Novel bacterial lineages associated with boreal moss species

Hannah Holland-Moritz, Julia Stuart, Lily R. Lewis, Samantha Miller, Michelle C. Mack, Stuart F. McDaniel, Noah Fierer

Affiliations:
1 Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, CO, USA
2 Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, Boulder, CO, USA
3 Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ USA
4 Department of Biology, University of Florida, Gainesville, FL 32611-8525, USA

* Corresponding Author

Abstract
Mosses are critical components of boreal ecosystems where they typically account for a large proportion of net primary productivity and harbor diverse bacterial communities that can be the major source of biologically-fixed nitrogen in these ecosystems. Despite their ecological importance, we have limited understanding of how microbial communities vary across boreal moss species and the extent to which local environmental conditions may influence the composition of these bacterial communities. We used marker gene sequencing to analyze bacterial communities associated with eight boreal moss species collected near Fairbanks, AK USA. We found that host identity was more important than site in determining bacterial community composition and that mosses harbor diverse lineages of potential N2-fixers as well as an abundance of novel taxa assigned to understudied bacterial phyla (including candidate phylum WPS-2). We performed shotgun metagenomic sequencing to assemble genomes from the WPS-2 candidate phylum and found that these moss-associated bacteria are likely anoxygenic phototrophs capable of carbon fixation via RuBisCo with an ability to utilize byproducts of photorespiration from hosts via a glyoxylate shunt. These results give new insights into the metabolic capabilities of understudied bacterial lineages that associate with mosses and the importance of plant hosts in shaping their microbiomes.
Introduction

Mosses, like their cousins, the vascular plants, associate with a broad diversity of microbes, including, bacteria, fungi, and other microbial eukaryotes (Lindo and Gonzalez, 2010). These moss-microbe associations are particularly relevant to terrestrial nitrogen (N) and carbon (C) cycling in northern ecosystems, where mosses are ubiquitously distributed and can be responsible for as much as 50% of ecosystem net primary productivity (Turetsky et al., 2012). Moss-associated N$_2$-fixing bacteria are often the primary source of ecosystem N inputs in boreal forests (DeLuca et al., 2002) and moss-associated microbes can have important influences on ecosystem C dynamics via methane oxidation, especially in peatlands, one of the largest natural sources of atmospheric methane (Kip et al., 2010). Together, the importance of moss-microbial communities to terrestrial biogeochemistry and unique features of bryophyte biology make boreal moss communities a useful system for investigating the interactions between host species identity, microbial community structure, and ecosystem function.

Mosses are ubiquitous across boreal forests, with distributions spanning ecologically important environmental gradients (Lindo and Gonzalez, 2010; Turetsky et al., 2012). Moss diversity in these forests can be quite high (Geffert et al., 2013), and different moss species often grow interspersed at a given location, creating many abundant and naturally occurring ‘common-garden’ experiments for testing how host identity and local environmental conditions influence the assembly of moss-associated microbial communities. Furthermore, mosses, unlike more commonly-studied vascular plants, have simpler leaf-tissue structures and do not possess roots thus representing a comparatively homogeneous host environment, reducing the need to control for inter-tissue spatial variation (a problem when studying vascular plants, e.g. Leff et al., 2015). Mosses are also small enough that an entire plant can be sampled for microbial analyses, an impossible task for most larger plants. Together these traits make mosses a useful study system for investigating the impacts of environmental and host factors on
microbial community structure and the contributions of these moss-associated microbial communities to ecosystem function.

Despite their potential biogeochemical importance and their utility as a model system, we still know surprisingly little about the structure and function of those microbial communities associated with boreal mosses. Much of the previous research has focused on a handful of abundant boreal moss species, in particular, members of the peat-moss genus *Sphagnum*, and the dominant feather mosses, *Pleurozium schreberi* and *Hylocomium splendens* (e.g. DeLuca *et al.*, 2002; Opelt *et al.*, 2007; Zackrisson *et al.*, 2009; Ininbergs *et al.*, 2011; Bragina, Berg, *et al.*, 2012; Bragina *et al.*, 2015). While these species are among the most abundant mosses in many boreal forests, the microbes living in association with less abundant hosts may be equally as important contributors to key ecosystem processes (Rousk *et al.*, 2015). Previous work on the abundant species makes it clear that boreal mosses possess surprisingly diverse microbial communities, containing not only well-studied *Cyanobacteria* but also novel and undescribed lineages within the *Alphaproteobacteria* sub-phylum and the *Verrucomicrobia* phylum among others (Bragina *et al.*, 2015). However, it is unclear whether local environmental factors govern the microbial community assembly, or whether a moss species hosts a characteristic microbial community, regardless of its abiotic environment.

It is well-known that mosses can harbor N$_2$-fixing bacteria and N$_2$-fixation rates of *Sphagnum* and the feather mosses, *Pleurozium schreberi* and *Hylocomium splendens*, can be quite variable across season (DeLuca *et al.*, 2002), forest type (Zackrisson *et al.*, 2004), and moss species (Leppänen *et al.*, 2015) with reported rates ranging from 0.3-4 kg N ha$^{-1}$ yr$^{-1}$ (DeLuca *et al.*, 2007). A number of previous studies have focused on selected moss-associated N$_2$-fixing bacterial taxa, particularly taxa within the *Cyanobacteria* phylum (reviewed in Rousk, Jones, *et al.*, 2013), and how the composition or diversity of these cyanobacteria relate to N$_2$-fixation rates (Ininbergs *et al.*, 2011). However, *Cyanobacteria* are unlikely to be the only moss-associated bacteria capable of N$_2$-fixation. For
example, Bragina et al. (2012) found that nitrogenase sequence libraries from bacteria on two *Sphagnum* species were dominated by alphaproteobacterial sequences, indicating that non-cyanobacterial N₂-fixers may play a more important role than previously recognized. In short, it remains unclear which bacteria are responsible for N₂-fixation in boreal mosses and how the abundances of these N₂-fixing taxa may vary across different moss species.

