bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Novel bacterial lineages associated with boreal moss species**

2 Hannah Holland-Moritz[1,2*], Julia Stuart[3], Lily R. Lewis[4], Samantha Miller[3], Michelle C. Mack[3], Stuart
3 F. McDaniel[4], Noah Fierer[1,2*]
4 Affiliations:
5 [1]Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder,
6 Boulder, CO, USA
7 [2]Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, Boulder, CO,
8 USA
9 [3]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ USA
10 [4]Department of Biology, University of Florida, Gainesville, FL 32611-8525, USA
11 [*]Corresponding Author
12

13 **Abstract**

14 Mosses are critical components of boreal ecosystems where they typically account for a large
15 proportion of net primary productivity and harbor diverse bacterial communities that can be the major
16 source of biologically-fixed nitrogen in these ecosystems. Despite their ecological importance, we have
17 limited understanding of how microbial communities vary across boreal moss species and the extent to
18 which local environmental conditions may influence the composition of these bacterial communities.
19 We used marker gene sequencing to analyze bacterial communities associated with eight boreal moss
20 species collected near Fairbanks, AK USA. We found that host identity was more important than site in
21 determining bacterial community composition and that mosses harbor diverse lineages of potential N_2-
22 fixers as well as an abundance of novel taxa assigned to understudied bacterial phyla (including
23 candidate phylum WPS-2). We performed shotgun metagenomic sequencing to assemble genomes from
24 the WPS-2 candidate phylum and found that these moss-associated bacteria are likely anoxygenic
25 phototrophs capable of carbon fixation via RuBisCo with an ability to utilize byproducts of
26 photorespiration from hosts via a glyoxylate shunt. These results give new insights into the metabolic
27 capabilities of understudied bacterial lineages that associate with mosses and the importance of plant
28 hosts in shaping their microbiomes.

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

29 **Introduction**

30 Mosses, like their cousins, the vascular plants, associate with a broad diversity of microbes,
31 including, bacteria, fungi, and other microbial eukaryotes (Lindo and Gonzalez, 2010). These moss-
32 microbe associations are particularly relevant to terrestrial nitrogen (N) and carbon (C) cycling in
33 northern ecosystems, where mosses are ubiquitously distributed and can be responsible for as much as
34 50% of ecosystem net primary productivity (Turetsky *et al.*, 2012). Moss-associated N_2-fixing bacteria
35 are often the primary source of ecosystem N inputs in boreal forests (DeLuca *et al.*, 2002) and moss-
36 associated microbes can have important influences on ecosystem C dynamics via methane oxidation,
37 especially in peatlands, one of the largest natural sources of atmospheric methane (Kip *et al.*, 2010).
38 Together, the importance of moss-microbial communities to terrestrial biogeochemistry and unique
39 features of bryophyte biology make boreal moss communities a useful system for investigating the
40 interactions between host species identity, microbial community structure, and ecosystem function.

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

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

52 microbial community structure and the contributions of these moss-associated microbial communities
53 to ecosystem function.

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

68 It is well-known that mosses can harbor N_2-fixing bacteria and N_2-fixation rates of *Sphagnum* and
69 the feather mosses, *Pleurozium schreberi* and *Hylocomium splendens,* can be quite variable across
70 season (DeLuca *et al.*, 2002), forest type (Zackrisson *et al.*, 2004), and moss species (Leppänen *et al.*,
71 2015) with reported rates ranging from 0.3-4 kg N ha^{-1} yr^{-1} (DeLuca *et al.*, 2007) . A number of
72 previous studies have focused on selected moss-associated N_2-fixing bacterial taxa, particularly taxa
73 within the *Cyanobacteria* phylum (reviewed in Rousk, Jones, *et al.*, 2013), and how the composition or
74 diversity of these cyanobacteria relate to N_2-fixation rates (Ininbergs *et al.*, 2011). However,
75 *Cyanobacteria* are unlikely to be the only moss-associated bacteria capable of N_2-fixation. For

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

76 example, Bragina *et al.* (2012) found that nitrogenase sequence libraries from bacteria on two
77 *Sphagnum* species were dominated by alphaproteobacterial sequences, indicating that non-
78 cyanobacterial N_2-fixers may play a more important role than previously recognized. In short, it
79 remains unclear which bacteria are responsible for N_2-fixation in boreal mosses and how the
80 abundances of these N_2-fixing taxa may vary across different moss species.

81 Here we characterized the bacterial communities associated with eight common boreal moss
82 species at three different sites using marker gene (16S rRNA gene) amplicon sequencing. We used this
83 dataset to address the following questions: 1) How do bacterial communities vary across different moss
84 species and different sites? and 2) Which microbial taxa from these communities are potential N_2-
85 fixers? Based on our observation that many of the taxa found in association with the moss species were
86 representatives of bacterial lineages for which little is known (including members of the candidate
87 phylum WPS-2), we then assembled genomes from shotgun metagenomic data to determine the
88 functional attributes of these abundant, ubiquitous, and previously undescribed members of moss-
89 associated microbial communities.

90 Experimental Procedures

91 Sample Collection

92 To characterize and compare the bacterial communities associated with different moss host species
93 and their potential contributions to N_2-fixation, we collected samples of eight boreal moss species (the
94 closely related pleurocarpous mosses *Pleurozium schreberi* (Brid.) Mitt., *Hylocomnium splendens*
95 (Hedw.) Shimp, *Sanionia uncinata* (Hedw.) Loeske, and *Tomenthypnum nitens* (Hedw.) Loeske, and the
96 successively more distantly related species, *Aulacomnium palustre* (Hedw.) Schwägr. and *A. turgidum*
97 (Wahlenb.) Schwägr., *Dicranum elongatum* Schleich. ex Schwägr., and *Sphagnum capillifolium* (Ehrh.)
98 Hedw.). The samples were collected from three black spruce (*Picea mariana*)-dominated sites that were

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

99 each at least a kilometer apart within the arboretum of the University of Alaska Fairbanks during late
100 July 2014. The selected moss species are both common in boreal forests and represent lineages
101 spanning the phylogenetic diversity of mosses. At each of the three sites we collected one sample for
102 each of the eight species (except *Aulacomnium turgidum* and *Dicranum elongatum*, which could only
103 be found at two sites), resulting in a total of 22 samples for analysis. Each sample was divided into four
104 subsamples of ten stems each for microbial community analysis, isotopic enrichment, natural
105 abundance voucher, and taxonomic voucher specimens (See Table S1). For each species within each
106 site, a clump of ramets were removed from a monospecific patch and carefully sorted and cleaned with
107 a gloved hand. Brown or decaying material was removed from the bottom so that each ramet was
108 approximately 5 cm in length and included the apical meristem. Each sample was then divided into
109 four subsamples of ten stems each for microbial community analysis, isotopic enrichment, natural
110 abundance voucher, and taxonomic voucher specimens (See Table S1). Samples collected for microbial
111 analyses were placed in a cooler on blue ice in the field and returned to the lab where they were frozen
112 (-20°C) within 2 h of collection. Samples for the N_2-fixation rate assays were also placed in the coolers
113 until the measurements were started, usually within 2 h of sample collection. Natural abundance
114 voucher specimens were returned to the lab and dried at 60°C for 48 h within 2 h of collection.

