N., Woyke, T., C Kyrpides, N., Barry, K., Singan, V., Lindquist, E., Ansong, C., Purvine, S.O., Brewer, H.M., Weyman, P.D., Dupont, C.L., Rasmussen, U.: Feathermoss and epiphytic *Nostoc* cooperate differently: Expanding the spectrum of plant-cyanobacteria symbiosis. *ISME Journal* **11**(12), 2821–2833 (2017). doi:10.1038/ismej.2017.134
23. Bragina, A., Maier, S., Berg, C., Müller, H., Chobot, V., Hadacek, F., Berg, G.: Similar diversity of *Alphaproteobacteria* and nitrogenase gene amplicons on two related *Sphagnum* mosses. *Frontiers in Microbiology* **2**(JAN), 1–10 (2012). doi:10.3389/fmicb.2011.00275
24. Vile, M.A., Kelman Wieder, R., Živković, T., Scott, K.D., Vitt, D.H., Hartsock, J.A., Iosue, C.L., Quinn, J.C., Petix, M., Fillingim, H.M., Popma, J.M.A., Dynarski, K.A., Jackman, T.R., Albright, C.M., Wykoff, D.D.: N<sub>2</sub>-fixation by methanotrophs sustains carbon and nitrogen accumulation in pristine peatlands. *Biogeochemistry* **121**(2), 317–328 (2014). doi:10.1007/s10533-014-0019-6
25. Kox, M.A.R., van den Elzen, E., Lamers, L.P.M., Jetten, M.S.M., van Kessel, M.A.H.J.: Microbial nitrogen fixation and methane oxidation are strongly enhanced by light in *Sphagnum* mosses. *AMB Express* **10**(1) (2020). doi:10.1186/s13568-020-00994-9
26. Liebner, S., Zeyer, J., Wagner, D., Schubert, C., Pfeiffer, E.M., Knoblauch, C.: Methane oxidation associated with submerged brown mosses reduces methane emissions from Siberian polygonal tundra. *Journal of Ecology* **99**(4), 914–922 (2011). doi:10.1111/j.1365-2745.2011.01823.x
27. Dedysh, S.N.: Cultivating uncultured bacteria from northern wetlands: Knowledge gained and remaining gaps. *Frontiers in Microbiology* **2**(SEP), 1–15 (2011). doi:10.3389/fmicb.2011.00184. 25136
28. Stuart, J.E.M., Holland-Moritz, H., Lewis, L.R., Jean, M., Miller, S.N., McDaniel, S.F., Fierer, N., Ponciano, J.M., Mack, M.C.: Host Identity as a Driver of Moss-Associated N<sub>2</sub> Fixation Rates in Alaska. *Ecosystems* (2020). doi:10.1007/s10021-020-00534-3
29. Fick, S.E., Hijmans, R.J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**(12), 4302–4315 (2017). doi:10.1002/joc.5086
30. Jean, M., Mack, M.C., Johnstone, J.F.: Spatial and temporal variation in moss-associated dinitrogen fixation in coniferous- and deciduous-dominated Alaskan boreal forests. *Plant Ecology* **219**(7), 837–851 (2018). doi:10.1007/s11258-018-0838-y
31. Breinholt, J., McDaniel, S.F., Carey, S., Davis, C., Nieves, L., Sessa, E., von Konrat, M., Burleigh, G.: Targeted enrichment probe set for resolving the flagellate plant tree of life. *bioRxiv* (2020)
32. Stamatakis, A.: RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**(9), 1312–1313 (2014). doi:10.1093/bioinformatics/btu033. bioinformatics/btu033
33. Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., Deyholos, M.K., Gitzendanner, M.A., Graham, S.W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., Porsch, M., Quint, M., Rensing, S.A., Soltis, D.E., Soltis, P.S., Stevenson, D.W., Ullrich, K.K., Wickert, N.J., DeGironimo, L., Edger, P.P., Jordon-Thaden, I.E., Joya, S., Liu, T., Melkonian, B., Miles, N.W., Pokorny, L., Quigley, C., Thomas, P., Villarreal, J.C., Augustin, M.M., Barrett, M.D., Baucom, R.S., Beerling, D.J., Benstein, R.M., Biffin, E., Brockington, S.F., Burge, D.O., Burris, J.N., Burris, K.P., Burtet-Sarramegna, V., Caicedo, A.L., Cannon, S.B., Çebi, Z., Chang, Y., Chater, C., Cheeseman, J.M., Chen, T., Clarke, N.D., Clayton, H., Covshoff, S., Crandall-Stotler, B.J., Cross, H., DePamphilis, C.W., Der, J.P., Determann, R., Dickson, R.C., Di Stilio, V.S., Ellis, S., Fast, E., Feja, N., Field, K.J., Filatov, D.A., Finnegan, P.M., Floyd, S.K., Fogliani, B., García, N., Gâteblé, G., Godden, G.T., Goh, F.Q.Y., Greiner, S., Harkess, A., Heaney, J.M., Helliwell, K.E., Heyduk, K., Hibberd, J.M., Hodel, R.G.J., Hollingsworth, P.M., Johnson, M.T.J., Jost, R., Joyce, B., Kapralov, M.V., Kazamia, E., Kellogg, E.A., Koch, M.A., Von Konrat, M., Könyves, K., Kutchan, T.M., Lam, V., Larsson, A., Leitch, A.R., Lentz, R., Li, F.W., Lowe, A.J., Ludwig, M., Manos, P.S., Mavrodiev, E., McCormick, M.K., McKain, M., McLellan, T., McNeal, J.R., Miller, R.E., Nelson, M.N., Peng, Y., Ralph, P., Real, D., Riggins, C.W., Ruhsam, M., Sage, R.F., Sakai, A.K., Scascitella, M., Schilling, E.E., Schlösser, E.M., Sederoff, H., Servick, S., Sessa, E.B., Shaw, A.J., Shaw, S.W., Sigel, E.M., Skema, C., Smith, A.G., Smithson, A., Stewart, C.N., Stinchcombe, J.R., Szövényi, P., Tate, J.A., Tiebel, H., Trapnell, D., Villegente, M., Wang, C.N., Weller, S.G., Wenzel, M., Weststrand, S., Westwood, J.H., Whigham, D.F., Wu, S., Wulff, A.S., Yang, Y., Zhu, D., Zhuang, C., Zuidof, J., Chase, M.W., Pires, J.C., Rothfels, C.J., Yu, J., Chen, C., Chen, L., Cheng, S., Li, J., Li, R., Li, X., Lu, H., Ou, Y., Sun, X., Tan, X., Tang, J., Tian, Z., Wang, F., Wang, J., Wei, X., Xu, X., Yan, Z., Yang, F., Zhong, X., Zhou, F., Zhu, Y., Zhang, Y., Ayyampalayam, S., Barkman, T.J., phuong Nguyen, N., Matasci, N., Nelson, D.R., Sayyari, E., Wafula, E.K., Walls, R.L., Warnow, T., An, H., Arrigo, N., Baniaga, A.E., Galuska, S., Jorgensen, S.A., Kidder, T.I., Kong, H., Lu-Irving, P., Marx, H.E., Qi, X., Reardon, C.R., Sutherland, B.L., Tiley, G.P., Welles, S.R., Yu, R., Zhan, S., Gramzow, L., Theißen, G., Wong, G.K.S.: One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**(7780), 679–685 (2019). doi:10.1038/s41586-019-1693-2
34. Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R.: Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal* **6**(8), 1621–1624 (2012). doi:10.1038/ismej.2012.8. arXiv:1408.1149
35. Leff, J.: prep\_samples\_for\_uparse\_paired.py (2016). <https://github.com/leffj/helper-code-for-uparse> Accessed Jan 10 2019
36. Edgar, R.C.: Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**(19), 2460–2461 (2010). doi:10.1093/bioinformatics/btq461