Here we characterized the bacterial communities associated with eight common boreal moss species at three different sites using marker gene (16S rRNA gene) amplicon sequencing. We used this dataset to address the following questions: 1) How do bacterial communities vary across different moss species and different sites? and 2) Which microbial taxa from these communities are potential N₂-fixers? Based on our observation that many of the taxa found in association with the moss species were representatives of bacterial lineages for which little is known (including members of the candidate phylum WPS-2), we then assembled genomes from shotgun metagenomic data to determine the functional attributes of these abundant, ubiquitous, and previously undescribed members of moss-associated microbial communities.

**Experimental Procedures**

**Sample Collection**

To characterize and compare the bacterial communities associated with different moss host species and their potential contributions to N₂-fixation, we collected samples of eight boreal moss species (the closely related pleurocarpous mosses *Pleurozium schreberi* (Brid.) Mitt., *Hylocomnium splendens* (Hedw.) Shimp, *Sphagnum capillifolium* (Ehrh.) Hedw.), the successively more distantly related species, *Aulacomnium palustre* (Hedw.) Schwägr. and *A. turgidum* (Wahlenb.) Schwägr., *Dicranum elongatum* Schleich. ex Schwägr., and the *Tomenthypnum nitens* (Hedw.) Loeske, and the successively more distantly related species, *Aulacomnium palustre* (Hedw.) Schwägr. and *A. turgidum* (Wahlenb.) Schwägr., *Dicranum elongatum* Schleich. ex Schwägr., and *Sphagnum capillifolium* (Ehrh.) Hedw.). The samples were collected from three black spruce (*Picea mariana*)-dominated sites that were
each at least a kilometer apart within the arboretum of the University of Alaska Fairbanks during late July 2014. The selected moss species are both common in boreal forests and represent lineages spanning the phylogenetic diversity of mosses. At each of the three sites we collected one sample for each of the eight species (except *Aulacomnium turgidum* and *Dicranum elongatum*, which could only be found at two sites), resulting in a total of 22 samples for analysis. Each sample was divided into four subsamples of ten stems each for microbial community analysis, isotopic enrichment, natural abundance voucher, and taxonomic voucher specimens (See Table S1). For each species within each site, a clump of ramets were removed from a monospecific patch and carefully sorted and cleaned with a gloved hand. Brown or decaying material was removed from the bottom so that each ramet was approximately 5 cm in length and included the apical meristem. Each sample was then divided into four subsamples of ten stems each for microbial community analysis, isotopic enrichment, natural abundance voucher, and taxonomic voucher specimens (See Table S1). Samples collected for microbial analyses were placed in a cooler on blue ice in the field and returned to the lab where they were frozen (-20°C) within 2 h of collection. Samples for the N\textsubscript{2}-fixation rate assays were also placed in the coolers until the measurements were started, usually within 2 h of sample collection. Natural abundance voucher specimens were returned to the lab and dried at 60°C for 48 h within 2 h of collection.

**N\textsubscript{2}-Fixation Measurements**

To quantify field rates of N\textsubscript{2}-fixation for each sample, we used an isotopic enrichment approach modified from Ruess *et al.* (2009, 2013). Each sample was placed in a 60 mL translucent polycarbonate syringe that was depressed to contain 10 ml of air. We then added 10 mL of \textsuperscript{15}N\textsubscript{2} (98% enriched, Cambridge Isotope Laboratories, Inc., U.S.A.) and sealed the syringe with a stopcock. Sealed syringes from all sites were placed in site A, where they were incubated for 24 h under similar light and temperature conditions. After 24 h, mosses were removed from the syringes and placed in a 60 °C oven for 48 h. Both enriched incubation samples and natural abundance samples were ground to a
homogenous powder. Nitrogen and carbon concentrations and atom% $^{15}$N and $^{13}$C values for both enriched incubation samples and natural abundance voucher specimens were measured on a Costech ECS4010 coupled to a Thermo Scientific Delta V Advantage Isotope ratio mass spectrometer in Mack’s lab at Northern Arizona University. Nitrogen fixation was calculated by comparing the $^{15}$N values from enriched and control samples (see Jean et al. (2017) for details).

**Amplicon-Based Bacterial Community Analysis**

To analyze the bacterial communities associated with each of the collected moss specimens, we PCR amplified and sequenced a portion of the bacterial and archaeal 16S rRNA marker gene. First, to minimize any bias from micro-spatial differences along the moss tissue, we homogenized each sample (ten moss stems per sample, approximately 0.25 g of tissue) with liquid N$_2$ under sterile conditions. We then extracted DNA from each homogenized moss sample using the MoBio Power Soil DNA extraction kit (MoBio Laboratories, Carlsbad, CA). After extracting DNA, we used the 515f / 806r primers to PCR amplify the V4-V5 region of the 16S rRNA gene (Caporaso et al., 2012). For each sample, we used a unique primer pair that included a 12-bp barcode and Illumina sequencing adapters to allow for multiplexed sequencing. To minimize amplification of mitochondrial and chloroplast DNA, we used PNA (peptide nucleic acid) clamps during PCR amplification (Lundberg et al., 2013). During both DNA extraction and PCR amplification, we included negative controls to check for potential contaminants introduced during those steps. We prepared the samples for sequencing by normalizing the concentrations of PCR products across all samples using the ThermoFisher Scientific SequalPrep Normalization plate (Thermo Fisher Scientific Inc. USA) and pooled the amplicons together. We sequenced all samples on the Illumina MiSeq platform running the 2 x 150 bp paired-end chemistry at the University of Colorado Next Generation Sequencing Facility. The *Hylocomium splendens* amplicon sequences were sequenced on a separate run. Because we cannot control for run-to-run variation we have excluded these samples from downstream statistical analyses. A cursory analysis
indicates that the *Hylocomniun* microbiome is statistically distinguishable from the other mosses in our sample, however the community was not marked by novel species relative to the other samples we analyzed.