115 **N_2-Fixation Measurements**

116 To quantify field rates of N_2-fixation for each sample, we used an isotopic enrichment approach
117 modified from Ruess *et al.* (2009, 2013). Each sample was placed in a 60 mL translucent polycarbonate
118 syringe that was depressed to contain 10 ml of air. We then added 10 mL of $^{15}N_2$ (98% enriched,
119 Cambridge Isotope Laboratories, Inc., U.S.A.) and sealed the syringe with a stopcock. Sealed syringes
120 from all sites were placed in site A, where they were incubated for 24 h under similar light and
121 temperature conditions. After 24 h, mosses were removed from the syringes and placed in a 60 °C oven
122 for 48 h. Both enriched incubation samples and natural abundance samples were ground to a

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

123 homogenous powder. Nitrogen and carbon concentrations and atom% ^{15}N and ^{13}C values for both
124 enriched incubation samples and natural abundance voucher specimens were measured on a Costech
125 ECS4010 coupled to a Thermo Scientific Delta V Advantage Isotope ratio mass spectrometer in Mack's
126 lab at Northern Arizona University. Nitrogen fixation was calculated by comparing the ^{15}N values from
127 enriched and control samples (see Jean *et al.* (2017) for details).

128 **Amplicon-Based Bacterial Community Analysis**

129 To analyze the bacterial communities associated with each of the collected moss specimens, we
130 PCR amplified and sequenced a portion of the bacterial and archaeal 16S rRNA marker gene. First, to
131 minimize any bias from micro-spatial differences along the moss tissue, we homogenized each sample
132 (ten moss stems per sample, approximately 0.25 g of tissue) with liquid N_2 under sterile conditions. We
133 then extracted DNA from each homogenized moss sample using the MoBio Power Soil DNA extraction
134 kit (MoBio Laboratories, Carlsbad, CA). After extracting DNA, we used the 515f / 806r primers to
135 PCR amplify the V4-V5 region of the 16S rRNA gene (Caporaso *et al.*, 2012). For each sample, we
136 used a unique primer pair that included a 12-bp barcode and Illumina sequencing adapters to allow for
137 multiplexed sequencing. To minimize amplification of mitochondrial and chloroplast DNA, we used
138 PNA (peptide nucleic acid) clamps during PCR amplification (Lundberg *et al.*, 2013). During both
139 DNA extraction and PCR amplification, we included negative controls to check for potential
140 contaminants introduced during those steps. We prepared the samples for sequencing by normalizing
141 the concentrations of PCR products across all samples using the ThermoFisher Scientific SequalPrep
142 Normalization plate (Thermo Fisher Scientific Inc. USA) and pooled the amplicons together. We
143 sequenced all samples on the Illumina MiSeq platform running the 2 x 150 bp paired-end chemistry at
144 the University of Colorado Next Generation Sequencing Facility. The *Hylocomium splendens*
145 amplicon sequences were sequenced on a separate run. Because we cannot control for run-to-run
146 variation we have excluded these samples from downstream statistical analyses. A cursory analysis

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

147 indicates that the *Hylocomnium* microbiome is statistically distinguishable from the other mosses in our
148 sample, however the community was not marked by novel species relative to the other samples we
149 analyzed.

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

163 To generate phylotype counts, we mapped the merged reads back to the *de novo* database
164 (command "-usearch_global", -id 0.97) and used a custom python script
165 ("create_otu_table_from_uc_file.py" at: https://github.com/leffj/helper-code-for-uparse/) to generate a
166 table of phylotype counts per sample from the USEARCH mapping output. We classified the reads
167 with the RDP Naïve Bayesian classifier (Wang *et al.*, 2007) against the Greengenes database
168 (McDonald *et al.*, 2012) and removed any residual chloroplast and mitochondrial sequences. To control
169 for differences in read depth (sequencing "effort") across samples, we randomly selected 6000 bacterial
170 and archaeal 16S rRNA gene reads per sample prior to downstream analyses. The 6000 read cut-off

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

171 was chosen based on the read depth of the sample with fewest reads after mitochondria and chloroplast
172 removal. The filtered reads can be found on FigShare (DOI:
173 https://doi.org/10.6084/m9.figshare.5594527).

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

179 **Identifying Nearest Isolated Representatives and Classifying Putative N_2-fixing Bacteria**

180 To identify the nearest isolated representatives of the 30 most abundant bacterial phylotypes, we
181 used the RDP (Ribosomal Database Project) SeqMatch tool which finds the nearest relatives by
182 comparing the percent of shared sub-sequences between a query and the RDP isolate database (Cole *et*
183 *al.*, 2014). We then downloaded the nearest-neighbor representative sequences from RDP and used
184 MUSCLE (Edgar, 2004) to create an alignment of the isolate sequences and representative sequences
185 from the 30 most abundant bacterial phylotypes, using FastTree (Price *et al.*, 2010) to generate the final
186 tree. We defined “putative N_2-fixing phylotypes” as those phylotypes whose nearest relative (≥97%
187 similarity in the 16S rRNA gene region) has been found to be capable of N_2-fixation when cultivated in
188 isolation under laboratory conditions (four phylotypes). Alternatively, if the nearest neighbor was <
189 97% similar but came from a lineage which included only previously-described representative strains
190 known to be capable of N_2-fixation, the phylotype was also defined as a “putative N_2-fixer” (two
191 phylotypes).

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

192 **Shotgun Metagenomic Sequencing and Analysis**

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

204 To assemble near complete genomes from the shotgun metagenomic data, we used the
205 metagenomic *de novo* assembler, MEGAHIT (Li *et al.*, 2014) (default settings: --min-count 2 --k-min
206 21 --k-max 99 --k-step 20) to co-assemble all the filtered reads without regard to sample origin
207 (pooled-assembly). Then we sorted the assembled contigs into bins each representing a preliminary
208 genome using the software MaxBin (Wu *et al.*, 2014) (-min_contig_length 1000 -max_iteration 50
209 -prob_threshold 0.9 -markerset 40). Since MaxBin can take advantage of differences in abundance of
210 identical contigs across samples, we were able to leverage differential abundances of reads from
211 individual bacterial taxa across the pooled assembly to identify contigs belonging to the same
212 organisms, thus facilitating the binning process. After binning, we used Bowtie2 (Langmead and
213 Salzberg, 2012) to map the filtered reads back to each bin. We used CheckM (Parks *et al.*, 2015) to
214 verify the completeness and contamination of each bin (Table S2). Briefly, CheckM measures
215 metagenomic completeness and contamination on the basis of presence and number of conserved
216 single-copy marker genes (genes which typically are present in bacterial genomes only once). The

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

217 number of marker genes present compared to the number of expected marker genes for a particular
218 bacterial lineage is used as a measure of completeness while the number of copies of a marker gene
219 indicates contamination. We considered bins that were greater than 70% complete and less than 10%
220 contaminated to be "high-quality" bins (as per the top two levels of quality identified in Parks *et al.*
221 (2015).