37. Edgar, R.C.: Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics* **34**(14), 2371–2375 (2018). doi:10.1093/bioinformatics/bty113
38. McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., Hugenholtz, P.: An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal* **6**(3), 610–618 (2012). doi:10.1038/ismej.2011.139
39. Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R.: Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* **73**(16), 5261–5267 (2007). doi:10.1128/AEM.00062-07
40. R Core Team: R: A Language and Environment for Statistical Computing, Vienna, Austria (2017). <https://www.r-project.org/>
41. Csardi, G., Nepusz, T.: The igraph software package for complex network research. *InterJournal Complex Sy*, 1695 (2006)
42. Wickham, H.: *ggplot2: Elegant Graphics for Data Analysis*. Springer, New York (2016). <http://ggplot2.org>
43. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H.: *vegan: Community Ecology Package*. (2019). <https://cran.r-project.org/package=vegan>
44. Mazel, F., Davis, K.M., Loudon, A., Kwong, W.K., Groussin, M., Parfrey, L.W.: Is Host Filtering the Main Driver of Phyllosymbiosis across the Tree of Life? *mSystems* **3**(5), 00097–18 (2018). doi:10.1128/msystems.00097-18
45. Goslee, S.C., Urban, D.L.: The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* **22**(7), 1–19 (2007)
46. Chen, I.M.A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J.R., Seshadri, R., Smirnova, T., Kirton, E., Jungbluth, S.P., Woyke, T., Eloë-Fadrosh, E.A., Ivanova, N.N., Kyrpides, N.C.: IMG/M v.5.0: An integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Research* **47**(D1), 666–677 (2019). doi:10.1093/nar/gky901
47. Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloë-Fadrosh, E.A., Tringe, S.G., Ivanova, N.N., Copeland, A., Clum, A., Becraft, E.D., Malmstrom, R.R., Birren, B., Podar, M., Bork, P., Weinstock, G.M., Garrity, G.M., Dodsworth, J.A., Yooseph, S., Sutton, G., Glöckner, F.O., Gilbert, J.A., Nelson, W.C., Hallam, S.J., Jungbluth, S.P., Etema, T.J.G., Tighe, S., Konstantinidis, K.T., Liu, W.T., Baker, B.J., Rattei, T., Eisen, J.A., Hedlund, B., McMahon, K.D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G.W., Rinke, C., Lapidus, A., Meyer, F., Yilmaz, P., Parks, D.H., Eren, A.M., Schriml, L., Banfield, J.F., Hugenholtz, P., Woyke, T.: Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology* **35**(8), 725–731 (2017). doi:10.1038/nbt.3893
48. Pollock, F.J., McMinds, R., Smith, S., Bourne, D.G., Willis, B.L., Medina, M., Thurber, R.V., Zaneveld, J.R.: Coral-associated bacteria demonstrate phyllosymbiosis and cophylogeny. *Nature Communications* **9**(1), 4921 (2018). doi:10.1038/s41467-018-07275-x
49. Laenen, B., Shaw, B., Schneider, H., Goffinet, B., Paradis, E., Désamorie, A., Heinrichs, J., Villarreal, J.C., Gradstein, S.R., McDaniel, S.F., Long, D.G., Forrest, L.L., Hollingsworth, M.L., Crandall-Stotler, B., Davis, E.C., Engel, J., Von Konrat, M., Cooper, E.D., Patiño, J., Cox, C.J., Vanderpoorten, A., Shaw, A.J.: Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts. *Nature Communications* **5**, 1–6 (2014). doi:10.1038/ncomms6134
50. Blomberg, S.P., Garland, T., Ives, A.R.: Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. *Evolution* **57**(4), 717–745 (2003). doi:10.1111/j.0014-3820.2003.tb00285.x
51. Vacher, C., Hampe, A., Porté, A., Sauer, U., Compant, S., Morris, C.E.: The phyllosphere: microbial jungle at the plant-climate interface. *Annual Review of Ecology Evolution and Systematics* **47**(June), 1–24 (2016). doi:10.1146/annurev-ecolsys-121415-032238
52. Remus-Emsermann, M.N.P., Schlechter, R.O.: Phyllosphere microbiology: at the interface between microbial individuals and the plant host. *New Phytologist* **218**(4), 1327–1333 (2018). doi:10.1111/nph.15054
53. Redford, A.J., Bowers, R.M., Knight, R., Linhart, Y., Fierer, N.: The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environmental microbiology* **12**(11), 2885–93 (2010). doi:10.1111/j.1462-2920.2010.02258.x
54. Kim, M., Singh, D., Lai-Hoe, A., Go, R., Rahim, R.A., Ainuddin, A.N., Chun, J., Adams, J.M.: Distinctive Phyllosphere Bacterial Communities in Tropical Trees. *Microbial Ecology* **63**(3), 674–681 (2012). doi:10.1007/s00248-011-9953-1
55. Kembel, S.W., O'Connor, T.K., Arnold, H.K., Hubbell, S.P., Wright, S.J., Green, J.L.: Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. *Proceedings of the National Academy of Sciences* **111**(38), 13715–13720 (2014). doi:10.1073/pnas.1216057111
56. Clymo, R.S., Hayward, P.M.: *The Ecology of Sphagnum*. In: Smith, A.J.E. (ed.) *Bryophyte Ecology*, pp. 229–289. Springer, Dordrecht (1982)
57. Ininbergs, K., Bay, G., Rasmussen, U., Wardle, D.A., Nilsson, M.C.: Composition and diversity of *nifH* genes of nitrogen-fixing cyanobacteria associated with boreal forest feather mosses. *New Phytologist* **192**(2), 507–517 (2011). doi:10.1111/j.1469-8137.2011.03809.x
58. Belova, S.E., Pankratova, T.A., Detkova, E.N., Kaparullina, E.N., Dedys, S.N.: *Acidisoma tundrae* gen. nov., sp. nov. and *Acidisoma sibiricum* sp. nov., two acidophilic, psychrotolerant members of the *Alphaproteobacteria* from acidic northern wetlands. *International Journal of Systematic and Evolutionary Microbiology* **59**(9), 2283–2290 (2009). doi:10.1099/ijs.0.009209-0
59. Atamna-Ismaeel, N., Finkel, O.M., Glaser, F., Sharon, I., Schneider, R., Post, A.F., Spudich, J.L., Mering, C.V., Vorholt, J.A., Iluz, D.: Microbial rhodopsins on leaf surfaces of terrestrial plants. *Environmental Microbiology* **14**(1), 140–146 (2012). doi:10.1111/j.1462-2920.2011.02554.x
60. Van Den Elzen, E., Kox, M.A.R., Harpenslager, S.F., Hensgens, G., Fritz, C., Jetten, M.S.M., Ettwig, K.F., Lamers, L.P.M.: Symbiosis revisited: Phosphorus and acid buffering stimulate N<sub>2</sub> fixation but not *Sphagnum*