We used the approach described by Leff *et al.* (2015) to analyze the 16S rRNA sequence data. Briefly, we removed adapters from the raw reads using cutadapt (paired, -O 1) (Martin, 2011) and demultiplexed the reads using a custom in-house python script (“prep_fastq_for_uparse_paired.py” at: [https://github.com/leffj/helper-code-for-uparse/](https://github.com/leffj/helper-code-for-uparse/)). Then, using USEARCH v.8 (Edgar, 2010), we merged, quality filtered (“maxee rate” = 0.005) and dereplicated the reads to create a fasta file containing each unique amplicon sequence. We also removed singletons (sequences found only once across the entire sample set) from the dereplicated reads. We created a de novo database from our sequences using the UPARSE pipeline (Edgar, 2013) (implemented with USEARCH v.8) by clustering reads at ≥97% sequence similarity level (which we refer to as ‘phylotypes’). As a quality-control measure, we removed representative sequences from phylotypes that were less than 75% similar to any sequence in the Greengenes database (version August 2013) (McDonald *et al.*, 2012). We assumed these highly divergent sequences to be chimeric, a product of non-specific amplification, or of insufficient quality.

To generate phylotype counts, we mapped the merged reads back to the *de novo* database (command “-usearch_global”, -id 0.97) and used a custom python script (“create_otu_table_from_uc_file.py” at: [https://github.com/leffj/helper-code-for-uparse/](https://github.com/leffj/helper-code-for-uparse/)) to generate a table of phylotype counts per sample from the USEARCH mapping output. We classified the reads with the RDP Naïve Bayesian classifier (Wang *et al.*, 2007) against the Greengenes database (McDonald *et al.*, 2012) and removed any residual chloroplast and mitochondrial sequences. To control for differences in read depth (sequencing “effort”) across samples, we randomly selected 6000 bacterial and archaeal 16S rRNA gene reads per sample prior to downstream analyses. The 6000 read cut-off
was chosen based on the read depth of the sample with fewest reads after mitochondria and chloroplast
removal. The filtered reads can be found on FigShare (DOI: https://doi.org/10.6084/m9.figshare.5594527).

To measure differences between communities across host species and sites, we calculated pairwise
Bray-Curtis dissimilarities on the square-root transformed phylotype table and tested for differences
between species and sites using a PERMANOVA test. All statistical analyses were conducted in R (R
Core Team, 2017, package “Vegan”, R Core Team, 2017). We used metaMDS (R package “Vegan”) to
generate NMDS plots from the dissimilarity matrix.

**Identifying Nearest Isolated Representatives and Classifying Putative N\(_2\)-fixing Bacteria**

To identify the nearest isolated representatives of the 30 most abundant bacterial phylotypes, we
used the RDP (Ribosomal Database Project) SeqMatch tool which finds the nearest relatives by
comparing the percent of shared sub-sequences between a query and the RDP isolate database (Cole et
al., 2014). We then downloaded the nearest-neighbor representative sequences from RDP and used
MUSCLE (Edgar, 2004) to create an alignment of the isolate sequences and representative sequences
from the 30 most abundant bacterial phylotypes, using FastTree (Price et al., 2010) to generate the final
tree. We defined “putative N\(_2\)-fixing phylotypes” as those phylotypes whose nearest relative (≥97%
similarity in the 16S rRNA gene region) has been found to be capable of N\(_2\)-fixation when cultivated in
isolation under laboratory conditions (four phylotypes). Alternatively, if the nearest neighbor was <
97% similar but came from a lineage which included only previously-described representative strains
known to be capable of N\(_2\)-fixation, the phylotype was also defined as a “putative N\(_2\)-fixer” (two
phylotypes).
Shotgun Metagenomic Sequencing and Analysis

Based on the preponderance of bacterial lineages identified from the 16S rRNA amplicon analyses that came from poorly described and novel bacterial lineages (see below), we generated shotgun metagenomic libraries from the samples of 3 moss species *Aulacomnium turgidum* (2 samples), *Pleurozium schreberi* (3 samples), and *Tomenthypnum nitens* (3 samples). We prepared the metagenomic libraries following the method described in Baym et al. (2015). All 8 libraries were then sequenced on the Illumina NextSeq platform running the 2 x 150 bp chemistry at the University of Colorado Next Generation Sequencing Facility. Each sample had an average of 58 million reads. After sequencing, we filtered the raw reads with Sickle (-q 20 -l 50) (Joshi and Fass, 2011) and used Metaxa2 (Bengtsson-Palme et al., 2015) on the filtered reads to verify that the 16S rRNA amplicon data was consistent with the taxonomic composition of the bacterial communities as inferred from the metagenomic data.

To assemble near complete genomes from the shotgun metagenomic data, we used the metagenomic de novo assembler, MEGAHIT (Li et al., 2014) (default settings: --min-count 2 --k-min 21 --k-max 99 --k-step 20) to co-assemble all the filtered reads without regard to sample origin (pooled-assembly). Then we sorted the assembled contigs into bins each representing a preliminary genome using the software MaxBin (Wu et al., 2014) (-min_contig_length 1000 -max_iteration 50 -prob_threshold 0.9 -markerset 40). Since MaxBin can take advantage of differences in abundance of identical contigs across samples, we were able to leverage differential abundances of reads from individual bacterial taxa across the pooled assembly to identify contigs belonging to the same organisms, thus facilitating the binning process. After binning, we used Bowtie2 (Langmead and Salzberg, 2012) to map the filtered reads back to each bin. We used CheckM (Parks et al., 2015) to verify the completeness and contamination of each bin (Table S2). Briefly, CheckM measures metagenomic completeness and contamination on the basis of presence and number of conserved single-copy marker genes (genes which typically are present in bacterial genomes only once). The
number of marker genes present compared to the number of expected marker genes for a particular bacterial lineage is used as a measure of completeness while the number of copies of a marker gene indicates contamination. We considered bins that were greater than 70% complete and less than 10% contaminated to be “high-quality” bins (as per the top two levels of quality identified in Parks et al. (2015)).