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

236 Results and Discussion

237 Moss Species Identity Drives Microbiome Composition

238 Our analyses of moss-associated bacterial 16S rRNA amplicons show that mosses host diverse and
239 species-specific bacterial communities. After quality filtering, but prior to chloroplast and

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

259 Moss species identity was more important than site in determining the composition of the moss-
260 associated bacterial communities. Across all samples, moss species contributed to 63% of the variation
261 in community composition (Permanova $R^2 = 0.63$, $p < 0.001$), while site was not a significant source of
262 variation (Permanova $R^2 = 0.076$, $p = 0.21$) (Figure 2). This finding is consistent with studies in
263 vascular plants which find that species identity is generally more important than site in shaping
264 phyllosphere community composition (Redford *et al.*, 2010; Laforest-Lapointe *et al.*, 2016). However,

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

265 sites in this study were situated within a relatively small area. It is possible that across larger scales,
266 differences in local environmental conditions may have a stronger influence on the observed bacterial
267 communities. Across all moss species studied here, *Sanionia uncinata* harbored particularly distinct
268 bacterial communities. More than half (59%) of phylotypes found on *Sanionia uncinata* were not found
269 in any of the other moss species and, of the 30 most abundant phylotypes highlighted in Table S3, only
270 77% were present on *S. uncinata* compared to more than 90% found across every other species (Figure
271 3). While most of the mosses were dominated by *Acetobacteraceae* (*Alphaproteobacteria*),
272 *Acidobacteriaceae* (*Acidobacteria*), and *Methylacidiphilales* (*Verrucomicrobia*), the moss species
273 *Sanionia uncinata* had a low abundance or complete absence of these taxa. Instead, *Sanionia uncinata*
274 was dominated by *Comamonadaceae* (*Betaproteobacteria*), *Nostocacaceae* (*Cyanobacteria*) and
275 *Chitinophagageae* (*Bacteroidetes*). The distinctiveness of the *S. uncinata* microbiome does not appear
276 to reflect any aspect of phylogenetic relatedness among the sampled moss species. For instance, the
277 pleurocarpous *S. uncinata* has much closer phylogenetic affinities with *P. schreberi* and *T. nitens*, yet
278 the phylogenetically distant *Sphangnum capillifolium* hosts a microbiome much more similar to *P.*
279 *schreberi* and *T. nitens* than does *S. uncinata*. The factors driving these differences between *S. uncinata*
280 and the other moss species remain uncertain.

281 **Evidence for multiple groups of N_2-fixing bacteria in boreal mosses**

282 Using stable-isotope enrichment, we measured N_2-fixation rates and found that all of the moss
283 species were actively fixing N_2 (Figure S1), though rates varied by a factor of 100 across samples (from
284 0.98 to 100 µg N • g dry weight moss^{-1} day^{-1}). Importantly, we find that many moss microbiomes which
285 were assumed to be non-N_2-fixing (the genus *Dicranum*, for example, Gundale *et al.* (2011) do, in fact,
286 appear to associate with bacteria capable of fixing N_2 at rates comparable to those of feather mosses,
287 which are well-known to host bacteria capable of sustaining relatively high rates of N_2-fixation
288 (DeLuca *et al.*, 2002; Turetsky *et al.*, 2012). Consistent with other studies, *Pleurozium schreberi* and

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

289 *Sanionia uncinata* had the highest average rates of N_2-fixation at 46.1 and 52.4 µg N • g dry weight
290 moss^{-1} day^{-1}, respectively. Our measured rates of fixation are within the expected range previously
291 reported for *Sphagnum* mosses and *Pleurozium schreberi* (Vile *et al.*, 2014). However, because we used
292 a $\partial^{15}N$ method as opposed to an acetylene reductase assay, and because of the lack of reported N_2
293 fixation studies for some of the moss species in our study, it is difficult to directly compare all of our
294 measurements of N_2-fixation rates to those reported previously (e.g. DeLuca *et al.*, 2002; Zackrisson *et*
295 *al.*, 2004; Gundale *et al.*, 2012).

296 Given that all of the sampled moss species had microbial communities with measurable N_2-fixation
297 activities, we then used the 16S rRNA gene survey data to determine which taxa might be contributing
298 to the N_2-fixation. We found three cyanobacterial and three alphaproteobacterial lineages that were
299 reasonably abundant across nearly all moss species, with all of these lineages having previously been
300 identified as capable of N_2-fixation. Specifically, the putative N_2-fixers in our samples were from the
301 *Nostocaeae*, *Acetobacteraceae*, *Bradyrhizobiaceae*, and *Methylocystaceae* families (Figure 3). Previous
302 studies have focused on the role of *Cyanobacteria* in bryophyte-associated N_2- fixation, but one study
303 in *Sphagnum*, found that 50% of bacterial cells colonizing *Sphagnum* were from *Alphaproteobacteria*
304 and that the *nifH* gene libraries for these species were dominated by alphaproteobacterial, rather than
305 cyanobacterial sequences (Bragina, Maier, *et al.*, 2012). Interestingly, these two groups of putative N_2-
306 fixers showed contrasting patterns of abundance between *Sanionia uncinata* and *Pluerozium schreberi*
307 (the two moss species with the highest rates of N_2-fixation). These two hosts showed relatively similar
308 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
309 dominated by *Cyanobacteria* (79%) and showed low relative abundances of the alphaproteobacterial
310 N_2-fixers (7.2%), while *Pleurozium schreberi* showed the opposite pattern (6.8% cyanobacterial, 21.4%
311 alphaproteobacterial). Together, these results suggest that multiple bacterial taxa may contribute to the
312 measured N_2-fixation activity in these boreal moss species and, even though cyanobacterial N_2-fixers
313 have received the bulk of the attention in previous studies (e.g. Ininbergs *et al.*, 2011; Rousk, Rousk, *et*

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

314 *al.*, 2013), they are not the only N_2-fixing bacteria that associate with boreal mosses. To identify which
315 of these putative N_2-fixing lineages are responsible for the measured N_2-fixation activities in these
316 moss species, future work using stable isotope probing-based approaches would be necessary (e.g.
317 Buckley *et al.*, 2007; Jehmlich *et al.*, 2010).

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

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

328 *Verrucomicrobia* are widely distributed in soils (Brewer *et al.*, 2016), acidic geothermal
329 environments (Op den Camp *et al.*, 2009), and have also been found in boreal mosses, particularly in
330 those mosses growing in moist, acidic environments such as bogs and peatlands (Dedysh, 2011; Sharp
331 *et al.*, 2014; Bragina *et al.*, 2015). However, the ecologies of these moss-associated taxa remain poorly
332 known. Among our samples, the most common verrucomicrobial divisions were the orders
333 *Methylacidiphilales* (3.24%) and *Spartobacteria* (3.79%). One phylotype from *Methylacidiphilales*,
334 was the third most abundant phylotype overall. Several earlier studies hypothesized that bacteria from
335 *Methylacidiphilales* contribute to both N_2-fixation and methanotrophy in peatland environments
336 (Bragina *et al.*, 2015; Ho and Bodelier, 2015). We searched the assembled metagenomic contigs for the
337 methanotrophic marker gene (*pmoA*), and we were unable to find any evidence of the methanotrophic

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

338 marker gene (*pmoA*) associated with *Methylacidiphilales* in our metagenomic sequence data. It is
339 therefore unlikely that *Methylacidiphilales* in these samples are oxidizing methane unless they are
340 using genes that are not currently recognized in databases as being linked to methane oxidation.

