Boreal moss-microbe interactions are revealed through metagenome assembly of novel bacterial species

Sarah Ishak[13*+], Jonathan Rondeau-Leclaire[1+], Maria Faticov[123], Sébastien Roy[12], and Isabelle Laforest-Lapointe[123*]

[1]Département de biologie, Université de Sherbrooke, Sherbrooke, Québec, Canada

[2]Centre SÈVE, Université de Sherbrooke, Sherbrooke, Québec, Canada

[3]Centre d'Étude de la Forêt, Université du Québec à Montréal, Montréal, Québec, Canada

[+]These authors contributed equally

*Co-corresponding authors: isabelle.laforest.lapointe@gmail.com ; s.isaac1905@protonmail.com

ORCID NUMBERS

Sarah Ishak	(0000-0003-3412-3642)
Jonathan Rondeau-Leclaire	(0000-0002-4433-3566)
Maria Faticov	(0000-0001-8206-9332)
Sébastien Roy	(0000-0001-7183-0691)
Isabelle Laforest-Lapointe	(0000-0003-3630-9475)

Abstract (max 200 words)

Moss-microbe interactions play important roles for ecosystem processes in boreal forests. Yet, the extent to which the environment and host-specific characteristics drive the composition and metabolic potential of boreal moss microbiomes is still poorly understood. In this study, we use shotgun metagenomics to identify the taxonomy and metabolic potential of the bacterial communities of four moss species in the boreal forests of Northern Québec, Canada. To better characterize moss bacterial community composition and diversity, we assembled the genomes of 110 potentially novel bacterial species. Our results highlight that moss genus, species, gametophyte section, and key abiotic factors drive moss-associated bacterial community composition and diversity. Green gametophyte sections harbored differentially abundant bacteria (e.g., *Bradyrhizobia*) exhibiting significant carbon fixation capacities, particularly through the phosphate acetyltransferase-acetate kinase pathway. On the other hand, brown gametophyte sections harbored differentially abundant Cyanobacteria (e.g., *Stigonema* spp.) demonstrating partial pathways for anoxygenic photosynthesis and nitrogen fixation, which are important for nitrogen cycling in boreal ecosystems. Overall, our findings provide insights into the relationships between boreal mosses and their bacterial communities, for which data remains scarce. As global change accelerates in boreal ecosystems, identifying the roles of moss microbial symbionts in influencing their hosts and ecosystem functions becomes critical.

Key words (6): moss, bacteria, metabolic potential, abiotic conditions, boreal forests, metagenome.

Introduction

Boreal forests are one of Earth's largest biomes, covering 29% of land surfaces [1] and storing at least 30% of global carbon [2]. Within this ecosystem, a diverse understory vegetation thrives, consisting of vascular plants, lichens, and bryophytes (small non-vascular plants including liverworts, hornworts, and mosses). Ubiquitous in boreal forests, mosses are crucial contributors to key ecosystem processes, notably to nitrogen (N) fixation and carbon (C) sequestration [3], which are also facilitated by moss symbiotic microorganisms [4, 5]. Moss microbes contribute around 50% of N inputs to boreal forests, play crucial roles in facilitating C degradation in peatlands, and facilitate host adaptation in degraded landscapes [6]. Moss-associated microbial communities are notably dominated by bacteria from the Proteobacteria, Acidobacteria, Actinobacteria, and Cyanobacteria phyla [7–9]. Driven by the pressures of accelerating climate change, boreal ecosystems face a temperature increase of ~2.5ºC before the end of the century, which could alter moss-microbe interactions with major implications for global N [10] and C cycles [11]. In addition, boreal forests are often harvested intensively for forestry or disturbed by mining activities [12, 13]. Considering the strong pressures of climate change on boreal forests combined with the ever-increasing anthropogenic impacts on land use, improving our understanding of the drivers and functions of moss-microbe interactions is imperative to mitigate the risks of altering this precious habitat [14, 15].

Moss microbial communities can be affected by many local abiotic and biotic factors [8, 16]. For example, soil pH and temperature were shown to structure bacterial communities associated with *Sphagnum* moss species [9, 17, 18]. Phylogenetic relationships and host species are also important drivers of moss bacterial community diversity and community composition [5, 9, 20]. While different moss species have been shown to associate with contrasting bacterial communities [18, 21, 22], closely related moss species are likely to

harbor more similar microbial taxa than more distantly related ones. This convergence can be attributed to shared host physiological traits and ecological niches. For example, Holland-Moritz *et al*. [23], showed that phylogenetic relationship between *Sphagnum* species explained up to 65% of the variation in bacterial community composition. Interestingly, moss identity also influences the effects of environmental factors on bacterial diversity and community composition [18]. While previous studies have often focused on characterizing the microbial taxonomy and functions of dominant boreal mosses such as *Pleurozium schreberi* [24] and peat mosses of the genus *Sphagnum* [25], scant research is available on other moss genera such as *Polytrichum* and *Dicranum* (but see [26]), which are common in Canadian boreal forests [27]. Importantly, these taxa are key candidates for remediation and restoration efforts as they often thrive in degraded landscapes [28]. There is thus an important knowledge gap on moss-microbe interactions beyond common moss genera. Considering the urgency to develop conservation strategies for bryophytes, it appears crucial to deepen our understanding of moss-microbe interactions and of their functional roles in unaltered boreal ecosystems by investigating lesser studied moss species, their phylogenetic relationships, and local environmental factors.

Beyond moss identity, physiological differences within single host, such as disparities in tissue types along the moss gametophyte, could also influence moss-microbe interactions. Gametophytes are the dominant life stage of mosses, which are separated along their vertical gradient in two distinct sections: an above-ground living “green” section, and an underlying, decaying “brown” section [29]. These sections can differ in microbial diversity, community composition, and functional profiles [30, 31]. The green section harbors photosynthetically active microbes [32, 33] and is associated with N-fixing cyanobacteria [34]; it is often the focus of studies on moss-bacteria associations. Meanwhile, the brown section is much less studied, even though it is thought to be an important source of nutrients for specific groups of

microbes, such as bacteria [35]. Notably, green and brown sections of the moss gametophyte are exposed to different microenvironmental conditions [36], where variations in light, soil pH, moisture, and temperature can impact moss microbial diversity and functions. Overall, to gain a detailed overview of moss-microbe interactions, it is necessary to measure how gametophyte sections assemble divergent microbial communities and functional profiles, and whether these differences are consistent among moss species as well as along environmental gradients.

In recent years, metagenome-resolved community profiling became a widespread approach to estimate environmental microbial composition and functions [37]. The advent of whole metagenome sequencing (WMS) revolutionized the field of microbial ecology, making it possible to assemble metagenome-assembled genomes (MAGs) and thus delve into the so-called "dark matter" of microbiomes [38]. By leveraging the entire genomic material of a sample, WMS not only enables the discovery of novel or uncultured microbial species; it also provides a more complete picture of the metabolic potential of community members. WMS is therefore particularly useful in characterizing understudied microbiomes, such as that of boreal mosses. The discovery of novel microbial species combined with an improved understanding of how the metabolic potential of microbes differs across gametophytes and along environmental gradients, could be key to predict changes in moss-microbe interactions as climate change and human activities continue to impact boreal ecosystems.

Here, we measure the influence of moss identity, gametophyte section, and abiotic environment on moss bacterial communities and metabolic potential. We focus on four boreal moss species in the Eeyou Istchee (James Bay) region of Québec, Canada: *Polytrichum juniperinum*, *P. commune*, *P. piliferum*, and *Dicranum undulatum* (Fig. 1ab). We use WMS and a combination of cutting-edge bioinformatic tools to (i) characterize moss-microbe interactions, (ii) assemble metagenomes, and (iii) identify novel bacterial species associated

with boreal mosses. Our objectives are to (1) provide a detailed portrait of moss-associated bacteria and their metabolic potential; and (2) quantify the relative importance of moss species, gametophyte section, soil pH, and temperature as drivers of moss-bacteria interactions.

Results and Discussion

Bacterial community composition and diversity are driven by moss identity, gametophyte section, soil pH, and temperature

First, we investigated the influence of moss genus, moss species (nested within moss genus), gametophyte section, soil pH, and temperature on bacterial community composition and diversity (Table 1a; Fig. 2). We then narrowed our focus on the three *Polytrichum* spp. (excluding *Dicranum* sp.; Table 1b). This sequential approach allowed us to disentangle whether closely related moss species are likely to harbor more similar microbial taxa than more distantly related ones. Both moss genus and species (nested within genera) were strong drivers of bacterial community composition (Genera: R^2 = 13.8%, p-value = 0.001; Species: R^2 = 12.3%, p-value = 0.001; Fig. 2a, Table 1a). Thus, our results highlight that differences in bacterial community composition are driven by both moss genus and, within these genera, by moss species. These findings are in line with previous studies in which moss phylogeny has been shown to be an important driver of bacterial community composition [4, 8, 25]. Notably, when restricting the analysis to the three closely related *Polytrichum* spp., the effect of moss species on bacterial community composition remained strong and significant (R^2 = 19.7%, p-value = 0.001; Table 1b). This finding illustrates a potential co-evolutionary relationship between mosses and some of the bacteria members within the community [17, 39]. Yet, more studies are needed to fully understand the nature and mechanisms driving these genus- and species-specific interactions. We also found that bacterial community composition differed between green and brown gametophyte sections, regardless of whether *Dicranum* sp. was included in the model or not (R^2 = 12.7%, p-value = 0.001; R^2 = 17.2%, p-value = 0.001, respectively; Fig. 2b, Table 1ab). This variation between sections can be explained by differences in physiology of the gametophyte sections [40] and micro-environmental

conditions (e.g., soil pH, moisture, temperature) [36]. In addition, we observed a significant interaction between gametophyte sections and moss genus (R^2 = 2.6%, p-value = 0.001; Table 1a), as well as moss species (R^2 = 1.3%, p-value = 0.042; Table 1b), highlighting that the effect of gametophyte section on bacterial community composition differs among moss species. In both models, whether *Dicranum* sp. was included or not, we found that soil temperature (with *Dicranum:* R^2 = 7.4%, p-value = 0.001; without *Dicranum*: R^2 = 6.9, p-value = 0.001) and pH (R^2 = 5.1%, p-value = 0.001 and R^2 = 4.1%, p-value = 0.001 respectively) were significant drivers of bacterial community composition. Interestingly, the effect of abiotic factors on community composition was weaker than that of moss genus and species (Table 1). Finally, several 2nd level interactions of soil pH and temperature impacted moss bacterial community composition (Table 1), in coherence with the important roles of abiotic factors in structuring moss-microbe interactions across moss species [9, 17, 18].