- growth. *Biogeosciences* **14**(5), 1111–1122 (2017). doi:10.5194/bg-14-1111-2017
61. Sorensen, P.L., Lett, S., Michelsen, A.: Moss-specific changes in nitrogen fixation following two decades of warming, shading, and fertilizer addition. *Plant Ecology* **213**(4), 695–706 (2012). doi:10.1007/s11258-012-0034-4
  62. Stuiver, B.M., Gundale, M.J., Wardle, D.A., Nilsson, M.C.: Nitrogen fixation rates associated with the feather mosses *Pleurozium schreberi* and *Hylocomium splendens* during forest stand development following clear-cutting. *Forest Ecology and Management* **375**, 309 (2016). doi:10.1016/j.foreco.2016.05.038
  63. Johnstone, J.F., Hollingsworth, T.N., Chapin, F.S., Mack, M.C.: Changes in fire regime break the legacy lock on successional trajectories in Alaskan boreal forest. *Global Change Biology* **16**(4), 1281–1295 (2010). doi:10.1111/j.1365-2486.2009.02051.x
  64. De Groot, W.J., Cantin, A.S., Flannigan, M.D., Soja, A.J., Gowman, L.M., Newbery, A.: A comparison of Canadian and Russian boreal forest fire regimes. *Forest Ecology and Management* **294**, 23–34 (2013). doi:10.1016/j.foreco.2012.07.033
  65. Foster, A.C., Armstrong, A.H., Shuman, J.K., Shugart, H.H., Rogers, B.M., Mack, M.C., Goetz, S.J., Ranson, K.J.: Importance of tree- and species-level interactions with wildfire, climate, and soils in interior Alaska: Implications for forest change under a warming climate. *Ecological Modelling* **409**(July), 108765 (2019). doi:10.1016/j.ecolmodel.2019.108765
  66. Rousk, K., Degboe, J., Michelsen, A., Bradley, R., Bellenger, J.P.: Molybdenum and phosphorus limitation of moss-associated nitrogen fixation in boreal ecosystems. *New Phytologist* **214**(1), 97–107 (2017). doi:10.1111/nph.14331
  67. Kardol, P., Spitzer, C.M., Gundale, M.J., Nilsson, M.C., Wardle, D.A.: Trophic cascades in the bryosphere: the impact of global change factors on top-down control of cyanobacterial N<sub>2</sub>-fixation. *Ecology letters* **19**(8), 967–976 (2016). doi:10.1111/ele.12635
  68. Heger, T.J., Giesbrecht, I.J.W., Gustavsen, J., del Campo, J., Kellogg, C.T.E., Hoffman, K.M., Lertzman, K., Mohn, W.W., Keeling, P.J.: High-throughput environmental sequencing reveals high diversity of litter and moss associated protist communities along a gradient of drainage and tree productivity. *Environmental Microbiology* **20**(3), 1185–1203 (2018). doi:10.1111/1462-2920.14061
  69. Putkinen, A., Larmola, T., Tuomivirta, T., Siljanen, H.M.P., Bodrossy, L., Tuittila, E.S., Fritze, H.: Water dispersal of methanotrophic bacteria maintains functional methane oxidation in *Sphagnum* mosses. *Frontiers in Microbiology* **3**(JAN), 1–10 (2012). doi:10.3389/fmicb.2012.00015