341 In addition to relatively high abundances of poorly-described verrucomicrobial phylotypes, the
342 mosses were also dominated by a single phylotype from the candidate phylum WPS-2 which
343 represented 3.5% of the 16S rRNA sequences analyzed (Figure 3), making it one of the most abundant
344 phylotypes in six of the seven moss species (with the exception of *Sanionia uncinata*). No members of
345 WPS-2 have ever been cultured, and no published genome exists for any member of this candidate
346 phylum. Given how little we know about this phylum, we can only speculate about the ecology of this
347 group from a few previous studies in which members of this phylum have been detected.
348 Representatives from WPS-2 have previously been found in acidic and cold environments, including
349 alpine bog vegetation (Bragina *et al.*, 2015), mineral deposits from a low-temperature acidic spring
350 (Grasby *et al.*, 2013), and acidic natural gas extraction shale (Trexler *et al.*, 2014). Interestingly, Trexler
351 *et al.* (2014) also observed that WPS-2 seems to co-occur with methanotrophs in aquatic mosses and
352 speculated that WPS-2 may be using derivatives of methanotrophy such as carbon dioxide,
353 formaldehyde, or formate. However, we found only two instances of the methane oxidation marker
354 gene (*pmoA*) in our shotgun metagenomes (out of more than 926 million reads), therefore it seems
355 unlikely that the highly abundant members of the WPS-2 phylum found in our samples associate with
356 methanotrophs.

357 **Draft genome of WPS-2 recovered from metagenomic data**

358 To learn more about the abundant and poorly studied taxa found in these mosses and to provide
359 further insight into the functional attributes of these bacteria, we chose moss samples with particularly
360 high abundances of diverse taxa of interest (*Methylacidiphilae* and WPS-2) for metagenomic analysis.
361 We assembled 61 genome bins from these moss-associated bacterial communities, seven of which

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

362 passed contamination and completeness standards (Table S2). Despite the apparent abundance of
363 *Methylacidiphilae*, none of the recovered bins were from this phylum. The inability to recover highly
364 abundant strains from metagenomes is fairly common and can be a product of high strain variation
365 (intraspecific variation) (Miller *et al.*, 2011).

366 Four phyla were represented in the recovered genome bins: *Acidobacteria* (1 bin), *Proteobacteria*
367 (1 bin, *Alphaproteobacteria*), *Cyanobacteria* (2 bins), and WPS-2 (3 bins). The three bins from WPS-2
368 ranged from 84 – 89% complete and represent the first genomic data available for members of the
369 candidate phylum WPS-2. We estimate that the full genome sizes for these WPS-2 representatives are
370 3.88, 3.41, and 4.34 Mbp based on the presence of lineage-specific single-copy marker genes.

371 Because WPS-2 is one of the most abundant groups across our samples and we know so little about
372 the functional attributes of this group, we used gene annotation to try to reconstruct the potential
373 metabolic attributes of the most complete genome bin obtained for WPS-2. The gene annotations
374 revealed that WPS-2 is likely an anoxygenic phototroph, capable of carbon fixation, and able to
375 metabolize the bi-products of photorespiration making it well-suited to life on the surface of a plant
376 (Figure 4). We identified key genes involved in anoxygenic photosynthesis including those encoding
377 the M and L subunits of anoxygenic photosynthetic reaction centers (*pufM* and *pufL*) as well as the
378 gene for the Y subunit of chlorophyllide a reductase (*bchY*), a universal marker gene for BChl-
379 containing anoxygenic phototrophs (Yutin *et al.*, 2005, 2009). Anoxygenic photosynthesis was first
380 recognized for its importance in marine environments, but recently it has also been recognized as a
381 common trait of phyllosphere bacteria found on the surface of vascular plants (Atamna-Ismaeel *et al.*,
382 2013). There are three types of anoxygenic phototrophs that contain *pufM*-type reaction centers: purple
383 sulfur bacteria, purple non-sulfur bacteria, and aerobic anoxygenic phototrophs (AAP) (Yurkov and
384 Hughes, 2017). Of these three groups, we hypothesize that WPS-2 is an aerobic anoxygenic phototroph
385 since abundant sulfur sources are expected to be extremely limited in boreal forests and the exposed

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

386 environment of the moss phyllosphere is unlikely to provide the anaerobic environment necessary to
387 support purple non-sulfur bacteria. If WPS-2 is truly an AAP, it would be one of the few observed
388 lineages to possess RuBisCo. Until recently, all other known AAP were thought incapable of RuBisCo-
389 facilitated carbon fixation (Hughes *et al.*, 2017) however, a recent study (Graham *et al.*, 2017) has
390 found four Alphaproteobacteria in a marine metagenomic dataset that appear to possess the ability to
391 fix carbon via the Calvin-Benson-Bassham cycle.

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

401 In addition to its complementary absorption spectrum, WPS-2 may be able to take advantage of the
402 byproducts of its host's photorespiration to supplement its C requirements. Photorespiration takes place
403 when RuBisCo acts on O_2 rather than CO_2 producing phosphoglycolate, a potent inhibitor of the Calvin
404 cycle. Thus, in plants and other carbon fixers who use the Calvin cycle, metabolizing phosphoglycolate
405 is important. The moss-associated WPS-2 phylotypes appear to possess the enzymes necessary for
406 converting phosphoglycolate into glyoxylate and passing it into the TCA cycle. This process, known as
407 the "glyoxylate bypass", is a modified TCA cycle that allows organisms to bypass the normal electron-
408 generating steps of the cycle and devote it entirely to biomass production (Kornberg, 1966). Since
409 WPS-2 is likely phototrophic, during times of high light it is unlikely to need its TCA cycle for energy

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

415 **Conclusions**

416 We found that moss-associated bacterial communities are strongly structured by moss species
417 identity with different moss species harboring distinct bacterial communities regardless of collection
418 site. Surprisingly, all of the moss species studied possessed microbial communities that were capable of
419 N_2-fixation and hosted a broad diversity of putative N_2-fixing bacterial lineages. However, it remains
420 unclear which lineages are responsible for the measured N_2-fixation rates and whether distinct bacterial
421 lineages are driving N_2-fixation in different moss species. Future work should focus on how microbial
422 community variation between host species may influence N_2-fixation rates and how interactions both
423 between bacteria and between bacteria and the moss host might contribute to these rates. We also found
424 that mosses can harbor a number of poorly-described bacterial lineages, including a high relative
425 abundance of bacteria assigned to the candidate phylum WPS-2 with previously unknown ecological
426 and metabolic characteristics. Using shotgun metagenomic analyses, we were able to assemble a nearly
427 complete genome representative of this WPS-2 lineage and the genomic analyses suggest that WPS-2 is
428 an anoxygenic phototroph that is uniquely adapted to living in close-association with mosses in this
429 ecosystem.

430 **Acknowledgements**
431 Financial support for this work was provided by the U.S. National Science Foundation Dimensions of
432 Biodiversity program (DEB 1542609). The authors acknowledge infrastructural support provided by
433 the University of Colorado Next Generation Sequencing Facility and the University of Alaska,
434 Fairbanks. Helpful feedback on the manuscript was provided by the University of Colorado, EBIO
435 writing co-op.
436

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

437 **References**

438 Atamna-Ismaeel, N., Finkel, O., Glaser, F., von Mering, C., Vorholt, J.A., Koblížek, M., et al. (2012)
439 Bacterial anoxygenic photosynthesis on plant leaf surfaces. *Environ. Microbiol. Rep.* **4**: 209–216.