Bacterial alpha diversity differed between moss genera, with alpha diversity being higher for *Polytrichum* than for *Dicranum* (χ^2 = 10.3, p-value = 0.001; Table 2a, Fig. 2c). However, we did not detect a significant association between moss species and bacterial alpha diversity, neither when nesting species within moss genera (χ^2 = 0.07, p-value = 0.97; Table 2a, Fig. 2c) nor when narrowing the model to *Polytrichum* spp. (χ^2 = 0.08, p-value = 0.96; Table 2b, Fig. 2c). Of note, alpha diversity in green section was consistently lower than in brown section of the gametophyte. This pattern was consistent between moss genera (χ^2 = 29.2, p-value < 0.001; Fig. 2c, Table 2a) and among *Polytrichum* spp. (χ^2 = 35.7, p-value < 0.001; Fig. 2c, Table 2b). This could be explained by (i) a phenomenon of niche dominance of a few taxa in green section or by (ii) the proximity of brown section to soil which is known to be extremely rich in microorganisms. As such, brown section could behave as an interface in which both soil and moss-associated bacteria can thrive. This section is also mostly shaded from UV by the overlying green section due to the boundary layer effect (i.e., the physical

phenomenon in which wind velocity is reduced); an effect that minimizes moisture loss [41]. Among abiotic factors, soil temperature was the only factor that showed a significant association with moss bacterial alpha diversity, evident in both models, with and without *Dicranum* sp. ($\chi^2 = 6.3$, p-value = 0.01 and $\chi^2 = 6.6$, p-value = 0.01; Fig. 2c, Table 2ab). Furthermore, we did not detect a significant effect of the interaction between gametophyte section and moss genus or species on bacterial alpha diversity (Table 2).

In summary, our results show that moss genus, gametophyte section, soil pH and temperature, are important drivers of moss bacterial community composition and diversity, but that moss species only affects the former. As boreal ecosystems face the heightened pressures of global change, our work demonstrates that moss green and brown sections within a moss gametophyte differ in bacterial community composition (Fig. 2d) and alpha diversity, which subsequently could affect gametophyte-specific microbial functions.

Metagenome assembly reveals that novel bacteria constitute most of moss bacterial community

By assembling our metagenomes, we discovered 110 potentially novel bacterial species whose genome was not represented in reference databases. Specifically, we performed metagenome co-assembly, binning, and dereplication which produced 157 non-redundant species-level MAGs. We filtered out MAGs with quality scores (QS) ≤ 0.50 (QS = completeness - 5 × contamination), yielding a subset of 118 MAGs of high quality. Then, in order to estimate which MAGs belong to undescribed species, we compared them to every known representative bacterial genome using a combination of tools based on mash distance and average nucleotide identity (see *Methods*). This resulted in a set of draft genomes from 110 putative novel species which belonged to 23 bacterial orders from nine phyla (Fig. 3). Hereafter, we refer to this set as the novel species-level MAGs (nMAGs), in

that they constitute genomes that are non-redundant with any published genomes at the species level. These novel bacterial genomes have a mean length of 4.0 ± 1.2 Mbp, QS of 0.73 ± 0.13, and L50 of 11.9 kb ± 10.5. Full quality statistics of the nMAGs are reported in Supplementary Data 1, including the number of contigs and proportion of GC content.

Among these potential novel species interacting with mosses, we recovered seven nMAGs from the order Baltobacterales [42] from the phylum Eremiobacterota [43] (Fig. 3) despite the presence of 200 species-representative genomes in the GTDB Release 08-RS214 database (28th April 2023). Many Eremiobacterota species were recently described by metagenome assembly from a number of boreal ecosystems such as Alaskan mosses [18], Canadian acidic springs [44], and Swedish peatlands [45]. While there is little to no information on the ecological traits of the species belonging to Eremiobacterota, several genera were shown to have the capacity for C fixation [39]. Species from the Baltobacterales order were highly prevalent in our samples, detected in 109/131 (~83%) samples, with two *Nyctobacter* MAGs being of very high quality (QS = 0.95 and 0.91 respectively; Supplementary Data 1). Our study thus contributes novel information about moss-associated members of this clade.

To gain a more detailed understanding of the microbial communities of boreal mosses, we added the 110 nMAGs to the 85,205-reference genome GTDB used for community profiling, resulting in an average 4.3-fold increase in sample containment (a measure of proportion of sample k-mers included in the minimal reference genomes set that encompasses all k-mers in a sample; see Methods; Supplementary Figure S1). This demonstrates the effectiveness of using *de novo* metagenome assembly to identify and describe previously unknown microbial species, which is crucial for environmental datasets, and to increase the proportion of sample information to perform ecological analyses.

Specific bacterial species associate with moss species

To test for differentially abundant taxa, we used ANCOM-BC2 [47], which estimates $\log_2$-fold changes in absolute abundances between sample groups through pairwise comparisons. To assess differences in abundance across moss species, we used the Dunnett's test, with *D. undulatum* samples as a reference group for comparison. *D. undulatum* was selected because of its contrasting bacterial community composition, which served as a baseline for evaluating differential abundant bacteria among the moss species (Fig. 4a). A total of 28 species were found to be differentially abundant across the three *Polytrichum* species compared to *D. undulatum* (Fig. 4a; Supplementary Figure S2). Of these 28 bacterial species, 24 were nMAGs, underlining the contribution of the assembly process to the depth of our analyses. Notably, we found that the order Baltobacterales was more abundant in *P. commune* by 2.7 to 3.3 $\log_2$-folds in comparison to other moss species (Fig. 3; Supplementary Figure S2). Previous metagenomic studies have shown that Baltobacterales encode genes necessary for hopanoid synthesis (a biochemical process in which hopanoid, steroid-like compounds, are produced) [46]. Hopanoid molecules have been shown to contribute to bacterial tolerance to low pH soils [48], temperature stress [49], microbial defense [50], and are linked to N-fixation [51, 52]. Interestingly, hopanoid structures were shown to act as biomarkers for the presence of specific microorganisms, including Cyanobacteria [53]. Notably, *P. commune* is known for its antibacterial properties [54] and specimens from this moss species were sampled from acidic soils (Supplementary Data S2) in our study. The stress-tolerance properties of hopanoid structures in Baltobacterales MAGs could provide an adaptive advantage for colonization and persistence of bacteria in acidic and anti-microbial habitats such as *P. commune* mosses.

Interestingly, two *Stigonema* spp. (cyanobacteria) were differentially abundant in *P. juniperinum* and *P. piliferum*, but not in *P. commune*. During sampling, the majority of

P. commune were found in shaded conditions (Supplementary Data S2). This suggests that shade, under which *P. commune* grows, may hinder the growth of Cyanobacteria. In general, many nMAGs that were differentially abundant in *Polytrichum* species belonged to the orders Acetobacterales and Terriglobales (Fig. 4a; Supplementary Figure S2). Of note, Terriglobales, one of the most overall relatively abundant taxa (20.8 ± 10.8% of attributed sequences across all samples, Fig. 2d), were also found to be highly present in sub-arctic and boreal peatland moss microbiomes [55].

Three species were found to be differentially abundant in *D. undulatum*, from the orders Solirubrobacterales, Steroidobacterales, and Streptosporangiales, respectively (Fig. 4a; Supplementary Figure S2). The most closely related species from these orders were incidentally all species identified through MAG assembly and were described in soil and roots in non-boreal ecosystems [56, 57]. A potential reason for the differences in bacterial community between *Dicranum* and *Polytrichum* genera could be due to environmental preferences (*Dicranum* sp. preferring moist soils), as well as moss N and phosphorus (P) content [58]. Although our results show that moss-associated microbial communities are influenced by moss identity, experimental studies are needed to disentangle the simultaneous effects of biotic and abiotic conditions on moss-associated bacteria and on their contribution to ecosystem functions.

Bacterial associations are specific to brown and green gametophyte sections

We identified 44 species as differentially abundant between green and brown gametophyte sections ($p < 0.01$), of which 18 are nMAGs (Fig. 4b). These species are hereafter described as "associated" to the section in which they were more abundant. Our work identified 27 species as brown-associated, including all nine *Bradyrhizobia*, thus suggesting a strong specificity for brown gametophytes for members of this genus (Fig. 4b). Of note, Holland-

Moritz, *et al*. [18] showed that *Bradyrhizobia* are putative N_2-fixers, thus potentially contributing to N cycling.

Bacterial species with the highest log_2-fold change in brown gametophytes included three nMAGs from the genus *Edaphobacter*; species from the order Solirubrobacterales (*VAYN01*) and Terriglobales (*CAIQPK01*), as well as two species from the order Streptosporangiales (Bog-532 sp019240215 and a *Trebonia* nMAG; Fig. 4b). Various *Edaphobacter* species, such as *E. lichenicola sp. nov.*, *Edaphobacter modestus gen. nov.*, and *Edaphobacter aggregans sp. nov.*, were previously described from alpine forest soils [59, 60]. Although Solirubrobacterales are understudied, several research have detected members of this order in diverse ecosystems including agricultural lands [61] and Antarctic soils [62]. Several other brown-associated taxa, such as *Rudaea* spp., *CAIMXF01*, *Trebonia*, and *Acidisarcina*, are also mainly associated with soils [63–65]. Finally, two cyanobacterial species (*Stigonema*) were differentially abundant in brown gametophytes (Fig. 4b). In line with our findings, Renaudin, *et al*. [10] found that cyanobacterial biomass increased towards the bottom of moss gametophyte, peaking at approximately 2-3 cm from the apex. In addition, Klarenberg *et al*. [66] suggested that there is no lateral transmission from soil to moss gametophytes. However, given the number of soil taxa that were also found in brown gametophyte sections, our results suggest that this section is an interface colonised by both soil and moss-associated microbes.

In the green gametophyte section, nMAGs from the genera *CAHJXF01*, *CAHJXG01*, *Lichenicola*, and *Palsa-465* had the highest log-fold change (log_2-fold > 1.29; Fig. 4b) in absolute abundance, compared to the brown section. Hay *et al*. [68] described MAGs of *CAHJXF01* and *CAHJXG01* from high Arctic glaciers, containing the gene form I RubisCo – an important gene for C fixation. We also found that green-associated Rhizobiales encompassed the genera *Methylobacteria*, *Lichenihabitans minor*, and *RH-AL1* spp., but not

Bradyrhizobium spp. (Fig 4). *Methylobacteria* consume methanol that has been released from the stomata of non-bryophyte leaves [69] and are associated with the phyllosphere of terrestrial vascular plants [70]. The presence of stomata on bryophytes is phylogenetically determined. However, detection of this taxa may indicate the capacity for some of the studied mosses to release methanol [71]. *Methylobacteria* were also found in the core microbiome of boreal shrubs [72] and were shown to colonize non-boreal mosses (e.g., *Didymodon tectorum* and *Funaria hygrometrica* [73, 74]). Similarly, *Lichenihabitans minor* and *RH-AL1* have been shown to harbor genes associated with methanol consumption [67]. Although *Lichenihabitans minor* tends to be associated with lichens, our detection of this species is not surprising, given the prevalence of lichens in the habitat of boreal mosses [75]. *Lichenicola* and *Lichenicoccus*, eponymously, are taxa that are associated with lichens, for example *Lichenicola cladoniae* has previously been found in Antarctic lichens [76]. Overall, through our analysis, we observed enrichments in several genera such as *CAHJXF01*, *CAHJXG01*, *Methylobacteria*, *Lichenihabitans*, and *RH-AL1* spp. in green gametophyte section. Future studies would be warranted to quantify their impact on C fixation and methanol consumption in boreal forests.