Because few of our high-quality bins contained assembled 16S rRNA sequences (a common problem, see Miller et al., 2011), we used three separate methods to confirm the taxonomic identity of bins. First, we used the program EMIRGE (Miller et al., 2011) to assemble full-length 16S rRNA sequences from the raw metagenomic reads and used USEARCH (Edgar, 2010) to match contigs from the bins to the EMIRGE-assembled full-length 16S rRNA sequences. Second, to verify this result with a technique that did not rely on genome assembly from metagenomic data, we used Metaxa2 (Bengtsson-Palme et al., 2015) to independently identify fragments of the 16S rRNA gene in the binned contigs and classified those fragments against the Greengenes database using USEARCH (Edgar, 2010). Finally, we used the automatically-generated concatenated marker gene phylogeny available from CheckM to confirm that bins were clustered on a tree in the appropriate clade as indicated by analysis of their 16S rRNA gene sequences. After identifying the organisms represented by the bins, we used the Joint Genome Institute’s Integrated Microbial Genomes (IMG) analysis pipeline (Markowitz et al., 2014) to identify and assign functions to the genes in the assembled bins. The assembled metagenomic reads are available through the IMG portal (IMG submission ID 115847).

Results and Discussion

Moss Species Identity Drives Microbiome Composition

Our analyses of moss-associated bacterial 16S rRNA amplicons show that mosses host diverse and species-specific bacterial communities. After quality filtering, but prior to chloroplast and
mitochondrial filtering, we obtained 7947 to 16590 16S rRNA gene reads per sample with an average sequencing depth of 11430 reads. The chloroplast and mitochondrial sequences that were removed made up only a small percentage of the reads (10% and 5%, respectively). Samples had an average richness of 924 phylotypes per sample and were dominated by eight bacterial phyla. Archaeal sequences were extremely rare and made up less than 0.001% of all reads. We found similar archaeal abundances in the primer-independent metagenomic data suggesting that the low archaeal abundance in the amplicon data was not a product of primer biases. The dominant bacterial phyla, as measured by average read abundance across all samples, were Proteobacteria (44.8% of reads across all samples), Acidobacteria (10.8%), Verrucomicrobia (9.8%), Bacteroidetes (9.3%), Cyanobacteria (6.5%), Candidate phyla WPS-2 (5.7%), Planctomycetes (5.2%), and Actinobacteria (4.2%) (Figure 1). The most abundant phylotypes identified across the entire sample set included those assigned to the Acetobacteraceae (9.5%), Acidobacteriaceae (8.2%), Sinobacteraceae (8.2%) and Nostocaceae (5.2%) families as well as many phylotypes that could only be classified to the phylum or class level of resolution (including those in the WPS-2, and Verrucomicrobia phyla, Figure 1). These phyla and families have been found in other microbial studies of mosses; notably Acetobacteraceae and Acidobacteriaceae were found to be abundant in sphagnum mosses (Bragina, Berg, et al., 2012) and Nostocaceae have long been studied in association with feather mosses (DeLuca et al., 2002) and Sphagnum (Bay et al., 2013; Kostka et al., 2016). Members of the WPS-2 and Verrucomicrobia phyla have previously been found in association with sphagnum mosses in bogs (Bragina et al., 2015).

Moss species identity was more important than site in determining the composition of the moss-associated bacterial communities. Across all samples, moss species contributed to 63% of the variation in community composition (Permanova $R^2 = 0.63$, $p < 0.001$), while site was not a significant source of variation (Permanova $R^2 = 0.076$, $p = 0.21$) (Figure 2). This finding is consistent with studies in vascular plants which find that species identity is generally more important than site in shaping phyllosphere community composition (Redford et al., 2010; Laforest-Lapointe et al., 2016). However,
sites in this study were situated within a relatively small area. It is possible that across larger scales, differences in local environmental conditions may have a stronger influence on the observed bacterial communities. Across all moss species studied here, Sanionia uncinata harbored particularly distinct bacterial communities. More than half (59%) of phylotypes found on Sanionia uncinata were not found in any of the other moss species and, of the 30 most abundant phylotypes highlighted in Table S3, only 77% were present on S. uncinata compared to more than 90% found across every other species (Figure 3). While most of the mosses were dominated by Acetobacteraceae (Alphaproteobacteria), Acidobacteriaceae (Acidobacteria), and Methylacidiphilales (Verrucomicrobia), the moss species Sanionia uncinata had a low abundance or complete absence of these taxa. Instead, Sanionia uncinata was dominated by Comamonadaceae (Betaproteobacteria), Nostocaceae (Cyanobacteria) and Chitinophagaceae (Bacteroidetes). The distinctiveness of the S. uncinata microbiome does not appear to reflect any aspect of phylogenetic relatedness among the sampled moss species. For instance, the pleurocarpous S. uncinata has much closer phylogenetic affinities with P. schreberi and T. nitens, yet the phylogenetically distant Sphagnum capillifolium hosts a microbiome much more similar to P. schreberi and T. nitens than does S. uncinata. The factors driving these differences between S. uncinata and the other moss species remain uncertain.

Evidence for multiple groups of N₂-fixing bacteria in boreal mosses

Using stable-isotope enrichment, we measured N₂-fixation rates and found that all of the moss species were actively fixing N₂ (Figure S1), though rates varied by a factor of 100 across samples (from 0.98 to 100 µg N • g dry weight moss⁻¹ day⁻¹). Importantly, we find that many moss microbiomes which were assumed to be non-N₂-fixing (the genus Dicranum, for example, Gundale et al. (2011) do, in fact, appear to associate with bacteria capable of fixing N₂ at rates comparable to those of feather mosses, which are well-known to host bacteria capable of sustaining relatively high rates of N₂-fixation (DeLuca et al., 2002; Turetsky et al., 2012). Consistent with other studies, Pleurozium schreberi and
Sanionia uncinata had the highest average rates of N$_2$-fixation at 46.1 and 52.4 µg N • g dry weight moss$^{-1}$ day$^{-1}$, respectively. Our measured rates of fixation are within the expected range previously reported for Sphagnum mosses and Pleurozium schreberi (Vile et al., 2014). However, because we used a δ$^{15}$N method as opposed to an acetylene reductase assay, and because of the lack of reported N$_2$ fixation studies for some of the moss species in our study, it is difficult to directly compare all of our measurements of N$_2$-fixation rates to those reported previously (e.g. DeLuca et al., 2002; Zackrisson et al., 2004; Gundale et al., 2012).