440 Atamna-Ismaeel, N., Finkel, O.M., Glaser, F., Sharon, I., Schneider, R., Post, A.F., et al. (2013)
441 Microbial rhodopsins on leaf surfaces of terrestrial plants. *Environ. Microbiol.* **14**: 140–146.

442 Bay, G., Nahar, N., Oubre, M., Whitehouse, M.J., Wardle, D.A., Zackrisson, O., et al. (2013) Boreal
443 feather mosses secrete chemical signals to gain nitrogen. *New Phytol.* **200**: 54–60.

444 Baym, M., Kryazhimskiy, S., Lieberman, T.D., Chung, H., Desai, M.M., and Kishony, R.K. (2015)
445 Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* **10**: 1–15.

446 Bengtsson-Palme, J., Hartmann, M., Eriksson, K.M., Pal, C., Thorell, K., Larsson, D.G.J., and Nilsson,
447 R.H. (2015) metaxa2: Improved identification and taxonomic classification of small and large
448 subunit rRNA in metagenomic data. *Mol. Ecol. Resour.* **15**: 1403–1414.

449 Bragina, A., Berg, C., and Berg, G. (2015) The core microbiome bonds the Alpine bog vegetation to a
450 transkingdom metacommunity. *Mol. Ecol.* **24**: 4795–4807.

451 Bragina, A., Berg, C., Cardinale, M., Shcherbakov, A., Chebotar, V., and Berg, G. (2012) *Sphagnum*
452 mosses harbour highly specific bacterial diversity during their whole lifecycle. *ISME J.* **6**: 802–13.

453 Bragina, A., Maier, S., Berg, C., Müller, H., Chobot, V., Hadacek, F., and Berg, G. (2012) Similar
454 diversity of Alphaproteobacteria and nitrogenase gene amplicons on two related *Sphagnum*
455 mosses. *Front. Microbiol.* **2**: 1–10.

456 Brewer, T., Handley, K., Carini, P., Gibert, J., and Fierer, N. (2016) Genome reduction in an abundant
457 and ubiquitous soil bacterium "*Candidatus Udaeobacter copiosus*." *Nat. Microbiol.* **2**: 16198.

458 Buckley, D.H., Huangyutitham, V., Hsu, S.F., and Nelson, T.A. (2007) Stable isotope probing with ^{15}N2
459 reveals novel noncultivated diazotrophs in soil. *Appl. Environ. Microbiol.* **73**: 3196–3204.

460 Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012) Ultra-
461 high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME*
462 *J.* **6**: 1621–1624.

463 Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., et al. (2014) Ribosomal Database
464 Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42**: 633–642.

465 Dedysh, S.N. (2011) Cultivating uncultured bacteria from northern wetlands: Knowledge gained and
466 remaining gaps. *Front. Microbiol.* **2**: 1–15.

467 DeLuca, T.H., Zackrisson, O., Gentili, F., Sellstedt, A., and Nilsson, M.C. (2007) Ecosystem controls
468 on nitrogen fixation in boreal feather moss communities. *Oecologia* **152**: 121–130.

469 DeLuca, T.H., Zackrisson, O., Nilsson, M.-C., and Sellstedt, A. (2002) Quantifying nitrogen-fixation in
470 feather moss carpets of boreal forests. *Nature* **419**: 917–920.

471 Edgar, R.C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput.
472 *Nucleic Acids Res.* **32**: 1792–1797.

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

473 Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**:
474 2460–2461.

475 Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat.*
476 *Methods* **10**: 996–8.

477 Geffert, J.L., Frahm, J.P., Barthlott, W., and Mutke, J. (2013) Global moss diversity: spatial and
478 taxonomic patterns of species richness. *J. Bryol.* **35**: 1–11.

479 Graham, E., Heidelberg, J.F., and Tully, B. (2017) Undocumented Potential For Primary Productivity In
480 A Globally-Distributed Bacterial Photoautotroph. *bioRxiv*.

481 Grasby, S.E., Richards, B.C., Sharp, C.E., Brady, A.L., Jones, G.M., Dunfield, P.F., and Williamson,
482 M.-C. (2013) The Paint Pots, Kootenay National Park, Canada — a natural acid spring analogue
483 for Mars 1, 2. *Can. J. Earth Sci.* **50**: 94–108.

484 Gundale, M.J., Deluca, T.H., and Nordin, A. (2011) Bryophytes attenuate anthropogenic nitrogen inputs
485 in boreal forests. *Glob. Chang. Biol.* **17**: 2743–2753.

486 Gundale, M.J., Nilsson, M., Bansal, S., and Jäderlund, A. (2012) The interactive effects of temperature
487 and light on biological nitrogen fixation in boreal forests. *New Phytol.* **194**: 453–463.

488 Ho, A. and Bodelier, P.L.E. (2015) Diazotrophic methanotrophs in peatlands: the missing link? *Plant*
489 *Soil* 419–423.

490 Hughes, E., Head, B., Maltman, C., Piercey-Normore, M., and Yurkov, V. (2017) Aerobic anoxygenic
491 phototrophs in gold mine tailings in. *Can. J. Microbiol.* **63**: 212–218.

492 Ininbergs, K., Bay, G., Rasmussen, U., Wardle, D.A., and Nilsson, M.C. (2011) Composition and
493 diversity of *nifH* genes of nitrogen-fixing cyanobacteria associated with boreal forest feather
494 mosses. *New Phytol.* **192**: 507–517.

495 Jean, M. (2017) Broadleaf litter and environmental effects on bryophytes in boreal forests of Alaska.

496 Jehmlich, N., Schmidt, F., Taubert, M., Seifert, J., Bastida, F., von Bergen, M., et al. (2010) Protein-
497 based stable isotope probing. *Nat. Protoc.* **5**: 1957–1966.

498 Joshi, N.A. and Fass, J.N. (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for
499 FastQ files (Version 1.33) [Software].

500 Kip, N., Winden, J.F. Van, Pan, Y., Bodrossy, L., Reichart, G., Smolders, A.J.P., et al. (2010) Global
501 prevalence of methane oxidation by symbiotic bacteria in peat-moss ecosystems. *Nat. Geosci.* **3**:
502 617–621.

503 Kornberg, H.L. (1966) The role and control of the glyoxylate cycle in *Escherichia coli*. *Biochem. J.* **99**:
504 1–11.

505 Kostka, J.E., Weston, D.J., Glass, J.B., Lilleskov, E.A., Shaw, A.J., and Turetsky, M.R. (2016) Tansley
506 insight The *Sphagnum* microbiome : new insights from an ancient plant lineage. *New Phytol.*

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

507 Laforest-Lapointe, I., Messier, C., and Kembel, S.W. (2016) Tree phyllosphere bacterial communities:
508 exploring the magnitude of intra- and inter-individual variation among host species. **4**: 1–18.

509 Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**:
510 357–359.

511 Leff, J.W., Del Tredici, P., Friedman, W.E., and Fierer, N. (2015) Spatial structuring of bacterial
512 communities within individual *Ginkgo biloba* trees. *Environ. Microbiol.* **17**: 2352–2361.

513 Leppänen, S.M., Rissanen, A.J., and Tiirola, M. (2015) Nitrogen fixation in *Sphagnum* mosses is
514 affected by moss species and water table level. *Plant Soil* **389**: 185–196.