Bacterial metabolic potential differs in green and brown gametophyte sections

Next, we investigated the metabolic potential of the differentially abundant bacterial species by analyzing which pathways their genomes have the potential to express. We focused on four pathway groups that are relevant for boreal ecosystem functioning: photosynthesis, nitrogen metabolism, methane metabolism, and carbon fixation.

First, we found that two brown-associated *Stigonema* species, also known as true-branched filamentous cyanobacteria, had partial pathway completeness for anoxygenic photosystems I and II (*Stigonema sp903903925* = 83.3% and *Stigonema* BrZ1 MAG = 66.7%) (Fig. 5; Supplementary Data S3). These were the only two differentially abundant

species capable of encoding these pathways at least partially. We also observed that two *Stigonema* spp. stood out because of their partially encoded N fixation (nifHDK + vnfDKGH) pathway (*Stigonema sp903903925* = 75% and *Stigonema* MAGs = 50%) (Fig. 5; Supplementary Data S3). This supports a previous finding that in feathermoss, *Stigonema* was shown to be an important contributor to N fixation [77, 78]. *Stigonema* spp. also show partial pathway completeness in all methane modules that are relevant to methanogenesis, and not methane oxidation, which is a well-described function of cyanobacteria [79] (Fig. 5; Supplementary Data S3). Overall, our findings suggest that cyanobacteria associated with the brown moss sections are photosynthetically active and have the potential for N fixation [80, 81].

Second, we detected complete assimilatory nitrate reduction pathways in both green- and brown-associated Rhizobiales, which have no capacity for N fixation (Fig. 5; Supplementary Data S3). Although it is known that Rhizobiales are symbiotic N-fixers for nodule-producing leguminous plants, Garrido-Oter *et al*. [82] proposed that the development of symbiotic genes, including N fixing genes, can fluctuate in non-nodule producing plants. Considering that mosses do not form nodules and our observation that moss-associated cyanobacteria can fix N, we believe that the N-fixing function of Rhizobiales associated with mosses may have become redundant and therefore lost. Furthermore, most methane metabolic pathways were also partially complete in green and brown-associated Rhizobiales genomes (Fig. 5; Supplementary Data S3). This aligns with the results of a previous study on *Sphagnum* mosses, which found that methanotrophy is primarily carried out by Rhizobiales species [83]. *Methylobacteria*, which were differentially abundant in green sections, showed full potential to encode the anoxygenic photosystem II pathway, except for the *Methylobacterium* MAG, which only encodes 50% of the required genes (Fig. 5; Supplementary Data S3). Notably, certain species of phyllosphere *Methylobacteria* were

described as anoxygenic photosynthetic bacteria [84]. Of note, the only pathway from the photosynthesis group that was significantly associated with the green gametophyte section was the anoxygenic photosystem II (Supplementary Table S1). Furthermore, in green-associated Rhizobiales, we did not detect the phosphate acetyltransferase-acetate kinase pathway, which contributes to carbon fixation, unlike in most of *Bradyrhizobia* (Fig. 5; Supplementary Data S3). However, among pathways in C fixation group, phosphate acetyltransferase-acetate kinase pathway was the only one significantly associated with the green gametophyte section (Fig. 5; Supplementary Data S3). This pathway is part of a central metabolism of bacteria [85]. Yet, the completeness of both the phosphate acetyltransferase-acetate kinase pathway, coupled with the Wood-Ljungdahl pathway in *Bradyrhizobia* implies the metabolic potential for C fixation [86]. To sum up, our findings suggest that green- and brown-associated *Rhizobiales* provide a range of metabolic contributions such as C fixation, methane metabolism, and nitrate reduction to their ecosystem.

Third, brown-associated Burkholderiales (*Caballeronia* spp.) showed potential activity for denitrification, dissimilatory and assimilatory nitrate reduction, and a complete nitrification pathway (Fig. 5; Supplementary Data S3). However, the only green-associated Burkholderiales species (CAHJXF01 nMAG) lacked all three nitrification pathways (denitrification, nitrification, and complete nitrification, Fig. 5). While methane oxidation pathways were absent, partial completeness was observed in several methane metabolism pathways (Fig. 5; Supplementary Data S3). These findings suggest the diverse metabolic capability within the Burkholderiales order and their potential roles in N cycling and methane metabolism [87, 88]. Curiously, pathways for coenzyme M and F420 biosynthesis were partially to fully complete in several brown-associated species (Fig. 5; Supplementary Data S3). F420 is a co-enzyme that is costly to synthesize; nevertheless, Ney *et al*. [89] found that it was commonly synthesized by aerobic bacteria. Meanwhile, coenzyme M is found in

methanotrophic archaea and used by bacteria in metabolizing of different substrates, for example alkenes [90], which are found in boreal forest soils [91]. More interestingly, alkenes are known to be produced by bacteria associated with lichens [92]. In our sampling sites, lichens frequently grew alongside mosses in the same patches. Based on this observation, our results suggest that the presence of coenzyme M moss-associated bacteria may be attributed to their utilization of alkenes commonly found in lichens.

Finally, prokaryotic C fixation pathways such as the Arnon-Buchanon cycle and the Wood-Ljungdahl pathway, and methane metabolism pathways such as formaldehyde assimilation, had the highest average pathway completeness (Fig. 5; Supplementary Data S3). Additionally, several methane metabolism pathways, such as ribulose monophosphate pathway, methanogenesis, acetate and coenzyme M biosynthesis, were significantly associated with the green gametophyte section (Supplementary Table S1). These results therefore highlight the importance of moss-associated bacteria in biogeochemical processes outside of N fixation and recapitulate the capacity for moss-bacteria to mediate C fixation [93, 94]. Overall, our findings demonstrate the diverse metabolic potential and adaptations of bacterial species associated with different sections of moss gametophyte, shedding light on their roles in N cycling, methane metabolism, and C fixation in boreal forests.

Conclusion

In this study, using a whole metagenome sequencing approach, we (1) characterized moss-associated bacteria and their metabolic potential and (2) explored the relative importance of moss identity, moss physiological characteristics, and abiotic environment as drivers of moss-bacteria interactions. Our results showed that moss genus, moss species, gametophyte section, and to a lesser extent soil pH and soil temperature, are important drivers of bacterial community composition and diversity (Tables 1-2; Fig. 2). Using metagenomic

assembly, we identified 110 potentially novel bacterial species (Fig. 3). Our study identified distinct bacterial communities associated with four different moss species, as well as contrasting associations with green and brown gametophyte sections, revealing insights into their ecological roles in boreal forests (Fig. 4). We demonstrated how green- and brown-associated bacteria of the moss gametophyte differ in taxonomy (Tables 1-2) but also in metabolic potential (Fig. 5). Green-associated bacteria were shown to exhibit significant C fixation capacities, particularly through the phosphate acetyltransferase-acetate kinase pathway, while brown-associated cyanobacteria (e.g., *Stigonema* spp.) demonstrated partial pathways for anoxygenic photosynthesis and N fixation, important for N cycling in boreal ecosystems. Meanwhile, Burkholderiales displayed diverse metabolic functions, including potential roles in nitrification and denitrification, crucial for maintaining N balance. Overall, our findings contribute to the understanding of moss-bacteria interactions and provide insights into how mosses may respond to environmental disturbances. As bryophytes are submitted to the accelerating global change in boreal ecosystems, additional research on the rich diversity of moss species is needed to grasp fully the roles of microbial symbionts in influencing host and ecosystem functions. Invaluable data rests in ecosystems that are currently exposed to low levels of anthropogenic stress. We suggest that collecting this data is urgent, as well as studying how bryophytes and their microbial partners will adapt under altered environmental conditions.

Methods

Study area

This study was conducted within the Eeyou Istchee (James Bay) region of Northern Québec (Fig. 1a; N51.7040721°, W76.0521982°). All sampling locations were part of Québec's boreal forest biome which is dominated by Jack pine (*Pinus banksiana*) and black spruce (*Picea mariana*) [95]. The monthly average annual temperature of this region is 0.3□, while

mean annual precipitation is 741 mm [96]. The soils are predominantly acidic and nutrient poor [97].

Study design and moss species

Samples were taken from four acrocarpous (stems harbor archegonia at the top) moss species: *Polytrichum juniperinum*, *Polytrichum commune*, *Polytrichum piliferum*, and *Dicranum undulatum* (Fig. 1b), which are endemic to the Eeyou-Istchee region of Québec [28]. These were sampled in September 2021, at five different locations (Fig. 1a). Three microsites, each 5 to 10 meters apart, were randomly selected and then, at each microsite, 3-by-3-inch moss patches were established (Fig. 1a). At each microsite/species, we measured soil temperature, by inserting the probe of the Fisherbrand™ Envirometer™ directly in the soil of where the moss was sampled after it had been removed. Soil pH was measured in the laboratory using the method described by Hendershot, Lalande and Duquette [98]. Mosses were first visually identified to species level in the field and then confirmed by microscopy. For each moss species, samples were taken from both the brown and green sections of the gametophyte, which were determined visually (Fig. 1c). To maintain the integrity of the samples, the green and brown sections were separated while on site, using ethanol and DNA-away treated tweezers and scissors. Samples were placed in sterile 50 mL tubes and immediately stored at -20□ for the return trip to the laboratory, where they were stored at -80□. A total of 131 samples were collected for brown and green gametophyte sections across *P. juniperinum* ($n_B = 24$, $n_G = 22$), *P. commune* ($n_B = 9$, $n_G = 9$), *P. piliferum* ($n_B = 26$, $n_G = 26$), and *D. undulatum* ($n_B = 8$, $n_G = 7$) from five sites (see Supplementary Table S2 for sample breakdown).