Given that all of the sampled moss species had microbial communities with measurable N$_2$-fixation activities, we then used the 16S rRNA gene survey data to determine which taxa might be contributing to the N$_2$-fixation. We found three cyanobacterial and three alphaproteobacterial lineages that were reasonably abundant across nearly all moss species, with all of these lineages having previously been identified as capable of N$_2$-fixation. Specifically, the putative N$_2$-fixers in our samples were from the Nostocaceae, Acetobacteraceae, Bradyrhizobiaceae, and Methylocystaceae families (Figure 3). Previous studies have focused on the role of Cyanobacteria in bryophyte-associated N$_2$-fixation, but one study in Sphagnum, found that 50% of bacterial cells colonizing Sphagnum were from Alphaproteobacteria and that the nifH gene libraries for these species were dominated by alphaproteobacterial, rather than cyanobacterial sequences (Bragina, Maier, et al., 2012). Interestingly, these two groups of putative N$_2$-fixers showed contrasting patterns of abundance between Sanionia uncinata and Pleurozium schreberi (the two moss species with the highest rates of N$_2$-fixation). These two hosts showed relatively similar average rates of N$_2$-fixation (52.4 µg N g$^{-1}$ day$^{-1}$ and 46.1 µgN g$^{-1}$ day$^{-1}$), yet Sanionia uncinata was dominated by Cyanobacteria (79%) and showed low relative abundances of the alphaproteobacterial N$_2$-fixers (7.2%), while Pleurozium schreberi showed the opposite pattern (6.8% cyanobacterial, 21.4% alphaproteobacterial). Together, these results suggest that multiple bacterial taxa may contribute to the measured N$_2$-fixation activity in these boreal moss species and, even though cyanobacterial N$_2$-fixers have received the bulk of the attention in previous studies (e.g. Ininbergs et al., 2011; Rousk, Rousk, et
al., 2013), they are not the only N₂-fixing bacteria that associate with boreal mosses. To identify which of these putative N₂-fixing lineages are responsible for the measured N₂-fixation activities in these moss species, future work using stable isotope probing-based approaches would be necessary (e.g. Buckley et al., 2007; Jehmlich et al., 2010).

**Boreal mosses harbor abundant, undescribed bacterial lineages**

The mosses we studied hosted an unexpected abundance of understudied bacterial lineages. Abundant phylotypes for which there were no close matches (<97% 16S rRNA sequence similarity) to previously cultivated and described bacteria are summarized in Table S1. Of these phylotypes, several of the most abundant phylotypes were from the phylum *Verrucomicrobia* and candidate phylum WPS-2. These phylotypes had no closely related cultivated representatives beyond the phylum-level of resolution, yet they were consistently among the most abundant phylotypes in many of the samples (Table S1, Figure S2). Further analysis of the metagenomic data confirmed the observed high relative abundances of WPS-2 and *Verrucomicrobia* and highlight that the abundances observed with the amplicon-based analyses were not a product of primer biases (Figure S3).

*Verrucomicrobia* are widely distributed in soils (Brewer et al., 2016), acidic geothermal environments (Op den Camp et al., 2009), and have also been found in boreal mosses, particularly in those mosses growing in moist, acidic environments such as bogs and peatlands (Dedysh, 2011; Sharp et al., 2014; Bragina et al., 2015). However, the ecologies of these moss-associated taxa remain poorly known. Among our samples, the most common verrucomicrobial divisions were the orders *Methylacidiphilales* (3.24%) and *Spartobacteria* (3.79%). One phylotype from *Methylacidiphilales*, was the third most abundant phylotype overall. Several earlier studies hypothesized that bacteria from *Methylacidiphilales* contribute to both N₂-fixation and methanotrophy in peatland environments (Bragina et al., 2015; Ho and Bodelier, 2015). We searched the assembled metagenomic contigs for the methanotrophic marker gene (*pmoA*), and we were unable to find any evidence of the methanotrophic
marker gene (\textit{pmoA}) associated with \textit{Methylacidiphilales} in our metagenomic sequence data. It is therefore unlikely that \textit{Methylacidiphilales} in these samples are oxidizing methane unless they are using genes that are not currently recognized in databases as being linked to methane oxidation.

In addition to relatively high abundances of poorly-described verrucomicrobial phylotypes, the mosses were also dominated by a single phylotype from the candidate phylum WPS-2 which represented 3.5\% of the 16S rRNA sequences analyzed (Figure 3), making it one of the most abundant phylotypes in six of the seven moss species (with the exception of \textit{Sanionia uncinata}). No members of WPS-2 have ever been cultured, and no published genome exists for any member of this candidate phylum. Given how little we know about this phylum, we can only speculate about the ecology of this group from a few previous studies in which members of this phylum have been detected.

Representatives from WPS-2 have previously been found in acidic and cold environments, including alpine bog vegetation (Bragina \textit{et al.}, 2015), mineral deposits from a low-temperature acidic spring (Grasby \textit{et al.}, 2013), and acidic natural gas extraction shale (Trexler \textit{et al.}, 2014). Interestingly, Trexler \textit{et al.} (2014) also observed that WPS-2 seems to co-occur with methanotrophs in aquatic mosses and speculated that WPS-2 may be using derivatives of methanotrophy such as carbon dioxide, formaldehyde, or formate. However, we found only two instances of the methane oxidation marker gene (\textit{pmoA}) in our shotgun metagenomes (out of more than 926 million reads), therefore it seems unlikely that the highly abundant members of the WPS-2 phylum found in our samples associate with methanotrophs.

\textbf{Draft genome of WPS-2 recovered from metagenomic data}

To learn more about the abundant and poorly studied taxa found in these mosses and to provide further insight into the functional attributes of these bacteria, we chose moss samples with particularly high abundances of diverse taxa of interest (\textit{Methylacidiphilae} and WPS-2) for metagenomic analysis. We assembled 61 genome bins from these moss-associated bacterial communities, seven of which
passed contamination and completeness standards (Table S2). Despite the apparent abundance of *Methylacidiphilae*, none of the recovered bins were from this phylum. The inability to recover highly abundant strains from metagenomes is fairly common and can be a product of high strain variation (intraspecific variation) (Miller et al., 2011).