## Figures

**Figure 1** A map summarizing the species and sampling locations of the 461 samples in this study. Images (left) from each transect location show the environmental heterogeneity across the sites. The table (top right) displays the list of species sampled, ordered by the total number of samples collected for each species (see Supplementary Tables S1 and S2). Dots indicate the sites where at least one sample of a particular species was collected. The blue boxes indicate the seven moss species identified as “abundant and ubiquitous” in our data set (see Methods). Images of these seven species are displayed at the bottom right corner of the figure.

**Figure 2** A heatmap showing the abundant bacterial families across different host species with a tree showing the phylogenetic relationships between the moss host species. Only 24 of the 26 moss species are included in this heatmap (see Methods). The outgroup host species, *Ptilidium ciliare*, is a liverwort.

**Figure 3** N<sub>2</sub>-fixation rates for the seven abundant and ubiquitous species in our data set as measured by isotope enrichment over a 1-day incubation period. Moss species are arranged from highest average fixation rate to lowest. All seven species hosted N<sub>2</sub>-fixing microbes however rates were highly variable across host species. Error bars represent 1 standard error above and below the mean. A more in-depth analysis of the differences in N<sub>2</sub>-fixation across moss species can be found in ref [28]

## Tables

### Additional Files

Additional file 1 — Supplementary Tables and Figures

Supplementary Microsoft Word Document (.docx) containing supplementary figures and and small supplementary tables.

List of supplementary tables and figures in Additional file 1:

**Table S1:** A list of all of the moss species sampled and the number of samples collected for each.

**Figure 4** Spearman correlations between the relative abundances of individual bacterial phylotypes and measured N<sub>2</sub>-fixation rates. Bacterial phylotypes used in this analysis were limited to those that had an average relative abundance of >0.1% across the more abundant and ubiquitous moss host species (156 phylotypes). The data is separated into panels based on the family-level taxonomy of the phylotypes. Colored points and lines represent bacterial phylotypes that were significantly positively correlated with N<sub>2</sub>-fixation in a particular moss species. Lines and points are colored according to the moss species in question. Grey dots are measurements for 137 phylotypes that were not found to be significantly positively correlated with N<sub>2</sub>-fixation rates. We note that, to improve clarity, the x and y scales vary between panels.

**Figure 5** Correlation network showing phylotypes (colored nodes) which showed significant positive Spearman correlations with N<sub>2</sub>-fixation rates ( $\rho$ : 0.33 - 0.78,  $p < 0.05$ ) within each of the abundant moss species. Colors of the nodes represent the family membership of each phylotype. Phylotypes that were confirmed as potential N<sub>2</sub>-fixers are labeled with their unique identifier and asterisks (see Table S5 for details).

**Table 1** PERMANOVA results showing the relative importance of host-identity and site in structuring the microbial communities of the 26 moss species.

	Df	F	R <sup>2</sup>	P
Site	23	5.862	0.192	0.001
Species	25	4.829	0.172	0.001
Site:Species	117	1.306	0.217	0.001
Total	-	-	0.581	-

**Table S3:** List of sites and site characteristics. 10 sites around Fairbanks, AK, USA, and four sites around Toolik Lake, AK, USA were sampled in 2016 and an additional 10 sites from around Anchorage, AK, USA, were added in 2017. Sites were chosen to capture a breadth of common habitat types in boreal and tundra ecosystems

**Table S4:** Table summarizing the results of a mantel test between moss microbial communities and moss phylogenetic relatedness. Microbial communities are significantly correlated with phylogenetic structure. An analysis of Sphagnum shows that these results are primarily driven by phylogenetic structuring within the Sphagnum branch of the phylogeny.

**Figure S1:** The best environmental predictors shaping microbial communities for each of the seven most abundant host species.

**Table S7:** Results of PERMANOVAs run on eight individual sites (a subset of the 24 total sites that had >4 moss species present in >3 replicates) showing the effect of species in explaining microbial community differences.

**Table S8:** Results of PERMANOVAs run on seven individual species showing the effect of site in explaining microbial community differences.