DNA extraction and whole metagenome sequencing (WMS)

Green and brown gametophyte sections were processed separately, but in the same manner. DNA extraction of moss samples was performed as modified versions of the protocols used by Holland-Moritz, et al. [4] and Klarenberg, et al. [16]. Using tweezers and scissors sterilized with 80% ethanol solution and 'DNA Away' solution (Thermo Fisher Scientific, MA, USA), 0.25 g of each moss sample were weighed out and divided into two sterile 2 mL bead beating tubes that also contained three sterilized stainless-steel beads. These tubes were then frozen in liquid N_2 for approximately 30s and mosses were broken down using the FisherbrandTM Bead Mill 24 Homogenizer for 30s at 1.75 m/s. DNA was subsequently extracted using the QIAGEN DNEasy PowerSoil DNA Extraction Kit with some minor changes. Each moss sample was divided into two 2 mL PowerBead extraction tubes. The kit buffer and beads were added to each tube, along with 60 uL of C1 Solution from the kit. All tubes with buffer were homogenized at 3.25 m/s for 30s and then at 2.40 m/s for 10 minutes using the FisherbrandTM Bead Mill 24 Homogenizer. After centrifugation as per the extraction kit protocol, supernatant from both tubes of one sample were combined in one tube with 250 uL of C2 Solution from the extraction kit. The rest of the extraction followed the manufacturer's instructions. All extractions were performed under a Type A2 Bio Safety Cabinet. Extracted DNA were subsequently quantified using the NanoDropTM One Spectrophotometer (Thermo Fisher Scientific, MA, USA). Samples under 20 ng/uL or with A260/230 ratios less than 1.8 were tested for presence of bacteria by PCR amplification of the V4-V5 region of bacterial 16S rRNA using 515F (GTGCCAGCMGCCGCGGTAA) as forward primer and 806R (GGACTACHVGGGTWTCTAAT) as reverse primer. The 25 uL PCR reaction mix for these tests consisted of 5 uL of 1X GC Buffer, 0.5 uL of 10uM dNTPs, 0.5 uL each of 10uM forward and reverse primers, 0.25 uL HotStart DNA polymerase and 17.25 uL RNAse free MilliQ water. PCR thermocycling conditions were an initial denaturation at 94□ for 3 minutes, followed by 35 cycles of: denaturation at 95□ for 30s,

annealing at 50□ for 30s, elongation at 72□ for 1 minute 30s; lastly there was a final elongation step at 72□ for 10 minutes. All tested samples showed amplification at the predicted length of 290 base pairs.

In total, 131 metagenomic libraries were generated. Samples were sequenced on a single lane of Illumina NovaSeq6000 with 2 × 150bp chemistry at Génome Québec (Montréal, Canada).

Sample quality control

The raw sample library size was $14.8 \times 10^6 \pm 2.2 \times 10^6$ reads. Raw reads were trimmed and quality-controlled using Kneaddata. At the time of the experiment, all four moss species sampled in this study lacked reference genomes. Thus, host DNA was partially removed by aligning reads against the closely related genome of *Polytrichum patens*, as well as the human genome hg38 to control for human contamination, using Bowtie2 (from within the Kneaddata pipeline). On average, clean samples contained $10 \times 10^6 \pm 1.6 \times 10^6$ paired-end reads.

Choice of a co-assembly strategy

Metagenome assembly quality depends on depth of coverage of each genome within samples [99]. A common strategy is to co-assemble samples based on biological similarities, by pooling sequences submitted to the assembler, increasing the coverage of genomes, and therefore retrieving more good quality genomes [100]. In this study, we concatenated samples taken from the same moss species and gametophyte section into 46 sample pools. We applied the pipeline described in the following section identically to both assembly and co-assembly strategies. Co-assembly generated 157 MAGs with completeness > 50% and contamination < 10%, compared to 139 through regular sample-level assembly. Given this, we focused on

the MAGs generated by co-assembly, which had a slightly higher average completeness, increasing from 73.6% ± 16.5 to 74.2% ± 15.4%.

Metagenome co-assembly and MAG generation

Samples were co-assembled using MetaWRAP modules [101]. First, paired-end sequences were assembled using metaSPAdes and re-assembled unassembled reads using MEGAHIT. Contigs were then binned with three binners (MetaBAT2, CONCOCT and MaxBin2). Afterwards, these three sets of bins were hybridized with MetaWRAP's bin_refinement module to produce a final set of refined bins for each co-assembly. Bin quality was then assessed using CheckM and bins with completeness < 50% and contamination > 10% were excluded [102], thus yielding a set of 260 bins (5.7 ± 3.3 bins per co-assembly).

To obtain a unique set of species-level MAGs for the whole study, bins were pooled from all co-assemblies and dereplicated them using Drep [103]. Drep uses a two-step clustering strategy, whereby bins are first clustered using Mash [104]; then, within each cluster, bins are dereplicated at a 95% average nucleotide identity (ANI) using MUMmer [105]. GUNC [106] was used to identify chimeric MAGs, which were then removed from the final set of species-level metagenome-assembled genomes (MAGs).

MAG novelty analysis

MAGs which potentially belonged to species with no published reference genome, and therefore were absent from the reference database, were identified by applying three bioinformatic validations. First, we taxonomically annotated MAGs using GTDB-Tk [107]. Second, we calculated the Mash distance between these 156 MAGs and the 31,332 non-redundant bacterial and archaeal genomes from EnsemblBacteria. Third, since Mash distance could be underestimated when comparing incomplete MAGs, we used skani [108] to

compute ANI between these MAGs and the 85,205 GTDB species representative genomes. MAGs were labelled as novel MAGs (nMAGs) if they had an ANI < 0.95 having a Mash distance > 0.05 and no species-level taxonomy assignment. Additionally, after adding the nMAGs to the reference genome index as described below, two were dropped because they were not detected by Sourmash [109]. nMAGs were only kept if they exhibited an assembly quality score (defined as Completeness – 5 × Contamination) > 0.50, yielding a final set of 110 nMAGs.

nMAG annotations

nMAGs were further annotated by predicting their coding sequences using Prodigal [110] and passing these to the MicrobeAnnotator [111] pipeline, generating an estimation of their metabolic potential through pathway completeness reconstruction (see *Pathway completeness analysis* below for further details). Finally, phylogeny was estimated using PhyloPhlAn [112] and built phylogenetic trees using ggtree [113], allowing a cross-validation with their attributed GTDB taxonomy. nMAGs statistics (length, L50, number of contigs and GC contents) were assessed using bbmap.

Taxonomy characterization and abundance estimation of bacterial community

Sample community composition and relative abundance at the species level was estimated using Sourmash gather [109] with a k-mer size of 31, using the GTDB representative species genome index [107] to which nMAGs were added. This approach also measures a sample's containment, which represents the proportion of sample k-mers found in the smallest set of reference genomes required to cover all sample k-mers with an exact match. Of note, k-mers-based community profiling tools generate a *sequence* abundance table, which is not normalized for genome length (as opposed to marker-gene-based tools). However, the

necessity to introduce nMAGs in the reference database precludes the use of marker-gene referencing for taxonomic assignment.

Containment was used to estimate the improvement in sample coverage caused by the addition of nMAGs to the GTDB-rs214 reference index. The GenBank reference index (comprising 1.22M microbial genomes) was also compared and the containment increase is reported for both scenarios (Supplementary Figure S1). Finally, the sequence abundance matrix was built into a phyloseq object using the raw counts; this object was then used as a starting point for all subsequent analyses.

Beta- and alpha diversity analyses

All statistical analyses and plots were made using the open-access software R 4.2.1. For community composition (multivariate) analyses, uneven sequencing depth and heteroskedasticity were addressed by applying variance stabilizing transformations (VST, DESeq2 [114]) to the abundance matrix. This matrix was then used to estimate pairwise Bray-Curtis dissimilarity between samples [115], and to compute principal coordinates using the *capscale* function in the vegan package [116]. Finally, *adonis2* (vegan) was used to run PERMANOVAs [116], assessing the sensitivity of model outputs to the order of its parameters.

In brief, we explored the effect of moss genus, moss species, gametophyte section, and soil abiotic factors on bacterial community composition. We ran models with and without *Dicranum* samples, thus testing the differences in bacterial community composition between moss genera and among *Polytrichum* spp. (*P. juniperinum*, *P. commune* and *P. piliferum*), respectively. This approach allowed us to disentangle whether closely related moss species are likely to harbor more similar microbial taxa than more distantly related ones, because of shared physiological traits and similar ecological niches. In the models with *Dicranum*

samples, we modelled multivariate bacterial community composition as a function of categorical variables '*moss genus*', '*moss species*' nested within '*moss genus*', '*gametophyte section*', and continuous variables '*soil pH*' and '*soil temperature*', and their two-way interactions. We nested '*moss species'* within the genera to account for the hierarchical taxonomic structure and to explore species-specific effects on bacterial community composition and alpha diversity (see below). In the models without *Dicranum* samples, we modelled multivariate bacterial community composition as a function of categorical variables '*moss species*', '*gametophyte section*', and continuous variables '*soil pH*' and '*soil temperature*', and their two-way interactions. A fixed effect '*moss species*' was included in the models, to quantify the portion of variance in bacterial community composition explained by *Polytrichum* spp. In all multivariate models, permutations were restricted within blocks (i.e., 'Location'), using the function *how* of the permute package [117].

For alpha diversity analyses, reads were rarefied to 24,900 (slightly less than the lowest sample sum of reads) using *rarefy_even_depth* (phyloseq) to account for uneven sequencing depth [39, 118]. Linear models were run on Shannon index using function *lmer* in the lme4 [119] package. Then the *Anova* function in the *car* package was used to assess model significance. Final models were further simplified by removing insignificant interactions. For each of the alpha diversity models, the functions *simulateResiduals* and *plot* in the package DHARMa were used to assess model fit [120].

To explore the differences in bacterial diversity among moss identities, gametophyte section and abiotic factors, bacterial diversity (Shannon index) was modelled as a function of categorical variables 'moss *genus*', '*moss species*' (nested within '*moss genus*'), '*gametophyte section*', and continuous variables '*soil pH*' and '*soil temperature*', and their two-way interactions. Finally, models excluding *Dicranum* samples were run to confirm effect persistence intra-genera, for example by including fixed factor '*moss species*'. In all

models, to account for variation among sampling locations, the random effect 'Location' was included, and to account for nested design, we nested 'Microsite' under 'Location'. See Supplementary Table S3 for more details on model structures, fixed and random effects tested, and link functions.