Four phyla were represented in the recovered genome bins: *Acidobacteria* (1 bin), *Proteobacteria* (1 bin, *Alphaproteobacteria*), *Cyanobacteria* (2 bins), and WPS-2 (3 bins). The three bins from WPS-2 ranged from 84 – 89% complete and represent the first genomic data available for members of the candidate phylum WPS-2. We estimate that the full genome sizes for these WPS-2 representatives are 3.88, 3.41, and 4.34 Mbp based on the presence of lineage-specific single-copy marker genes.

Because WPS-2 is one of the most abundant groups across our samples and we know so little about the functional attributes of this group, we used gene annotation to try to reconstruct the potential metabolic attributes of the most complete genome bin obtained for WPS-2. The gene annotations revealed that WPS-2 is likely an anoxygenic phototroph, capable of carbon fixation, and able to metabolize the bi-products of photorespiration making it well-suited to life on the surface of a plant (Figure 4). We identified key genes involved in anoxygenic photosynthesis including those encoding the M and L subunits of anoxygenic photosynthetic reaction centers (*pufM* and *pufL*) as well as the gene for the Y subunit of chlorophylide a reductase (*bchY*), a universal marker gene for BChl-containing anoxygenic phototrophs (Yutin et al., 2005, 2009). Anoxygenic photosynthesis was first recognized for its importance in marine environments, but recently it has also been recognized as a common trait of phyllosphere bacteria found on the surface of vascular plants (Atamna-Ismaeel et al., 2013). There are three types of anoxygenic phototrophs that contain *pufM*-type reaction centers: purple sulfur bacteria, purple non-sulfur bacteria, and aerobic anoxygenic phototrophs (AAP) (Yurkov and Hughes, 2017). Of these three groups, we hypothesize that WPS-2 is an aerobic anoxygenic phototroph since abundant sulfur sources are expected to be extremely limited in boreal forests and the exposed
environment of the moss phyllosphere is unlikely to provide the anaerobic environment necessary to
support purple non-sulfur bacteria. If WPS-2 is truly an AAP, it would be one of the few observed
lineages to possess RuBisCo. Until recently, all other known AAP were thought incapable of RuBisCo-
facilitated carbon fixation (Hughes et al., 2017) however, a recent study (Graham et al., 2017) has
found four Alphaproteobacteria in a marine metagenomic dataset that appear to possess the ability to
fix carbon via the Calvin-Benson-Bassham cycle.

If WPS-2 is an AAP possessing RuBisCo, how might it live in close association to moss and their
cyanobacteria without being out-competed for light and carbon resources? WPS-2 appears to possess
several traits which may allow it to effectively associate with mosses. As an anoxygenic phototroph,
WPS-2 would not compete with its host for light resources since anoxygenic phototrophs absorb light
in a complementary spectrum to that of plants and cyanobacteria, with a maximal absorption peak
between 500 and 550nm (the region where chlorophyll has its minimum absorption) (Atamna-Ismaeel
et al., 2012). Interestingly, an absorption peak in this area of the spectrum may also give WPS-2 a
photosynthetic advantage under light snow cover since the wavelengths with the greatest transmittance
through snow are also in that region of the light spectrum (Perovich, 2007).

In addition to its complementary absorption spectrum, WPS-2 may be able to take advantage of the
byproducts of its host’s photorespiration to supplement its C requirements. Photorespiration takes place
when RuBisCo acts on O₂ rather than CO₂ producing phosphoglycolate, a potent inhibitor of the Calvin
cycle. Thus, in plants and other carbon fixers who use the Calvin cycle, metabolizing phosphoglycolate
is important. The moss-associated WPS-2 phylotypes appear to possess the enzymes necessary for
converting phosphoglycolate into glyoxylate and passing it into the TCA cycle. This process, known as
the “glyoxylate bypass”, is a modified TCA cycle that allows organisms to bypass the normal electron-
generating steps of the cycle and devote it entirely to biomass production (Kornberg, 1966). Since
WPS-2 is likely phototrophic, during times of high light it is unlikely to need its TCA cycle for energy

generation and could instead take advantage of its host’s sugars for biomass production. In such a situation, WPS-2 would only need to use RuBisCo under conditions when sugars from the plants were scarce, such as early in the growing season or under snow pack. The complementary absorption spectrum of the moss-associated members of the WPS-2 phylum and their ability to metabolize glyoxylate appear to make this group uniquely equipped for living commensally on their moss hosts.

**Conclusions**

We found that moss-associated bacterial communities are strongly structured by moss species identity with different moss species harboring distinct bacterial communities regardless of collection site. Surprisingly, all of the moss species studied possessed microbial communities that were capable of N\textsubscript{2} fixation and hosted a broad diversity of putative N\textsubscript{2}-fixing bacterial lineages. However, it remains unclear which lineages are responsible for the measured N\textsubscript{2} fixation rates and whether distinct bacterial lineages are driving N\textsubscript{2}-fixation in different moss species. Future work should focus on how microbial community variation between host species may influence N\textsubscript{2} fixation rates and how interactions both between bacteria and between bacteria and the moss host might contribute to these rates. We also found that mosses can harbor a number of poorly-described bacterial lineages, including a high relative abundance of bacteria assigned to the candidate phylum WPS-2 with previously unknown ecological and metabolic characteristics. Using shotgun metagenomic analyses, we were able to assemble a nearly complete genome representative of this WPS-2 lineage and the genomic analyses suggest that WPS-2 is an anoxygenic phototroph that is uniquely adapted to living in close-association with mosses in this ecosystem.

**Acknowledgements**

Financial support for this work was provided by the U.S. National Science Foundation Dimensions of Biodiversity program (DEB 1542609). The authors acknowledge infrastructural support provided by the University of Colorado Next Generation Sequencing Facility and the University of Alaska, Fairbanks. Helpful feedback on the manuscript was provided by the University of Colorado, EBIO writing co-op.
References


R Core Team (2017) R: A Language and Environment for Statistical Computing.