Additional file 2 — Table S2

Supplementary Microsoft Excel Document (.xlsx) containing Table S2.

**Table S2:** A list of all of the moss species sampled and where and how many samples were collected for each. The species sampled include 26 moss species.

Additional file 3 — Table S5

Supplementary Microsoft Excel Document (.xlsx) containing Table S5.

**Table S5:** Summary of the evidence for N<sub>2</sub>-fixation for each of the 22 phylotypes positively correlated with N<sub>2</sub>-fixation rate.

Additional file 4 — Table S6

Supplementary Microsoft Excel Document (.xlsx) containing Table S6.

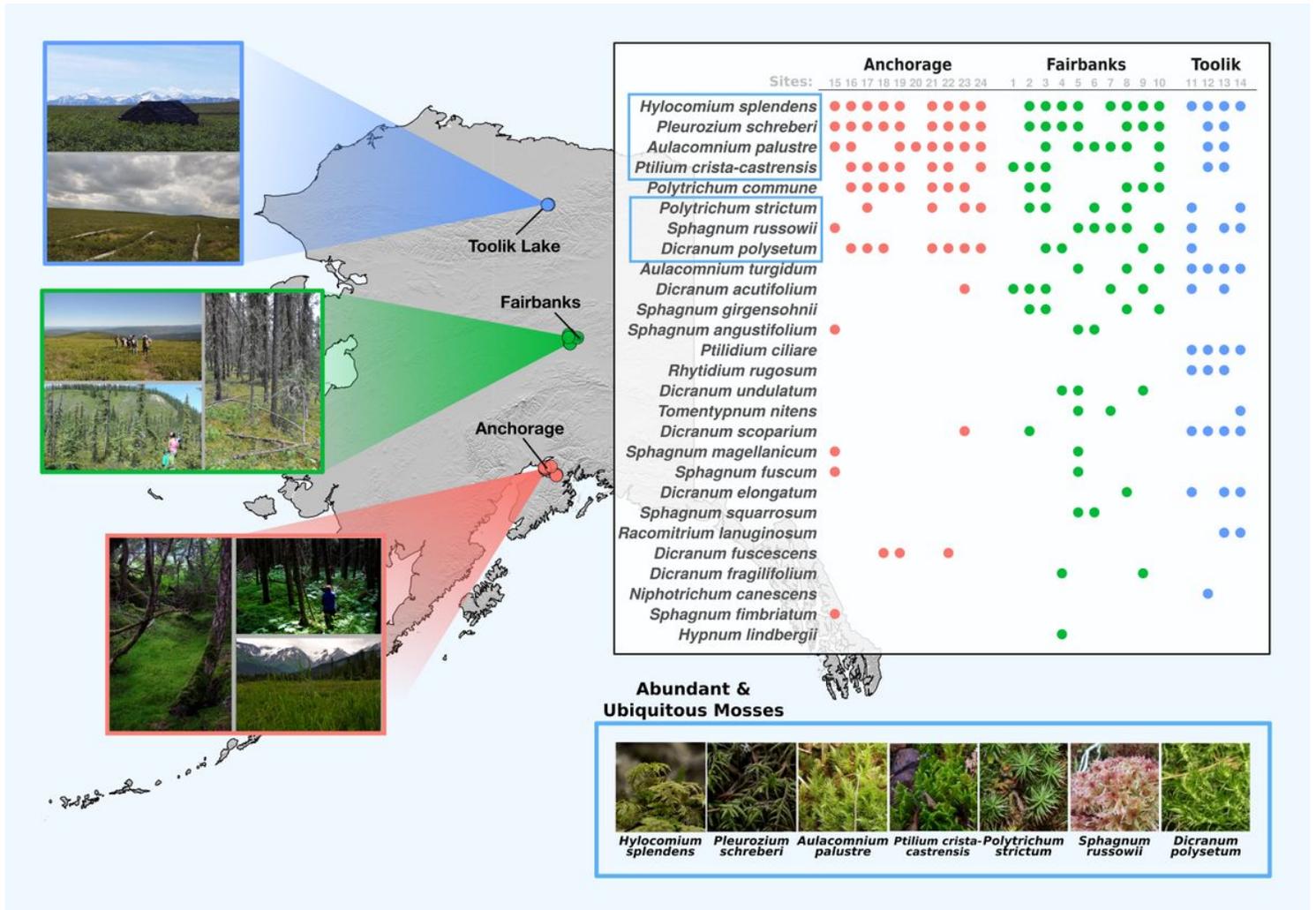
**Table S6:** Relative abundances of microbial families in each host species.

Additional file 5 — Table S9

Supplementary Microsoft Excel Document (.xlsx) containing Table S9.

**Table S9:** Table summarizing the findings from nine studies about prevalence and strength of host phylogenetic structuring of plant-microbial communities.