Differential abundance analyses (DAA)

To identify species associated with moss species and gametophyte section, DAA were performed using ANCOM-BC2 package [47]. This approach was chosen because it takes into account the compositionality (e.g., data is commonly expressed in terms of relative abundances) of next-generation sequencing data [121], controls for false discovery rate by accounting for multiple testing and multiple group comparisons, and corrects for taxon- and sample-specific biases (i.e., differences in extraction efficiency across taxa and sampling fraction across samples). In brief, differentially abundant taxa were identified between (1) gametophyte section (green vs. brown), controlling for host moss species, soil pH, and soil temperature; and (2) host moss species, controlling for gametophyte section, using Dunnett's test with *D. undulatum* as a reference group. In the latter, environmental covariates were omitted because they were correlated with location (SoilTemp ~ Location, R^2= 0.58, $p < 1e\text{-}10$) and moss species (SoilpH ~ Host, R^2= 0.62, $p < 1e\text{-}10$), and location is significantly associated with moss species ($\chi^2 = 179.1$, $p < 1e\text{-}10$). For both approaches, we only considered species that were present in at least 10% (n = 13) sample [122]. The Holm procedure was used to account for multiple comparisons. We only reported results that passed ANCOM-BC2's sensitivity test to pseudo-count addition and had a $p_{adj} < 0.01$.

Pathway completeness analysis

Pathway completeness is a value from 0 to 1 that represents the proportion of proteins making up a given pathway (or module) whose coding sequence could be found in a genome. The relationship between pathway completeness and gametophyte green and brown sections was investigated for the 44 differentially abundant species. The tests were performed on pathways from four groups of interest : nitrogen metabolism, photosynthesis, methane metabolism and carbon fixation (Supplementary Table S1), as described in Kyoto Encyclopedia of Genes and Genomes (KEGG) [123]. Pathways for which completeness was ≤ 20% in more than 80% of genomes were excluded, yielding a subset of 15 pathways. Since pathway completeness correlates with genome completeness [124], we controlled for genome completeness (attributing 100% to non-MAGs). Specifically, the model *Module proteins (%) ~ Gametophyte sections (green or brown) + MAG completeness* was tested, using a generalized linear model with a quasibinomial distribution and a logit link function to fit the response variable, which is a proportion that includes 0 and 100%. *p*-values were corrected using the Holm method and only pathways with $p_{adj} < 0.05$ were reported as significant. The heatmap in figure 5 was generated using the '*ComplexHeatmap*' package in R.

<u>Environmental DNA contamination assessment</u>

The quasi absence of non-bacterial genomes in our assemblies prompted us to investigate the high-level taxonomy of sample reads. 100,000 reads per sample were randomly selected and used to perform a BLASTn search against the entire *ncbi_nt* database (across all Kingdoms). Although this approach cannot be used to estimate the relative abundance of non-microbial reads, our data showed that numerous reads belonged to Eukaryotic, non-microbial species (see Supplementary Figures S3 and S4). However, short reads metagenome assembly pipelines generally do not perform well to assemble fungal, archaeal and protist genomes. Therefore, the effective sequencing depth of bacterial species, even after adding our nMAGs

is lower than the sample sequencing depth. Bioinformatic software versions and non-default options are reported in Supplementary Table S4.

Acknowledgements

First and foremost, the authors would like to thank the Cree Community of Nemaska for welcoming our team and allowing us to sample on their land. We thank Nemaska Lithium for their collaboration and support of the project. We would also like to acknowledge J. Chamard and members of S. Roy's Lab (L. Garneau, J. Beaudin, O. Alix-Paré, and S. Malo) for the much-appreciated help with sampling and moss dissections, E. Lussier for extensive support with laboratory work, J.-F. Lucier for assistance with metagenomic analysis, and M. Desrochers for map generation. We are grateful to Génome Québec and to the Digital Research Alliance of Canada for providing us with access to super computers to perform our bioinformatic analyses. We also thank J. Beaudin, P. Moffett, and P. B. Beauregard for critical reading of the manuscript. We are grateful to Isabel Ramirez for beautiful depictions of mosses presented in Figure 1b. This work was supported by funding from (1) Fonds de recherche du Québec – Nature et technologies (FRQNT) Programme partenariat sur le développement durable du secteur minier (Grant 285683 lead PI Pr. Roy and co-PI Pr. Laforest-Lapointe); and (2) Pr. Laforest-Lapointe's Canada Research Chair on Applied Microbial Ecology.

Data availability statement

Scripts for statistical analyses, along with taxonomy tables and metadata used for analysis are available in the Github repository (https://github.com/jorondo1/borealMoss). Raw sequence data are being deposited in the European Nucleotide Archive (ENA).

References

1. Kayes I, Mallik A. Boreal Forests: Distributions, Biodiversity, and Management. In: Leal Filho W, Azul AM, Brandli L, Lange Salvia A, Wall T (eds). *Life on Land*. 2020. Springer International Publishing, Cham, pp 1–12.
2. Walker XJ, Baltzer JL, Cumming SG, Day NJ, Ebert C, Goetz S, et al. Increasing wildfires threaten historic carbon sink of boreal forest soils. *Nature* 2019; **572**: 520–523.
3. Turetsky MR. The role of bryophytes in carbon and nitrogen cycling. *The Bryologist* 2003; **106**: 395–409.
4. Holland-Moritz H, Stuart J, Lewis LR, Miller S, Mack MC, McDaniel SF, et al. Novel bacterial lineages associated with boreal moss species. *Environmental Microbiology* 2018; **20**: 2625–2638.
5. Ininbergs K, Bay G, Rasmussen U, Wardle DA, Nilsson M-C. Composition and diversity of nifH genes of nitrogen-fixing cyanobacteria associated with boreal forest feather mosses. *New Phytologist* 2011; **192**: 507–517.
6. Rousk K, Michelsen A. Ecosystem nitrogen fixation throughout the snow-free period in subarctic tundra: effects of willow and birch litter addition and warming. *Global Change Biology* 2017; **23**: 1552–1563.
7. Klarenberg IJ, Keuschnig C, Russi Colmenares AJ, Warshan D, Jungblut AD, Jónsdóttir IS, et al. Long-term warming effects on the microbiome and nifH gene abundance of a common moss species in sub-Arctic tundra. *New Phytologist* 2022; **234**: 2044–2056.
8. Jean M, Holland-Moritz H, Melvin AM, Johnstone JF, Mack MC. Experimental assessment of tree canopy and leaf litter controls on the microbiome and nitrogen fixation rates of two boreal mosses. *New Phytologist* 2020; **227**: 1335–1349.
9. Tveit AT, Kiss A, Winkel M, Horn F, Hájek T, Svenning MM, et al. Environmental patterns of brown moss- and Sphagnum-associated microbial communities. *Sci Rep* 2020; **10**: 22412.

690 10. Renaudin M, Blasi C, Bradley RL, Bellenger J-P. New insights into the drivers of moss-associated
691 nitrogen fixation and cyanobacterial biomass in the eastern Canadian boreal forest. *Journal of*
692 *Ecology* 2022; **110**: 1403–1418.
693 11. Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Pean C, Chen Y, et al. Climate Change 2021:
694 The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of
695 the Intergovernmental Panel on Climate Change. 2021. Intergovernmental Panel on Climate
696 Change, Switzerland.
697 12. Martin M, Grondin P, Lambert M-C, Bergeron Y, Morin H. Compared to Wildfire, Management
698 Practices Reduced Old-Growth Forest Diversity and Functionality in Primary Boreal Landscapes
699 of Eastern Canada. *Front For Glob Change* 2021; **4**.
700 13. Yin X, Martineau C, Fenton NJ. How big is the footprint? Quantifying offsite effects of mines on
701 boreal plant communities. *Global Ecology and Conservation* 2023; **41**: e02372.
702 14. Jonsson M, Kardol P, Gundale MJ, Bansal S, Nilsson M-C, Metcalfe DB, et al. Direct and Indirect
703 Drivers of Moss Community Structure, Function, and Associated Microfauna Across a
704 Successional Gradient. *Ecosystems* 2015; **18**: 154–169.
705 15. Royles J, Griffiths H. Invited review: climate change impacts in polar regions: lessons from
706 Antarctic moss bank archives. *Global Change Biology* 2015; **21**: 1041–1057.
707 16. Klarenberg IJ, Keuschnig C, Colmenares AJR, Warshan D, Jungblut AD, Jónsdóttir IS, et al. Long-
708 term warming effects on the microbiome and nifH gene abundance of a common moss species
709 in sub-Arctic tundra. *New Phytologist* 2021; **234**: 2044–2056.
710 17. Bragina A, Berg C, Cardinale M, Shcherbakov A, Chebotar V, Berg G. Sphagnum mosses harbour
711 highly specific bacterial diversity during their whole lifecycle. *ISME J* 2012; **6**: 802–813.
712 18. Holland-Moritz H, Stuart JEM, Lewis LR, Miller SN, Mack MC, Ponciano JM, et al. The bacterial
713 communities of Alaskan mosses and their contributions to N2-fixation. *Microbiome* 2021; **9**: 53.

714 19. McLaren JR, Buckeridge KM, van de Weg MJ, Shaver GR, Schimel JP, Gough L. Shrub
715 encroachment in Arctic tundra: Betula nana effects on above- and belowground litter
716 decomposition. *Ecology* 2017; **98**: 1361–1376.
717 20. Cao W, Xiong Y, Zhao D, Tan H, Qu J. Bryophytes and the symbiotic microorganisms, the
718 pioneers of vegetation restoration in karst rocky desertification areas in southwestern China.
719 *Appl Microbiol Biotechnol* 2020; **104**: 873–891.
720 21. Tian Y, Li YH. Comparative analysis of bacteria associated with different mosses by 16S rRNA
721 and 16S rDNA sequencing. *J Basic Microbiol* 2017; **57**: 57–67.
722 22. Wang Y, Xue D, Chen X, Qiu Q, Chen H. Structure and Functions of Endophytic Bacterial
723 Communities Associated with Sphagnum Mosses and Their Drivers in Two Different Nutrient
724 Types of Peatlands. *Microb Ecol* 2024; **87**: 47.
725 23. Holland-Moritz H, Stuart JEM, Lewis LR, Miller SN, Mack MC, Ponciano JM, et al. The bacterial
726 communities of Alaskan mosses and their contributions to N2-fixation. *Microbiome* 2021; **9**: 1–
727 14.
728 24. Elzen E van den, Bengtsson F, Fritz C, Rydin H, Lamers LPM. Variation in symbiotic N2 fixation
729 rates among Sphagnum mosses. *PLOS ONE* 2020; **15**: e0228383.
730 25. Kostka JE, Weston DJ, Glass JB, Lilleskov EA, Shaw AJ, Turetsky MR. The Sphagnum microbiome:
731 new insights from an ancient plant lineage. *New Phytologist* 2016; **211**: 57–64.
732 26. Stuart JEM, Holland-Moritz H, Jean M, Miller SN, Ponciano JM, McDaniel SF, et al. The
733 relationship of C and N stable isotopes to high-latitude moss-associated N2 fixation. *Oecologia*
734 2021; **197**: 283–295.
735 27. Macdonald SE, Fenniak TE. Understory plant communities of boreal mixedwood forests in
736 western Canada: Natural patterns and response to variable-retention harvesting. *Forest*
737 *Ecology and Management* 2007; **242**: 34–48.
738 28. Liu C, Pouliot K, Roy S, Rochefort L. Moss regeneration for lithium mine waste rock
739 revegetation in Québec, Canada. *Botany* 2024; **102**: 26–40.