**Author Contributions**

The project was conceived by N.F., M.C.M, and S.F.M. The manuscript was written by H.H.M and N.F. with contributions from all co-authors. M.C.M. collected the samples. J.S. and S.M. performed the isotopic analysis. L.R.L. and S.F.M. identified the moss species and prepared the voucher specimens. H.H.M. and N.F. performed the molecular analyses.

**Conflict of Interest Statement**

The authors declare no conflict of interest in this study.

**Materials and Correspondence**

Requests for materials and correspondence should be addressed to N.F. or H.H.M.

**Figure Captions:**

**Figure 1:** Heat map showing the relative abundances of the most abundant bacterial families across the 7 moss species. Phylum-level classifications for each family are noted on the left. Higher relative abundances are indicated in brighter colors. Families with no official classification are noted as “Unclassified Family.” Host moss species are indicated along the x-axis and ordered according to overall bacterial community similarity.

**Figure 2:** Non-metric multidimensional scaling (NMDS) plot of bacterial community dissimilarities across samples (measured using the Bray-Curtis distance metric). Samples from the same host species tend to cluster together (A), while those from the same site were spread across ordination space (B). A PERMANOVA test revealed that species identity contributes to 63% of the variation in community composition ($R^2 = 0.63$, $p < 0.001$), while site is not a significant source of variation between these communities ($R^2 = 0.08$, $p = 0.21$).

**Figure 3:** A phylogenetic tree of the top 30 bacterial phylotypes (circles) and their $N_2$-fixing potential. Putative $N_2$-fixers (black circles) were identified by their relatedness to known $N_2$–fixers (see text).
The putative N\(_2\)-fixers observed in our moss samples include representatives from the cyanobacterial and alphaproteobacterial groups. The phylotypes represented here were chosen as they were consistently the most abundant phylotypes across all moss species and together account for 35% of 16S rRNA gene reads in the dataset. Each inner colored ring represents a different species of moss, from the inside ring to the outer ring they are: *Sanionia uncinata* (blue), *Dicranum elongatum* (yellow), *Aulacomnium turgidum* (bright red), *Pleurozium schreberi* (dark red), *Sphagnum capillifolium* (green), *Aulacomnium palustre* (purple), *Tomenthypnum nitens* (gray). The opacity of the colored rings represents the relative abundance of the different phylotypes (log transformed and scaled between 0 and 1) within each moss species. The black outer ring represents the total relative abundance of each phylotype across all host species (also log transformed and scaled between 0 and 1).