# Figures



**Figure 1**

A map summarizing the species and sampling locations of the 461 samples in this study. Images (left) from each transect location show the environmental heterogeneity across the sites. The table (top right) displays the list of species sampled, ordered by the total number of samples collected for each species (see Supplementary Tables S1 and S2). Dots indicate the sites where at least one sample of a particular species was collected. The blue boxes indicate the seven moss species identified as "abundant and ubiquitous" in our data set (see Methods). Images of these seven species are displayed at the bottom right corner of the figure. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

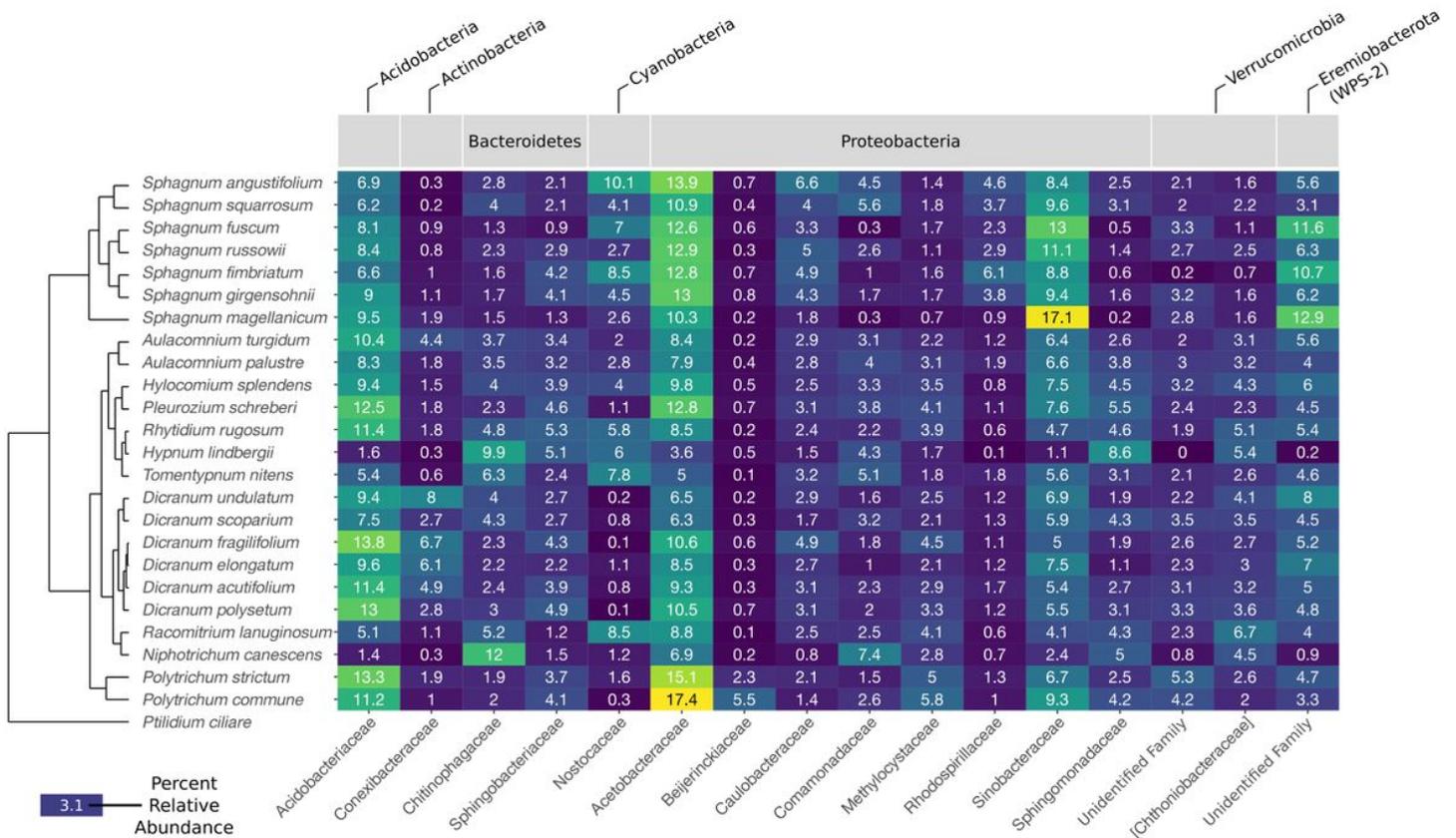


Figure 2

A heatmap showing the abundant bacterial families across different host species with a tree showing the phylogenetic relationships between the moss host species. Only 24 of the 26 moss species are included in this heatmap (see Methods). The outgroup host species, *Ptilidium ciliare*, is a liverwort.