740 29. Lindo Z, Gonzalez A. The bryosphere: an integral and influential component of the Earth's
741 biosphere. *Ecosystems* 2010; **13**: 612–627.
742 30. Arróniz-Crespo M, Bougoure J, Murphy DV, Cutler NA, Souza-Egipsy V, Chaput DL, et al.
743 Revealing the transfer pathways of cyanobacterial-fixed N into the boreal forest through the
744 feather-moss microbiome. *Front Plant Sci* 2022; **13**: 1036258.
745 31. Chen K-H, Liao H-L, Bellenger J-P, Lutzoni F. Differential gene expression associated with fungal
746 trophic shifts along the senescence gradient of the moss Dicranum scoparium. *Environmental*
747 *Microbiology* 2019; **21**: 2273–2289.
748 32. Solheim B, Zielke M. Associations Between Cyanobacteria and Mosses. In: Rai AN, Bergman B,
749 Rasmussen U (eds). *Cyanobacteria in Symbiosis*. 2002. Springer Netherlands, Dordrecht, pp
750 137–152.
751 33. Liu X, Rousk K. The moss traits that rule cyanobacterial colonization. *Annals of Botany* 2022;
752 **129**: 147–160.
753 34. Renaudin M, Blasi C, Bradley RL, Bellenger J-P. New insights into the drivers of moss-associated
754 nitrogen fixation and cyanobacterial biomass in the eastern Canadian boreal forest. *Journal of*
755 *Ecology* 2022; **110**: 1403–1418.
756 35. Shen JC, Zhang ZH, Liu R, Wang ZH. Ecological restoration of eroded karst utilizing pioneer moss
757 and vascular plant species with selection based on vegetation diversity and underlying soil
758 chemistry. *International Journal of Phytoremediation* 2018; **20**: 1369–1379.
759 36. Mohanasundaram B, Pandey S. Effect of environmental signals on growth and development in
760 mosses. *Journal of Experimental Botany* 2022; **73**: 4514–4527.
761 37. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling
762 to analysis. *Nat Biotechnol* 2017; **35**: 833–844.
763 38. Thomas AM, Segata N. Multiple levels of the unknown in microbiome research. *BMC Biology*
764 2019; **17**: 48.

765 39. Wicaksono WA, Cernava T, Berg C, Berg G. Bog ecosystems as a playground for plant–microbe
766 coevolution: bryophytes and vascular plants harbour functionally adapted bacteria.
767 *Microbiome* 2021; **9**: 170.
768 40. Budke JM, Bernard EC, Gray DJ, Huttunen S, Piechulla B, Trigiano RN. Introduction to the
769 Special Issue on Bryophytes. *Critical Reviews in Plant Sciences* 2018; **37**: 102–112.
770 41. Halpern CB, Dovčiak M, Urgenson LS, Evans SA. Substrates mediate responses of forest
771 bryophytes to a gradient in overstory retention. *Canadian Journal of Forest Research* 2014; **44**:
772 855–866.
773 42. Ward LM, Cardona T, Holland-Moritz H. Evolutionary Implications of Anoxygenic Phototrophy
774 in the Bacterial Phylum Candidatus Eremiobacterota (WPS-2). *Front Microbiol* 2019; **10**.
775 43. Ji M, Greening C, Vanwonterghem I, Carere CR, Bay SK, Steen JA, et al. Atmospheric trace gases
776 support primary production in Antarctic desert surface soil. *Nature* 2017; **552**: 400–403.
777 44. Sheremet A, Jones GM, Jarett J, Bowers RM, Bedard I, Culham C, et al. Ecological and genomic
778 analyses of candidate phylum WPS-2 bacteria in an unvegetated soil. *Environmental*
779 *Microbiology* 2020; **22**: 3143–3157.
780 45. Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, Zayed AAF, et al. Genome-centric
781 view of carbon processing in thawing permafrost. *Nature* 2018; **560**: 49–54.
782 46. Ji M, Williams TJ, Montgomery K, Wong HL, Zaugg J, Berengut JF, et al. Candidatus
783 Eremiobacterota, a metabolically and phylogenetically diverse terrestrial phylum with acid-
784 tolerant adaptations. *ISME J* 2021; **15**: 2692–2707.
785 47. Lin H, Peddada SD. Multigroup analysis of compositions of microbiomes with covariate
786 adjustments and repeated measures. *Nat Methods* 2024; **21**: 83–91.
787 48. Brenac L, Baidoo EEK, Keasling JD, Budin I. Distinct functional roles for hopanoid composition in
788 the chemical tolerance of Zymomonas mobilis. *Molecular Microbiology* 2019; **112**: 1564–1575.

789 49. Kulkarni G, Busset N, Molinaro A, Gargani D, Chaintreuil C, Silipo A, et al. Specific Hopanoid
790 Classes Differentially Affect Free-Living and Symbiotic States of Bradyrhizobium diazoefficiens.
791 *mBio* 2015; **6**: 10.1128/mbio.01251-15.
792 50. Schmerk CL, Bernards MA, Valvano MA. Hopanoid Production Is Required for Low-pH
793 Tolerance, Antimicrobial Resistance, and Motility in Burkholderia cenocepacia. *Journal of*
794 *Bacteriology* 2011; **193**: 6712–6723.
795 51. Dobritsa SV, Potter D, Gookin TE, Berry AM. Hopanoid lipids in Frankia: identification of
796 squalene-hopene cyclase gene sequences. *Can J Microbiol* 2001; **47**: 535–540.
797 52. Cornejo-Castillo FM, Zehr JP. Hopanoid lipids may facilitate aerobic nitrogen fixation in the
798 ocean. *Proc Natl Acad Sci U S A* 2019; **116**: 18269–18271.
799 53. Kiss L. Chapter 3 Intracellular mycoparasites in action: Interactions between powdery mildew
800 fungi and *Ampelomyces*. *British Mycological Society Symposia Series*. 2008. Elsevier, pp 37–52.
801 54. Klavina L, Springe G, Nikolajeva V, Martsinkevich I, Nakurte I, Dzabijeva D, et al. Chemical
802 Composition Analysis, Antimicrobial Activity and Cytotoxicity Screening of Moss Extracts (Moss
803 Phytochemistry). *Molecules* 2015; **20**: 17221–17243.
804 55. Ortiz-Rivero J, Garrido-Benavent I, Heiðmarsson S, de los Ríos A. Moss and Liverwort Covers
805 Structure Soil Bacterial and Fungal Communities Differently in the Icelandic Highlands. *Microb*
806 *Ecol* 2023; **86**: 1893–1908.
807 56. Ray AE, Zaugg J, Benaud N, Chelliah DS, Bay S, Wong HL, et al. Atmospheric chemosynthesis is
808 phylogenetically and geographically widespread and contributes significantly to carbon fixation
809 throughout cold deserts. *ISME J* 2022; **16**: 2547–2560.
810 57. Taş N, Conejo N, Salmon VG. Metagenomes and Metagenome-Assembled Genome Sequences
811 from Nitrogen-Fixing Alder Nodules. *Microbiol Resour Announc* 2023; **12**: e01266-22.
812 58. Gordon C, Wynn JM, Woodin SJ. Impacts of increased nitrogen supply on high Arctic heath: the
813 importance of bryophytes and phosphorus availability. *New Phytologist* 2001; **149**: 461–471.

814 59. Belova SE, Suzina NE, Rijpstra WIC, Sinninghe Damsté JS, Dedysh SN. Edaphobacter lichenicola
815 sp. nov., a member of the family Acidobacteriaceae from lichen-dominated forested tundra.
816 *International Journal of Systematic and Evolutionary Microbiology* 2018; **68**: 1265–1270.
817 60. Koch IH, Gich F, Dunfield PF, Overmann J. Edaphobacter modestus gen. nov., sp. nov., and
818 Edaphobacter aggregans sp. nov., acidobacteria isolated from alpine and forest soils.
819 *International Journal of Systematic and Evolutionary Microbiology* 2008; **58**: 1114–1122.
820 61. Shange RS, Ankumah RO, Ibekwe AM, Zabawa R, Dowd SE. Distinct Soil Bacterial Communities
821 Revealed under a Diversely Managed Agroecosystem. *PLOS ONE* 2012; **7**: e40338.
822 62. Chong CW, Convey P, Pearce DA, Tan IKP. Assessment of soil bacterial communities on
823 Alexander Island (in the maritime and continental Antarctic transitional zone). *Polar Biol* 2012;
824 **35**: 387–399.
825 63. Rapoport D, Sagova-Mareckova M, Sedláček I, Provaznik J, Králová S, Pavlinic D, et al. Trebonia
826 kvetii gen. nov., sp. nov., an acidophilic actinobacterium, and proposal of the new
827 actinobacterial family Treboniaceae fam. nov. *Int J Syst Evol Microbiol* 2020; **70**: 5106–5114.
828 64. Weon H-Y, Yoo S-H, Kim Y-J, Lee C-M, Kim B-Y, Jeon Y-A, et al. Rudaea cellulosilytica gen. nov.,
829 sp. nov., isolated from soil. *International Journal of Systematic and Evolutionary Microbiology*
830 2009; **59**: 2308–2312.
831 65. Mieszkin S, Trouche B, Ancousture J, Raouf Y, Uroz S, Alain K. Scleromatobacter humisilvae gen.
832 nov., sp. nov., a novel bacterium isolated from oak forest soil. *International Journal of*
833 *Systematic and Evolutionary Microbiology* 2023; **73**: 005793.
834 66. Klarenberg IJ, Keuschnig C, Salazar A, Benning LG, Vilhelmsson O. Moss and underlying soil
835 bacterial community structures are linked to moss functional traits. *Ecosphere* 2023; **14**: e4447.
836 67. Tagirdzhanova G, Saary P, Cameron E, Garber A, Díaz Escandón D, Goyette S, et al. Evidence for
837 a core set of microbial lichen symbionts from a global survey of metagenome. 2023.