515 Li, D., Liu, C.M., Luo, R., Sadakane, K., and Lam, T.W. (2014) MEGAHIT: An ultra-fast single-node
516 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
517 *Bioinformatics* **31**: 1674–1676.

518 Lindo, Z. and Gonzalez, A. (2010) The Bryosphere: An Integral and Influential Component of the
519 Earth's Biosphere. 612–627.

520 Lundberg, D.S., Yourstone, S., Mieczkowski, P., Jones, C.D., and Dangl, J.L. (2013) Practical
521 innovations for high-throughput amplicon sequencing. *Nat. Methods* **10**: 999–1002.

522 Markowitz, V.M., Chen, I.M.A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., et al. (2014) IMG/M 4
523 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* **42**: 568–
524 573.

525 Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
526 *EMBnet.journal* **17**: 10–12.

527 McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., et al. (2012) An
528 improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of
529 bacteria and archaea. *ISME J.* **6**: 610–618.

530 Miller, C.S., Baker, B.J., Thomas, B.C., Singer, S.W., and Banfield, J.F. (2011) EMIRGE:
531 reconstruction of full-length ribosomal genes from microbial community short read sequencing
532 data. *Genome Biol.* **12**: R44.

533 Op den Camp, H.J.M., Islam, T., Stott, M.B., Harhangi, H.R., Hynes, A., Schouten, S., et al. (2009)
534 Environmental, genomic and taxonomic perspectives on methanotrophic Verrucomicrobia.
535 *Environ. Microbiol. Rep.* **1**: 293–306.

536 Opelt, K., Chobot, V., Hadacek, F., Schönmann, S., Eberl, L., and Berg, G. (2007) Investigations of the
537 structure and function of bacterial communities associated with *Sphagnum* mosses. *Environ.*
538 *Microbiol.* **9**: 2795–2809.

539 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM:
540 assessing the quality of microbial genomes recovered from isolates, single cells, and
541 metagenomes. *Genome Res.* **25**: 1043–55.

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

542 Perovich, D.K. (2007) Light reflection and transmission by a temperate snow cover. *J. Glaciol.* **53**:
543 201–210.

544 Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 - Approximately maximum-likelihood trees
545 for large alignments. *PLoS One* **5**:.

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

547 Redford, A.J., Bowers, R.M., Knight, R., Linhart, Y., and Fierer, N. (2010) The ecology of the
548 phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves.
549 *Environ. Microbiol.* **12**: 2885–93.

550 Rousk, K., Jones, D.L., and Deluca, T.H. (2013) Moss-cyanobacteria associations as biogenic sources
551 of nitrogen in boreal forest ecosystems. **4**: 1–10.

552 Rousk, K., Rousk, J., Jones, D.L., Zackrisson, O., and DeLuca, T.H. (2013) Feather moss nitrogen
553 acquisition across natural fertility gradients in boreal forests. *Soil Biol. Biochem.* **61**: 86–95.

554 Rousk, K., Sorensen, P.L., Lett, S., and Michelsen, A. (2015) Across-habitat comparison of diazotroph
555 activity in the subarctic. *Microb. Ecol.* **69**: 778–787.

556 Ruess, R.W., Anderson, M.D., Mcfarland, J.M., Kielland, K., Olson, K., and Taylor, D.L. (2013)
557 Ecosystem-level consequences of symbiont partnerships in an N-fixing shrub from interior
558 Alaskan floodplains. *Ecol. Monogr.* **83**: 177–194.

559 Ruess, R.W., Mcfarland, J.M., Trummer, L.M., and Rohrs-Richey, J.K. (2009) Disease-mediated
560 declines in N-fixation inputs by *Alnus tenuifolia* to early-successional floodplains in interior and
561 south-central Alaska. *Ecosystems* **12**: 489–502.

562 Sharp, C.E., Smirnova, A. V., Graham, J.M., Stott, M.B., Khadka, R., Moore, T.R., et al. (2014)
563 Distribution and diversity of Verrucomicrobia methanotrophs in geothermal and acidic
564 environments. *Environ. Microbiol.* **16**: 1867–1878.

565 Trexler, R., Solomon, C., Brislawn, C.J., Wright, J.R., Rosenberger, A., Keddache, M., et al. (2014)
566 Assessing impacts of unconventional natural gas extraction on microbial communities in
567 headwater stream ecosystems in Northwestern Pennsylvania. *Front. Microbiol.* **5**: 1–13.

568 Turetsky, M.R., Bond-Lamberty, B., Euskirchen, E., Talbot, J., Frolking, S., McGuire, a. D., and
569 Tuittila, E.S. (2012) The resilience and functional role of moss in boreal and arctic ecosystems.
570 *New Phytol.* **196**: 49–67.

571 Vile, M.A., Kelman Wieder, R., Živković, T., Scott, K.D., Vitt, D.H., Hartsock, J.A., et al. (2014) N2-
572 fixation by methanotrophs sustains carbon and nitrogen accumulation in pristine peatlands.
573 *Biogeochemistry* **121**: 317–328.

574 Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naïve Bayesian classifier for rapid
575 assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**:
576 5261–5267.

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

577 Wu, Y.-W., Tang, Y.-H., Tringe, S.G., Simmons, B. a, and Singer, S.W. (2014) MaxBin: an automated
578 binning method to recover individual genomes from metagenomes using an expectation-
579 maximization algorithm. *Microbiome* **2**: 26.

580 Yurkov, V. and Hughes, E. (2017) Aerobic Anoxygenic Phototrophs: Four Decades of Mystery. In,
581 Hallenbeck,P.C. (ed), *Modern Topics in the Phototrophic Prokaryotes: Environmental and Applied*
582 *Aspects*. Springer International Publishing, Cham, pp. 193–214.Yutin, N., Suzuki, M.T., and Béjà,
583 O. (2005) Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs.
584 *Appl. Environ. Microbiol.* **71**: 8958–8962.

585 Yutin, N., Suzuki, M.T., Rosenberg, M., Rotem, D., Madigan, M.T., Süling, J., et al. (2009) BchY-based
586 degenerate primers target all types of anoxygenic photosynthetic bacteria in a single PCR. *Appl.*
587 *Environ. Microbiol.* **75**: 7556–7559.

588 Zackrisson, A.O., Deluca, T.H., Nilsson, M., Sellstedt, A., Berglund, L.M., and Zackrisson, O. (2004)
589 Nitrogen fixation increases with successional age in boreal forests. *Ecology* **85**: 3327–3334.