**Figure 4:** Diagram highlighting the inferred metabolic pathways linked to anoxygenic phototrophy and carbon cycling between the WPS-2 bacterial phylotype found to be abundant in mosses and the moss host cells. In the bacterial cell, light in a complementary spectrum to that absorbed by plants can be used to fuel cellular processes through anoxygenic phototrophy. CO\(_2\) fixation via RuBisCo in the bacterial cell may also occur, a novel feature for an aerobic anoxygenic phototroph. In the host cell, photorespiration (the action of RuBisCo on O\(_2\) rather than CO\(_2\)) can produce a Calvin Cycle-inhibiting byproduct, 2-phosphoglycolate, that the bacterial cell may have the ability to metabolize for biomass production though conversion to glyoxylate, a process known as the “glyoxylate bypass”. This process represents a possible mechanism for a mutualistic or commensal interactions between these WPS-2 bacteria and the moss host.
<table>
<thead>
<tr>
<th></th>
<th>Aulacomnium palustre</th>
<th>Tomentypnum nitens</th>
<th>Sphagnum capillifolium</th>
<th>Aulacomnium turgidum</th>
<th>Pleurozium schreberi</th>
<th>Dicranum elongatum</th>
<th>Sanionia uncinata</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proteobacteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetobacteraceae</td>
<td>0.058</td>
<td>0.069</td>
<td>0.128</td>
<td>0.125</td>
<td>0.135</td>
<td>0.164</td>
<td>0.013</td>
</tr>
<tr>
<td>Sinobacteraceae</td>
<td>0.093</td>
<td>0.079</td>
<td>0.076</td>
<td>0.094</td>
<td>0.098</td>
<td>0.135</td>
<td>0.014</td>
</tr>
<tr>
<td>Comamonadaceae</td>
<td>0.021</td>
<td>0.026</td>
<td>0.026</td>
<td>0.015</td>
<td>0.021</td>
<td>0.001</td>
<td>0.016</td>
</tr>
<tr>
<td>Caulobacteraceae</td>
<td>0.015</td>
<td>0.015</td>
<td>0.028</td>
<td>0.031</td>
<td>0.028</td>
<td>0.026</td>
<td>0.015</td>
</tr>
<tr>
<td>Sphingomonadaceae</td>
<td>0.009</td>
<td>0.017</td>
<td>0.007</td>
<td>0.016</td>
<td>0.012</td>
<td>0.002</td>
<td>0.083</td>
</tr>
<tr>
<td>Unclassified Family 1</td>
<td>0.031</td>
<td>0.018</td>
<td>0.019</td>
<td>0.012</td>
<td>0.004</td>
<td>0.017</td>
<td>0.003</td>
</tr>
<tr>
<td>Methylcystaceae</td>
<td>0.015</td>
<td>0.019</td>
<td>0.009</td>
<td>0.023</td>
<td>0.018</td>
<td>0.009</td>
<td>0.009</td>
</tr>
<tr>
<td>Burkholderiaceae</td>
<td>0.015</td>
<td>0.01</td>
<td>0.008</td>
<td>0.028</td>
<td>0.023</td>
<td>0.008</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>Rhodospirillaceae</td>
<td>0.015</td>
<td>0.012</td>
<td>0.025</td>
<td>0.01</td>
<td>0.01</td>
<td>0.014</td>
<td>0.004</td>
</tr>
<tr>
<td>Unclassified Family 4</td>
<td>0.015</td>
<td>0.016</td>
<td>0.018</td>
<td>0.008</td>
<td>0.008</td>
<td>0.006</td>
<td>0.014</td>
</tr>
<tr>
<td>Unclassified Family 3</td>
<td>0.011</td>
<td>0.012</td>
<td>0.01</td>
<td>0.012</td>
<td>0.018</td>
<td>0.005</td>
<td>0.008</td>
</tr>
<tr>
<td>Xanthomonadaceae</td>
<td>0.027</td>
<td>0.015</td>
<td>0.008</td>
<td>0.01</td>
<td>0.003</td>
<td>0.005</td>
<td>0.007</td>
</tr>
<tr>
<td>Coxiellaceae</td>
<td>0.012</td>
<td>0.006</td>
<td>0.016</td>
<td>0.013</td>
<td>0.003</td>
<td>0.022</td>
<td>0.001</td>
</tr>
<tr>
<td>Bradyrhizobiaceae</td>
<td>0.008</td>
<td>0.013</td>
<td>0.007</td>
<td>0.011</td>
<td>0.004</td>
<td>0.005</td>
<td>0.008</td>
</tr>
<tr>
<td>Unclassified Family 2</td>
<td>0.007</td>
<td>0.005</td>
<td>0.011</td>
<td>0.003</td>
<td>0.003</td>
<td>0.007</td>
<td>0.011</td>
</tr>
<tr>
<td>Hyphomicrobiaceae</td>
<td>0.005</td>
<td>0.008</td>
<td>0.008</td>
<td>0.001</td>
<td>&lt;0.005</td>
<td>0.006</td>
<td>0.122</td>
</tr>
<tr>
<td>Acidobacteriaeiae</td>
<td>0.063</td>
<td>0.056</td>
<td>0.052</td>
<td>0.129</td>
<td>0.158</td>
<td>0.122</td>
<td>0.006</td>
</tr>
<tr>
<td>Solibacteraceae</td>
<td>0.018</td>
<td>0.014</td>
<td>0.021</td>
<td>0.02</td>
<td>0.016</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>[Chthoniobacteraceae]</td>
<td>0.034</td>
<td>0.029</td>
<td>0.037</td>
<td>0.051</td>
<td>0.051</td>
<td>0.015</td>
<td>0.049</td>
</tr>
<tr>
<td>Unclassified Family 1</td>
<td>0.042</td>
<td>0.037</td>
<td>0.03</td>
<td>0.041</td>
<td>0.034</td>
<td>0.02</td>
<td>0.001</td>
</tr>
<tr>
<td>auto67_4W</td>
<td>0.032</td>
<td>0.022</td>
<td>0.02</td>
<td>0.003</td>
<td>&lt;0.005</td>
<td>0.009</td>
<td>0.002</td>
</tr>
<tr>
<td>Opitutaceae</td>
<td>0.009</td>
<td>0.011</td>
<td>0.014</td>
<td>0.004</td>
<td>0.002</td>
<td>0.011</td>
<td>0.008</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>0.035</td>
<td>0.054</td>
<td>0.036</td>
<td>0.022</td>
<td>0.017</td>
<td>0.015</td>
<td>0.087</td>
</tr>
<tr>
<td>Chitinophagaceae</td>
<td>0.027</td>
<td>0.029</td>
<td>0.015</td>
<td>0.033</td>
<td>0.058</td>
<td>0.019</td>
<td>0.025</td>
</tr>
<tr>
<td>Sphingobacteriaceae</td>
<td>0.005</td>
<td>0.011</td>
<td>0.011</td>
<td>0.001</td>
<td>&lt;0.005</td>
<td>0.003</td>
<td>0.046</td>
</tr>
<tr>
<td>Cytophagaceae</td>
<td>0.042</td>
<td>0.037</td>
<td>0.03</td>
<td>0.041</td>
<td>0.034</td>
<td>0.02</td>
<td>0.001</td>
</tr>
<tr>
<td>Cyanobacteria</td>
<td>0.091</td>
<td>0.069</td>
<td>0.046</td>
<td>0.028</td>
<td>0.039</td>
<td>0.001</td>
<td>0.082</td>
</tr>
<tr>
<td>Nostocaceae</td>
<td>0.015</td>
<td>0.009</td>
<td>0.007</td>
<td>0.011</td>
<td>0.001</td>
<td>0.013</td>
<td>0.001</td>
</tr>
<tr>
<td>WPS-2</td>
<td>0.064</td>
<td>0.065</td>
<td>0.06</td>
<td>0.077</td>
<td>0.075</td>
<td>0.062</td>
<td>0.001</td>
</tr>
<tr>
<td>Unclassified Family 1</td>
<td>0.018</td>
<td>0.017</td>
<td>0.022</td>
<td>0.018</td>
<td>0.019</td>
<td>0.025</td>
<td>0.01</td>
</tr>
<tr>
<td>Isophaeraceae</td>
<td>0.013</td>
<td>0.016</td>
<td>0.012</td>
<td>0.022</td>
<td>0.03</td>
<td>0.023</td>
<td>0.002</td>
</tr>
<tr>
<td>Gemmataceae</td>
<td>0.015</td>
<td>0.016</td>
<td>0.014</td>
<td>0.006</td>
<td>0.003</td>
<td>0.025</td>
<td>0.01</td>
</tr>
<tr>
<td>Conexibacteriaceae</td>
<td>0.004</td>
<td>0.005</td>
<td>0.001</td>
<td>0.012</td>
<td>0.11</td>
<td>0.03</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>Microbacteriaceae</td>
<td>0.004</td>
<td>0.013</td>
<td>0.004</td>
<td>0.009</td>
<td>0.012</td>
<td>0.003</td>
<td>0.007</td>
</tr>
</tbody>
</table>

The copyright holder for this preprint (which was posted November 16, 2017) is the author/funder. All rights reserved. No reuse allowed without permission.
Putative N₂-fixation

No evidence for N₂-fixation

Sanionia uncinata
Dicranum elongatum
Aulacomnium turgidum
Pleurozium schreberi
Sphagnum capillifolium
Aulacomnium palustre
Tomentypnum nitens

Total Mean Relative Abundance