838 68. Hay MC, Mitchell AC, Soares AR, Debbonaire AR, Mogrovejo DC, Els N, et al. Metagenome-
839 assembled genomes from High Arctic glaciers highlight the vulnerability of glacier-associated
840 microbiota and their activities to habitat loss. *Microb Genom* 2023; **9**: 001131.
841 69. Yurimoto H, Shiraishi K, Sakai Y. Physiology of Methylotrophs Living in the Phyllosphere.
842 *Microorganisms* 2021; **9**: 809.
843 70. Leducq J-B, Sneddon D, Santos M, Condrain-Morel D, Bourret G, Martinez-Gomez NC, et al.
844 Comprehensive Phylogenomics of Methylobacterium Reveals Four Evolutionary Distinct
845 Groups and Underappreciated Phyllosphere Diversity. *Genome Biology and Evolution* 2022; **14**:
846 evac123.
847 71. Harris BJ, Harrison CJ, Hetherington AM, Williams TA. Phylogenomic Evidence for the
848 Monophyly of Bryophytes and the Reductive Evolution of Stomata. *Current Biology* 2020; **30**:
849 2001-2012.e2.
850 72. Bragina A, Berg C, Berg G. The core microbiome bonds the Alpine bog vegetation to a
851 transkingdom metacommunity. *Molecular Ecology* 2015; **24**: 4795–4807.
852 73. Schauer S, Kutschera U. A novel growth-promoting microbe, Methylobacterium funariae sp.
853 nov., isolated from the leaf surface of a common moss. *Plant Signaling & Behavior* 2011; **6**:
854 510–515.
855 74. Gao Y-Q, Yu J-L, Ren D-M, Zhao D-P, Zhang D-J. Endophytic Methylobacterium in Tissue Culture
856 of the Moss Didymodon tectorum. *anbf* 2022; **59**: 117–122.
857 75. Noh H-J, Park Y, Hong SG, Lee YM. Diversity and Physiological Characteristics of Antarctic
858 Lichens-Associated Bacteria. *Microorganisms* 2021; **9**: 607.
859 76. Noh H-J, Shin SC, Park Y, Choi A, Baek K, Hong SG, et al. Lichenicola cladoniae gen. nov., sp.
860 nov., a member of the family Acetobacteraceae isolated from an Antarctic lichen. *International*
861 *Journal of Systematic and Evolutionary Microbiology* 2020; **70**: 5918–5925.

862 77. Warshan D, Bay G, Nahar N, Wardle DA, Nilsson M-C, Rasmussen U. Seasonal variation in nifH
863 abundance and expression of cyanobacterial communities associated with boreal feather
864 mosses. *ISME J* 2016; **10**: 2198–2208.
865 78. Pederson ERA, Warshan D, Rasmussen U. Genome Sequencing of Pleurozium schreberi: The
866 Assembled and Annotated Draft Genome of a Pleurocarpous Feather Moss. *G3*
867 *Genes|Genomes|Genetics* 2019; **9**: 2791–2797.
868 79. Bižić M, Klintzsch T, Ionescu D, Hindiyeh MY, Günthel M, Muro-Pastor AM, et al. Aquatic and
869 terrestrial cyanobacteria produce methane. *Science Advances* 2020; **6**: eaax5343.
870 80. Renaudin M, Laforest-Lapointe I, Bellenger J-P. Unraveling global and diazotrophic bacteriomes
871 of boreal forest floor feather mosses and their environmental drivers at the ecosystem and at
872 the plant scale in North America. *Science of The Total Environment* 2022; **837**: 155761.
873 81. Gavazov KS, Soudzilovskaia NA, van Logtestijn RSP, Braster M, Cornelissen JHC. Isotopic
874 analysis of cyanobacterial nitrogen fixation associated with subarctic lichen and bryophyte
875 species. *Plant Soil* 2010; **333**: 507–517.
876 82. Garrido-Oter R, Nakano RT, Dombrowski N, Ma K-W, AgBiome Team, McHardy AC, et al.
877 Modular Traits of the Rhizobiales Root Microbiota and Their Evolutionary Relationship with
878 Symbiotic Rhizobia. *Cell Host Microbe* 2018; **24**: 155-167.e5.
879 83. Kolton M, Weston DJ, Mayali X, Weber PK, McFarlane KJ, Pett-Ridge J, et al. Defining the
880 Sphagnum Core Microbiome across the North American Continent Reveals a Central Role for
881 Diazotrophic Methanotrophs in the Nitrogen and Carbon Cycles of Boreal Peatland Ecosystems.
882 *mBio* 2022; **13**: e03714-21.
883 84. Zervas A, Zeng Y, Madsen AM, Hansen LH. Genomics of Aerobic Photoheterotrophs in Wheat
884 Phyllosphere Reveals Divergent Evolutionary Patterns of Photosynthetic Genes in
885 Methylobacterium spp. *Genome Biology and Evolution* 2019; **11**: 2895–2908.
886 85. Ingram-Smith C, Martin SR, Smith KS. Acetate kinase: not just a bacterial enzyme. *Trends in*
887 *Microbiology* 2006; **14**: 249–253.

888 86. Smith AR, Kieft B, Mueller R, Fisk MR, Mason OU, Popa R, et al. Carbon fixation and energy
889 metabolisms of a subseafloor olivine biofilm. *ISME J* 2019; **13**: 1737–1749.
890 87. Estrada-De Los Santos P, Bustillos-Cristales R, Caballero-Mellado J. Burkholderia, a Genus Rich
891 in Plant-Associated Nitrogen Fixers with Wide Environmental and Geographic Distribution.
892 *Applied and Environmental Microbiology* 2001; **67**: 2790–2798.
893 88. Li J, Yu C, Liu Z, Wang Y, Wang F. Microplastic accelerate the phosphorus-related metabolism of
894 bacteria to promote the decomposition of methylphosphonate to methane. *Sci Total Environ*
895 2023; **858**: 160020.
896 89. Ney B, Ahmed FH, Carere CR, Biswas A, Warden AC, Morales SE, et al. The methanogenic redox
897 cofactor F420 is widely synthesized by aerobic soil bacteria. *ISME J* 2017; **11**: 125–137.
898 90. Wu H-H, Pun MD, Wise CE, Streit BR, Mus F, Berim A, et al. The pathway for coenzyme M
899 biosynthesis in bacteria. *Proceedings of the National Academy of Sciences* 2022; **119**:
900 e2207190119.
901 91. Yang Y, Cheng S, Fang H, Guo Y, Li Y, Zhou Y, et al. Linkages between the molecular composition
902 of dissolved organic matter and soil microbial community in a boreal forest during freeze–thaw
903 cycles. *Front Microbiol* 2023; **13**.
904 92. Ikeda MA, Nakamura H, Sawada K. Aliphatic hydrocarbons in the lichen class Lecanoromycetes
905 and their potential use as chemotaxonomic indicators and biomarkers. *Organic Geochemistry*
906 2023; **179**: 104588.
907 93. Lindo Z, Nilsson M-C, Gundale MJ. Bryophyte-cyanobacteria associations as regulators of the
908 northern latitude carbon balance in response to global change. *Glob Chang Biol* 2013; **19**:
909 2022–2035.
910 94. Leppänen SM, Salemaa M, Smolander A, Mäkipää R, Tiirola M. Nitrogen fixation and
911 methanotrophy in forest mosses along a N deposition gradient. *Environmental and*
912 *Experimental Botany* 2013; **90**: 62–69.

913 95. Moussaoui L, Leduc A, Girona MM, Bélisle AC, Lafleur B, Fenton NJ, et al. Success Factors for
914 Experimental Partial Harvesting in Unmanaged Boreal Forest: 10-Year Stand Yield Results.
915 *Forests* 2020; **11**: 1199.
916 96. Environment and Climate Change Canada Historical Climate Data. Monthly Climate Summaries.
917 2021.
918 97. Collin A, Messier C, Kembel SW, Bélanger N. Can sugar maple establish into the boreal forest?
919 Insights from seedlings under various canopies in southern Quebec. *Ecosphere* 2018; **9**:
920 e02022.
921 98. Carter MR, Gregorich EG (eds). Soil Sampling and Methods of Analysis, 2nd ed. 2007. CRC
922 Press, Boca Raton.
923 99. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key
924 considerations in genomic analyses. *Nat Rev Genet* 2014; **15**: 121–132.
925 100. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of
926 nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat*
927 *Microbiol* 2017; **2**: 1533–1542.
928 101. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved
929 metagenomic data analysis. *Microbiome* 2018; **6**: 158.
930 102. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al.
931 Minimum information about a single amplified genome (MISAG) and a metagenome-
932 assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 2017; **35**: 725–731.
933 103. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic
934 comparisons that enables improved genome recovery from metagenomes through de-
935 replication. *ISME J* 2017; **11**: 2864–2868.
936 104. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast
937 genome and metagenome distance estimation using MinHash. *Genome Biology* 2016; **17**: 132.

938 105. Cirkoska E, Kishore C, Bose SC, Mubarakali A, Reshmi TR, Marina N. A Study on Suffix Trees and
939 Their Applications in Genome Sequences Using MUMmer. In: Smys S, Palanisamy R, Rocha Á,
940 Beligiannis GN (eds). *Computer Networks and Inventive Communication Technologies*. 2021.
941 Springer Nature, Singapore, pp 15–28.
942 106. Orakov A, Fullam A, Coelho LP, Khedkar S, Szklarczyk D, Mende DR, et al. GUNC: detection of
943 chimerism and contamination in prokaryotic genomes. *Genome Biology* 2021; **22**: 178.
944 107. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with
945 the Genome Taxonomy Database. *Bioinformatics* 2020; **36**: 1925–1927.
946 108. Shaw J, Yu YW. Fast and robust metagenomic sequence comparison through sparse chaining
947 with skani. *Nat Methods* 2023; **20**: 1661–1665.
948 109. Brown CT, Irber L. sourmash: a library for MinHash sketching of DNA. *Journal of Open Source*
949 *Software* 2016; **1**: 27.
950 110. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene
951 recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**: 119.
952 111. Ruiz-Perez CA, Conrad RE, Konstantinidis KT. MicrobeAnnotator: a user-friendly,
953 comprehensive functional annotation pipeline for microbial genomes. *BMC Bioinformatics*
954 2021; **22**: 11.
955 112. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, et al. Precise phylogenetic
956 analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat*
957 *Commun* 2020; **11**: 2500.
958 113. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation
959 of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and*
960 *Evolution* 2017; **8**: 28–36.
961 114. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq
962 data with DESeq2. *Genome Biology* 2014; **15**: 550.