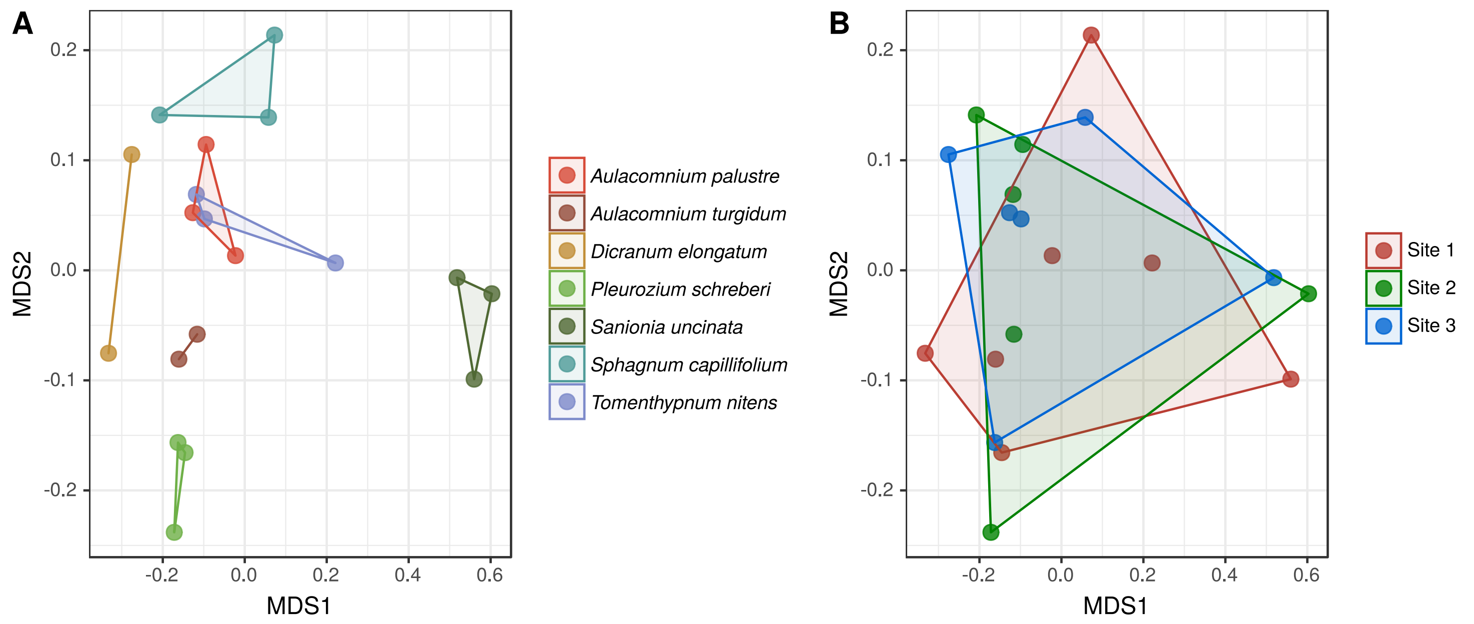
590 Zackrisson, O., Deluca, T.H., Gentili, F., Sellstedt, A., and Jäderlund, A. (2009) Nitrogen fixation in
591 mixed *Hylocomium splendens* moss communities. *Oecologia* **160**: 309–319.
592
593
594 **Author Contributions**
595 The project was conceived by N.F., M.C.M, and S.F.M. The manuscript was written by H.H.M. and
596 N.F. with contributions from all co-authors. M.C.M. collected the samples. J.S. and S.M. performed the
597 isotopic analysis. L.R.L. and S.F.M. identified the moss species and prepared the voucher specimens.
598 H.H.M. and N.F. performed the molecular analyses.
599
600 **Conflict of Interest Statement**
601 The authors declare no conflict of interest in this study.
602
603 **Materials and Correspondence**
604 Requests for materials and correspondence should be addressed to N.F. or H.H.M.
605
606 **Figure Captions:**
607
608 **Figure 1:** Heat map showing the relative abundances of the most abundant bacterial families across the
609 7 moss species. Phylum-level classifications for each family are noted on the left. Higher relative
610 abundances are indicated in brighter colors. Families with no official classification are noted as
611 "Unclassified Family." Host moss species are indicated along the x-axis and ordered according to
612 overall bacterial community similarity.
613
614 **Figure 2:** Non-metric multidimensional scaling (NMDS) plot of bacterial community dissimilarities
615 across samples (measured using the Bray-Curtis distance metric). Samples from the same host species
616 tend to cluster together (A), while those from the same site were spread across ordination space (B). A
617 PERMANOVA test revealed that species identity contributes to 63% of the variation in community
618 composition ($R^2 = 0.63$, $p < 0.001$), while site is not a significant source of variation between these
619 communities ($R^2 = 0.08$, $p = 0.21$).
620
621 **Figure 3:** A phylogenetic tree of the top 30 bacterial phylotypes (circles) and their N_2-fixing potential.
622 Putative N_2-fixers (black circles) were identified by their relatedness to known N_2–fixers (see text).

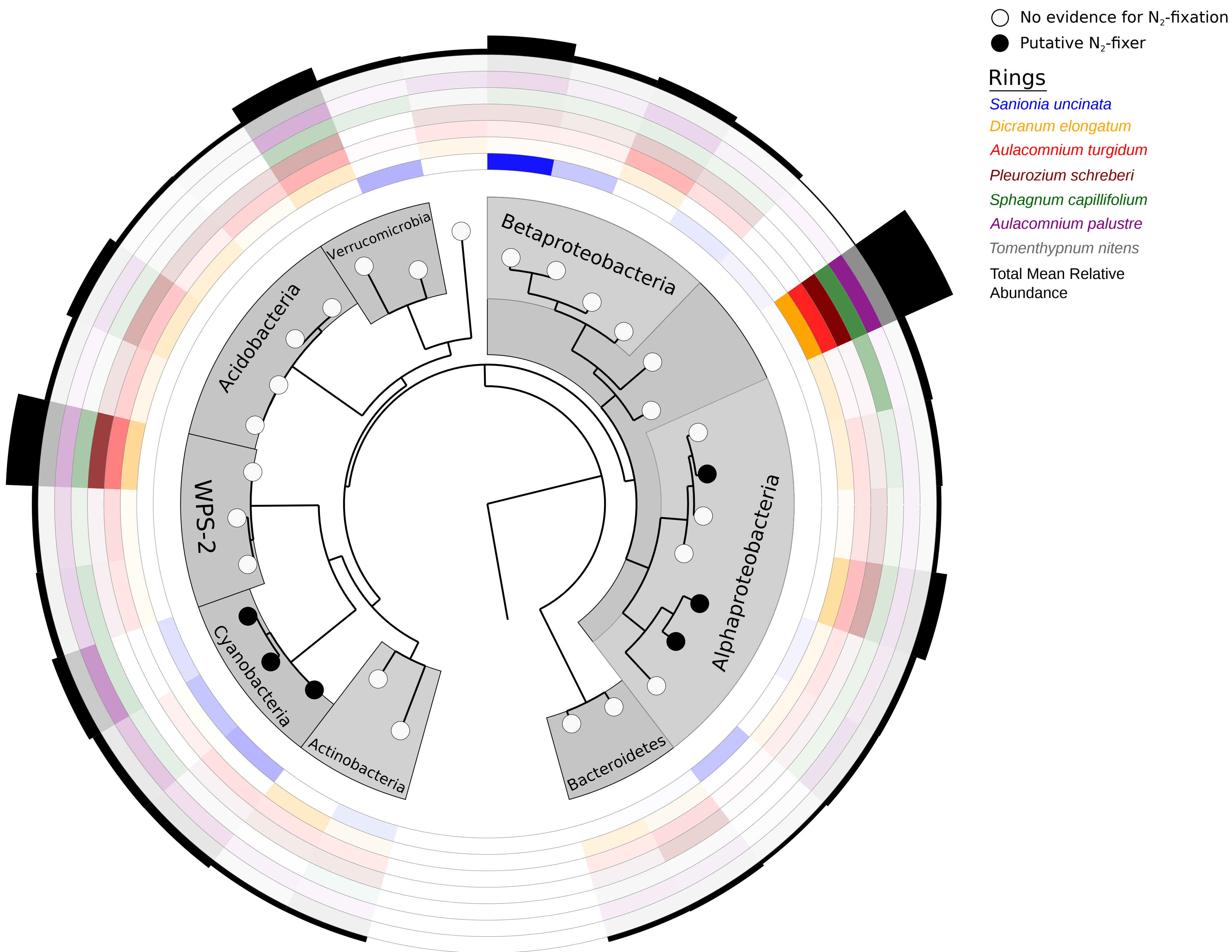
bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