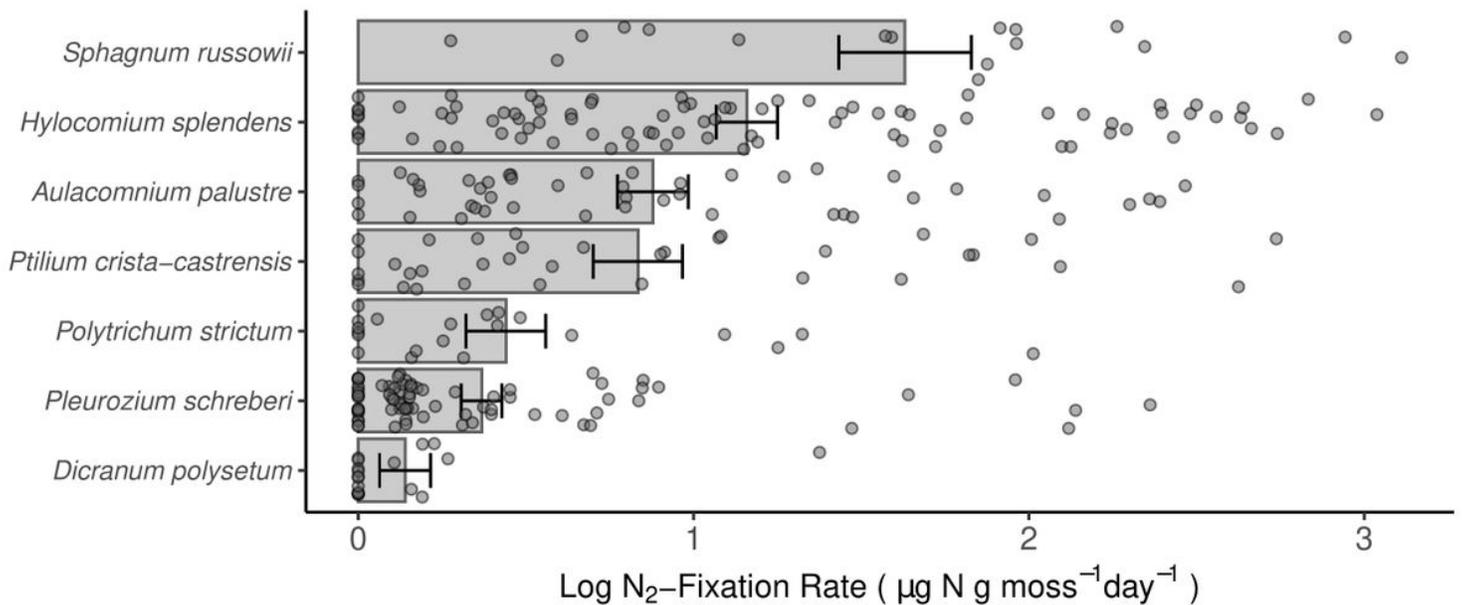
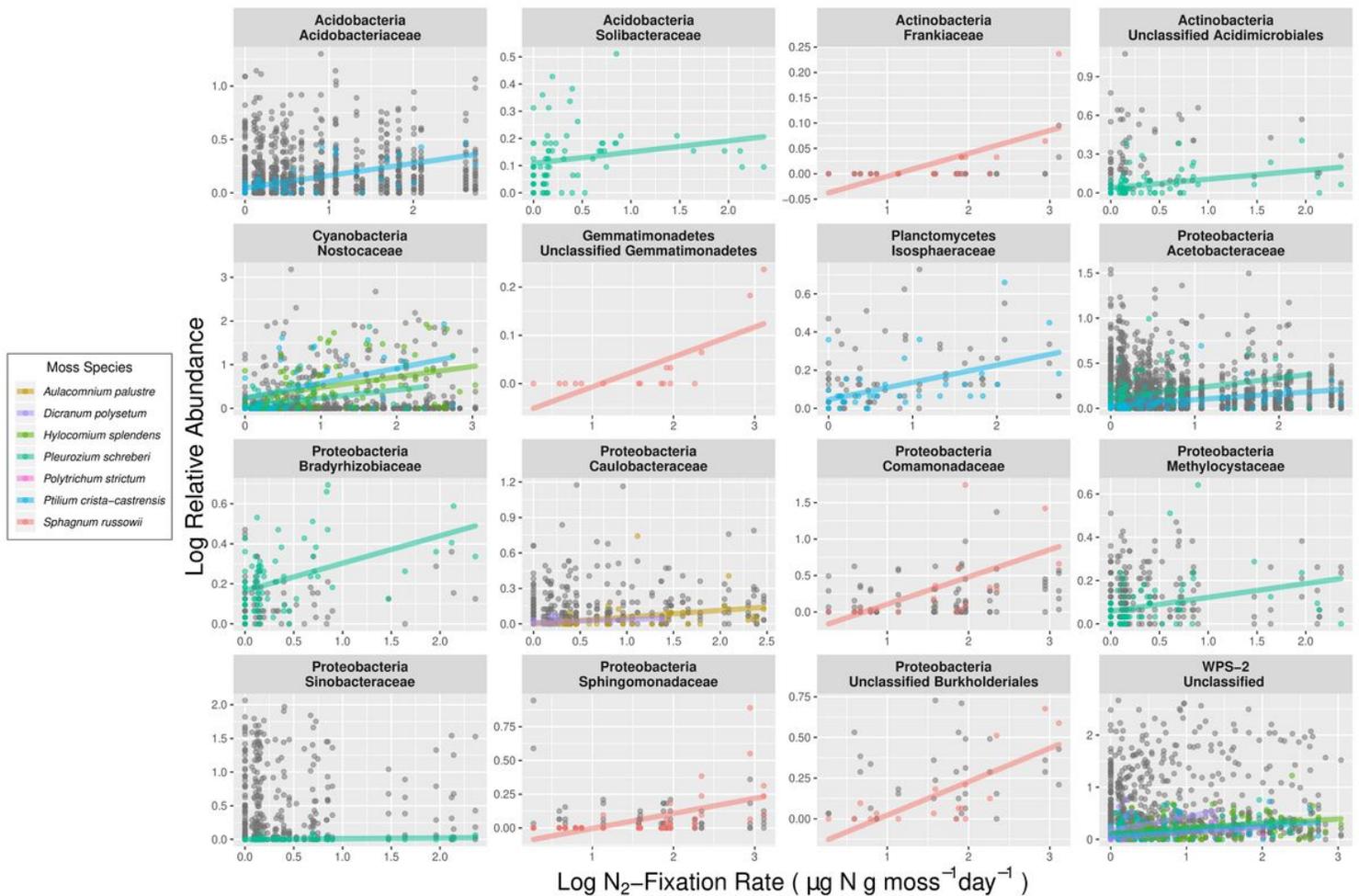


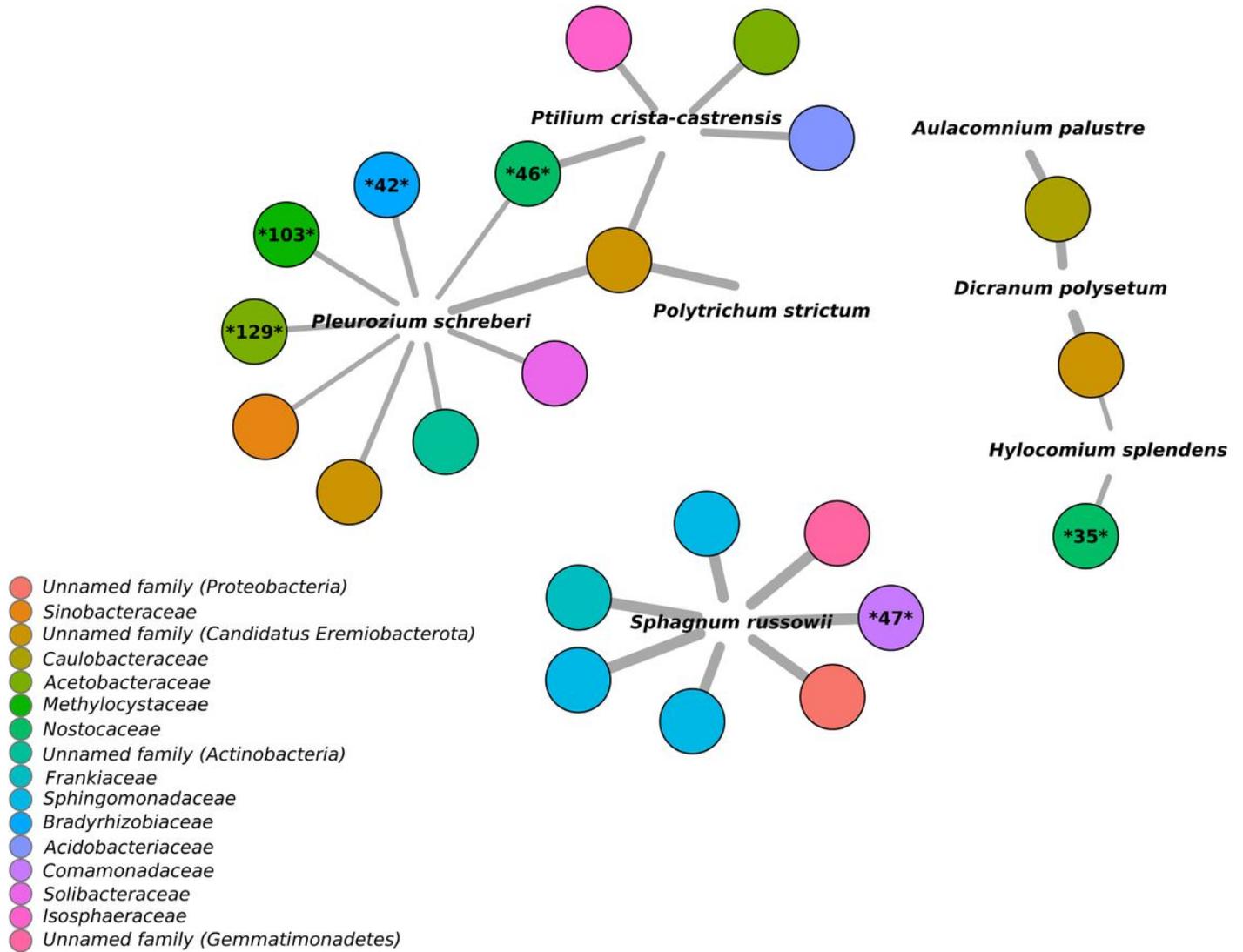
Figure 3

N<sub>2</sub>-fixation rates for the seven abundant and ubiquitous species in our data set as measured by isotope enrichment over a 1-day incubation period. Moss species are arranged from highest average fixation rate to lowest. All seven species hosted N<sub>2</sub>-fixing microbes however rates were highly variable across host species. Error bars represent 1 standard error above and below the mean. A more in-depth analysis of the differences in N<sub>2</sub>-fixation across moss species can be found in ref [28]



**Figure 4**

Spearman correlations between the relative abundances of individual bacterial phylotypes and measured N<sub>2</sub>-fixation rates. Bacterial phylotypes used in this analysis were limited to those that had an average relative abundance of >0.1% across the more abundant and ubiquitous moss host species (156 phylotypes). The data is separated into panels based on the family-level taxonomy of the phylotypes. Colored points and lines represent bacterial phylotypes that were significantly positively correlated with N<sub>2</sub>-fixation in a particular moss species. Lines and points are colored according to the moss species in question. Grey dots are measurements for 137 phylotypes that were not found to be significantly positively correlated with N<sub>2</sub>-fixation rates. We note that, to improve clarity, the x and y scales vary between panels.



**Figure 5**

Correlation network showing phylotypes (colored nodes) which showed significant positive Spearman correlations with N<sub>2</sub>-fixation rates ( $\rho$ : 0.33 - 0.78,  $p < 0.05$ ) within each of the abundant moss species. Colors of the nodes represent the family membership of each phylotype. Phylotypes that were confirmed as potential N<sub>2</sub>-Fixers are labeled with their unique identifier and asterisks (see Table S5 for details).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S09tableplantphylogeny papersclean.xlsx](#)
- [S06tablebacterialfamilyabundancesall.xlsx](#)
- [S05tablelsitannfixerclean.xlsx](#)

- [S02tablesamplerecountspersite.xlsx](#)
- [MossDim20172016papersuppfiguresV12formatted.docx](#)