623 The putative N_2-fixers observed in our moss samples include representatives from the cyanobacterial
624 and alphaproteobacterial groups. The phylotypes represented here were chosen as they were
625 consistently the most abundant phylotypes across all moss species and together account for 35% of 16S
626 rRNA gene reads in the dataset. Each inner colored ring represents a different species of moss, from the
627 inside ring to the outer ring they are: *Sanionia uncinata* (blue), *Dicranum elongatum* (yellow),
628 *Aulacomnium turgidum* (bright red), *Pleurozium schreberi* (dark red), *Sphagnum capillifolium* (green),
629 *Aulacomnium palustre* (purple), *Tomenthypnum nitens* (gray). The opacity of the colored rings
630 represents the relative abundance of the different phylotypes (log transformed and scaled between 0
631 and 1) within each moss species. The black outer ring represents the total relative abundance of each
632 phylotype across all host species (also log transformed and scaled between 0 and 1).
633
634 **Figure 4:** Diagram highlighting the inferred metabolic pathways linked to anoxygenic phototrophy and
635 carbon cycling between the WPS-2 bacterial phylotype found to be abundant in mosses and the moss
636 host cells. In the bacterial cell, light in a complementary spectrum to that absorbed by plants can be
637 used to fuel cellular processes through anoxygenic phototrophy. CO_2 fixation via RuBisCo in the
638 bacterial cell may also occur, a novel feature for an aerobic anoxygenic phototroph. In the host cell,
639 photorespiration (the action of RuBisCo on O_2 rather than CO_2) can produce a Calvin Cycle-inhibiting
640 byproduct, 2-phosphoglycolate, that the bacterial cell may have the ability to metabolize for biomass
641 production though conversion to glyoxylate, a process known as the “glyoxylate bypass”. This process
642 represents a possible mechanism for a mutalistic or commensal interactions between these WPS-2
643 bacteria and the moss host.

Phylum	Family	*Aulacomnium palustre*	*Tomenthypnum nitens*	*Sphagnum capillifolium*	*Aulacomnium turgidum*	*Pleurozium schreberi*	*Dicranum elongatum*	*Sanionia uncinata*
Proteobacteria	*Acetobacteraceae*	0.058	0.069	0.128	0.125	0.135	0.164	0.013
	Sinobacteraceae	0.093	0.079	0.076	0.094	0.098	0.135	0.014
	Comamonadaceae	0.021	0.026	0.026	0.015	0.021	0.001	0.146
	Caulobacteraceae	0.015	0.015	0.028	0.031	0.028	0.026	0.015
	Sphingomonadaceae	0.009	0.017	0.007	0.016	0.012	0.002	0.083
	Unclassified Family 1	0.031	0.018	0.019	0.012	0.004	0.017	0.003
	Methylocystaceae	0.015	0.019	0.009	0.023	0.018	0.009	0.003
	Burkholderiaceae	0.015	0.01	0.008	0.028	0.023	0.008	<0.005
	Rhodospirillaceae	0.015	0.012	0.025	0.01	0.01	0.014	0.004
	Unclassified Family 4	0.015	0.016	0.018	0.008	0.008	0.006	0.014
	Unclassified Family 3	0.011	0.012	0.01	0.012	0.018	0.005	0.008
	Xanthomonadaceae	0.027	0.015	0.008	0.01	0.003	0.005	0.007
	Coxiellaceae	0.012	0.006	0.016	0.013	0.003	0.022	0.001
	Bradyrhizobiaceae	0.008	0.013	0.007	0.011	0.004	0.005	0.008
	Unclassified Family 2	0.007	0.005	0.011	0.003	0.003	0.007	0.011
	Hyphomicrobiaceae	0.005	0.008	0.008	0.001	<0.005	0.006	0.012
Acidobacteria	*Acidobacteriaceae*	0.063	0.056	0.052	0.129	0.158	0.122	0.006
	Solibacteraceae	0.018	0.014	0.021	0.02	0.016	0.03	0.002
Verrucomicrobia	*[Chthoniobacteraceae]*	0.034	0.029	0.037	0.051	0.051	0.015	0.049
	Unclassified Family 1	0.042	0.037	0.03	0.041	0.034	0.02	0.001
	auto67_4W	0.032	0.022	0.02	0.003	<0.005	0.009	0.002
	Opitutaceae	0.009	0.011	0.014	0.004	0.002	0.011	0.008
Bacteroidetes	*Chitinophagaceae*	0.035	0.054	0.036	0.022	0.017	0.015	0.087
	Sphingobacteriaceae	0.027	0.029	0.015	0.033	0.058	0.019	0.025
	Cytophagaceae	0.005	0.011	0.011	0.001	<0.005	0.003	0.046
Cyanobacteria	*Nostocaceae*	0.091	0.069	0.046	0.028	0.039	0.001	0.082
	Unclassified Family 1	0.015	0.009	0.007	0.011	0.001	0.013	0.001
WPS–2	*Unclassified Family 1*	0.064	0.065	0.06	0.077	0.075	0.062	0.001
Planctomycetes	*Unclassified Family 1*	0.018	0.017	0.022	0.018	0.019	0.025	0.01
	Isosphaeraceae	0.013	0.016	0.012	0.022	0.03	0.023	0.002
	Gemmataceae	0.015	0.016	0.014	0.006	0.003	0.025	0.01
Actinobacteria	*Conexibacteraceae*	0.004	0.005	0.001	0.012	0.011	0.03	<0.005
	Microbacteriaceae	0.004	0.013	0.004	0.009	0.012	0.003	0.007

bioRxiv preprint doi: https://doi.org/10.1101/219659; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.




No evidence for N_2-fixation
Putative N_2-fixer
Rings
Sanionia uncinata
Dicranum elongatum
Aulacomnium turgidum
Pleurozium schreberi
Sphagnum capillifolium
Aulacomnium palustre
Tomenthypnum nitens
Total Mean Relative Abundance
Betaproteobacteria
Alphaproteobacteria
Bacteroidetes
Actinobacteria
Cyanobacteria
WPS-2
Acidobacteria
Verrucomicrobia

Moss Cell

Chloroplast

Calvin Cycle

Ribulose-1,5-bisphosphate

3-Phosphoglycerate

RuBisCo

O_2

2-Phosphoglycolate

Anoxygenic Phototrophy

WPS-2

Calvin Cycle

Ribulose-1,5-bisphosphate

RuBisCo

CO_2

Acetyl-CoA

Oxaloacetate

Malate

Citrate

Glyoxylate

Succinate

Isocitrate

Biomass production

TCA Cycle with Glyoxylate Bypass