963 115. Tian W, Wang H, Xiang X, Wang R, Xu Y. Structural Variations of Bacterial Community Driven by
964 Sphagnum Microhabitat Differentiation in a Subalpine Peatland. *Frontiers in Microbiology*
965 2019; **10**: 1–11.
966 116. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. vegan: Community
967 Ecology Package. 2022.
968 117. Simpson GL. permute: Functions for Generating Restricted Permutations of Data. R package
969 version 0.9-5. 2019.
970 118. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and
971 Graphics of Microbiome Census Data. *PLOS ONE* 2013; **8**: e61217.
972 119. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *Journal*
973 *of Statistical Software* 2015; **67**: 1–48.
974 120. Hartig F. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models.
975 2020.
976 121. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are
977 Compositional: And This Is Not Optional. *Front Microbiol* 2017; **8**.
978 122. Nearing JT, Douglas GM, Hayes MG, MacDonald J, Desai DK, Allward N, et al. Microbiome
979 differential abundance methods produce different results across 38 datasets. *Nat Commun*
980 2022; **13**: 342.
981 123. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000;
982 **28**: 27–30.
983 124. Eisenhofer R, Odriozola I, Alberdi A. Impact of microbial genome completeness on
984 metagenomic functional inference. *ISME Communications* 2023; **3**: 12.
985

Author contributions

SI, SR, and ILL designed the experiment. SI conducted the sample collection, processing, and molecular work. JRL developed the bioinformatic workflow and conducted the bioinformatic analyses with the contribution from SI, MF, and ILL. SI and JRL analyzed the data with support from MF and ILL. SI wrote the first draft with the contribution from JRL, MF, ILL and SR. All authors reviewed the final manuscript.

Additional Information

Authors declare no competing interests.

Table legends

Table 1. Moss genus, moss species nested within moss genus, gametophyte section, soil pH, soil temperature, and their interactions drive moss bacterial community composition. Shown are the PERMANOVA on Bray-Curtis dissimilarity (n = 131 and n=121 samples respectively) degrees of freedom (df), determination coefficients (R^2), and p-values (significant values are shown in bold).

Table 2. The effects of moss genus, moss species nested within moss genus, gametophyte section, soil pH, and temperature on moss bacterial alpha-diversity (Shannon index). Shown are the linear mixed-effects model (n = 131 and n=121 samples respectively) test statistics (χ^2), degrees of freedom (df), and p-values (significant values are shown in bold).

Figure legends

Figure 1. Sampling design, location, and collected moss species.

Panel (a) shows the sampling layout: at each location, three microsites were chosen at random (indicated by yellow circles). Within each microsite, three replicate samples were collected (shown as green circles). Inset in the top right of Panel (a) shows a map of Québec with the sampling sites located within the continuous boreal forest region. Panel (b) shows the illustrations of each of four moss species (drawings were created by Isabel Ramirez). Panel (c) shows the images of green and brown sections of the moss gametophyte.

Figure 2. (a, b) Two-dimensional principal coordinate analysis (PCoA) of bacterial communities based on Bray-Curtis dissimilarity index, calculated using vst-transformed abundance table. Each point represents one moss sample, and ellipses indicate a 95% confidence interval around moss samples that represent same moss species (a) or from the same moss gametophyte section (b). (c) Differences in alpha diversity (Shannon diversity) of moss bacterial communities. Boxplots are faceted according to moss species. Showed are

four moss species: *P. juniperinum* (nB = 24, nG = 22), *P. commune* (nB = 9, nG = 9), *P. piliferum* (nB = 26, nG = 26), and *D. undulatum* (nB = 8, nG = 7). (d) Bacterial taxonomic profiles between green and brown gametophyte sections and among moss species. Bar charts are faceted by moss species, and bar colours represent different microbial taxa.

Figure 3. Novel species-level metagenome-assembled genomes obtained with co-assembly. From inner to outer layers: proportion of guanine and cytidine in assembly nucleotides (%□GC), genome size (length, Mbp), number of contigs in each MAG, length of the shortest of the largest contigs required to cover 50% of the total assembly length (N50) and quality score (QC). The inner section shows the phylogenetic tree of 110 metagenome-assembled genomes constructed using *PhyloPhlAn* and built phylogenetic trees using *ggtree* packages, allowing a cross-validation with their assigned taxonomy in the GTDB database. The different colored dots on the branches of the tree correspond to the colors of the bacterial orders listed in the legend.

Figure 4. Differential abundance of bacterial species between moss species and between green and brown gametophyte sections.

LFC: $\log_2$-fold change in absolute abundances, with all values showing significant differential abundance ($p < 0.01$). (a) Differentially abundant bacterial species in moss species relative to *D. undulatum*. LFC are relative to *D. undulatum*. Positive LFC (blue) indicates enrichment in specified moss species relative to *D. undulatum*, whereas negative LFC (red) indicates enrichment in *D. undulatum* relative to specified moss species. (b) Bacterial species that are differentially abundant in either green or brown sections of moss gametophyte. Bar colours indicate in which section (green or brown) the species was differentially enriched, with LFC calculated in relation to the reference group. Blank cells

correspond to species for which no significant differences in abundance were detected, and therefore no LFC values reported.

Figure 5. Metabolic abilities of the differentially abundant species in brown and green gametophyte sections.

The pathway completeness of selected metabolic pathway groups, such as photosynthesis, nitrogen and methane metabolism, and prokaryotic carbon fixation, is indicated for each genome (blue scale). This figure also shows the taxonomic rank (e.g., bacterial order) for each of 44 differentially abundant genomes.

Table 1. Moss genus, moss species nested within moss genus, gametophyte section, soil pH, soil temperature, and their interactions drive moss bacterial community composition. Shown are the PERMANOVA on Bray-Curtis dissimilarity (n = 131 and n=121 samples respectively) degrees of freedom (df), determination coefficients (R^2), and p-values (significant values are shown in bold).

		a) with *Dicranum sp.*			b) without *Dicranum sp.*		
	Variables	df	R^2 (%)	p-value	df	R^2 (%)	p-value
	Moss Genus	1	13.8	**0.001**	-	-	-
	Moss Species	-	-	-	2	19.7	**0.001**
1st level	Gametophyte Section	1	12.7	**0.001**	1	17.2	**0.001**
	Soil Temperature	1	7.4	**0.001**	1	6.9	**0.001**
	Soil pH	1	5.1	**0.001**	1	4.1	**0.001**
	Genus × Species	2	12.3	**0.001**	-	-	-
	Genus × Gametophyte Section	1	2.6	**0.001**	-	-	-
	Genus × Soil Temperature	1	0.6	0.232	-	-	-
	Genus × Soil pH	1	0.2	0.789	-	-	-
2nd level	Species × Gametophyte Section	-	-	-	2	1.3	**0.042**
	Species × Soil Temperature	-	-	-	2	3.0	**0.003**
	Species × Soil pH	-	-	-	2	3.0	**0.001**
	Gametophyte Section × Soil Temp.	1	0.9	**0.003**	1	1.1	**0.011**
	Gametophyte Section × Soil pH	1	0.3	0.447	1	0.2	0.749
	Total (%)		**57.3**			59	

Table 2. The effects of moss genus, moss species nested within moss genus, gametophyte section, soil pH, and temperature on moss bacterial alpha diversity (Shannon index). Shown are the linear mixed-effects model (n = 131 and n=121 samples respectively) test statistics (χ^2), degrees of freedom (df), and p-values (significant values are shown in bold).

		a) *Dicranum sp.*			b) without *Dicranum sp.*		
	Variables	df	χ2 (%)	p-value	df	χ2 (%)	p-value
1st level	Moss Genus	1	8.6	**0.003**	-	-	-
	Moss Species	-	-	-	2	0.08	0.960
	Gametophyte Section	1	29.3	<0.001	1	35.7	**<0.001**
	Soil Temperature	1	6.3	**0.012**	1	6.6	**0.010**
	Soil pH	1	1.2	0.279	1	1.7	0.192
2nd level	Genus × Species	2	0.07	0.966	-	-	-
	Genus × Gametophyte Section	1	2.1	0.148			
	Genus × Soil Temperature	1	0.04	0.840	-	-	-
	Genus × Soil pH	1	0.08	0.784	-	-	-
	Species × Gametophyte Section	-	-	-	2	2.8	0.246
	Species × Soil Temperature	-	-	-	2	2.3	0.320
	Gametophyte Section × Soil Temp.	1	1.7	0.194	1	3.0	0.084
	Gametophyte Section × Soil pH	1	12.6	**<0.001**	1	6.8	**0.009**
	Total R^2 (%)		**62**			**59**	

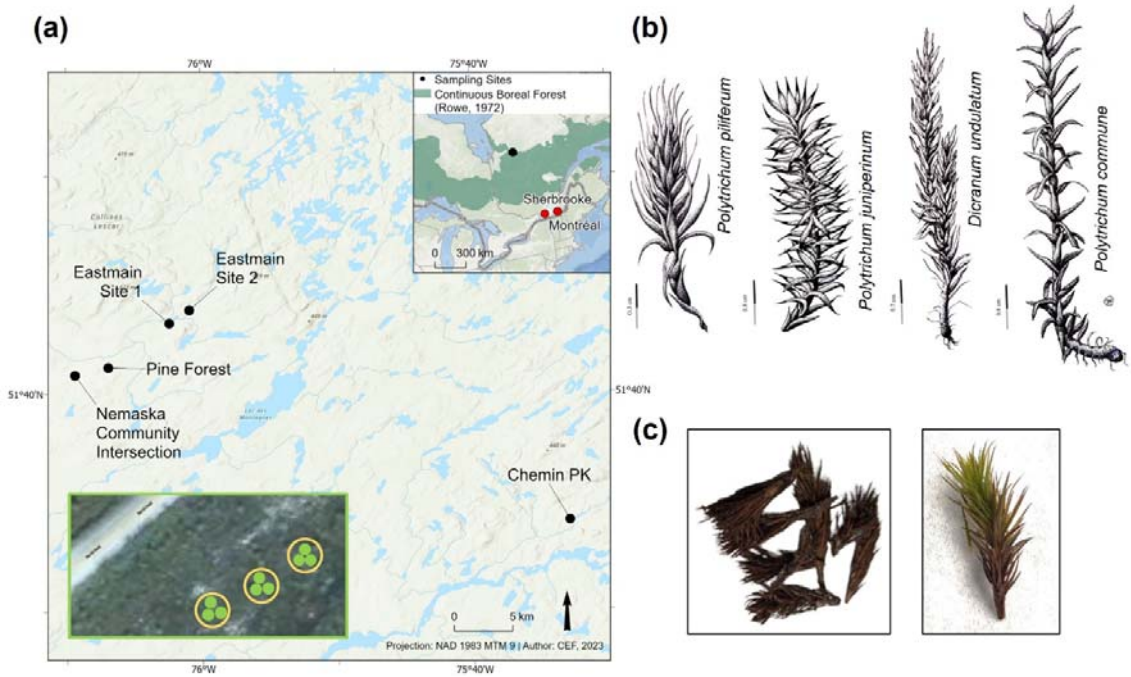


Figure 1. Sampling design, location, and collected moss species.

Panel (a) shows the sampling layout: at each location, three microsites were chosen at random (indicated by yellow circles). Within each microsite, three replicate samples were collected (shown as green circles). Inset in the top right of Panel (a) shows a map of Québec with the sampling sites located within the continuous boreal forest region. Panel (b) shows the illustrations of each of four moss species (drawings were created by Isabel Ramirez). Panel (c) shows the images of green and brown sections of the moss gametophyte.

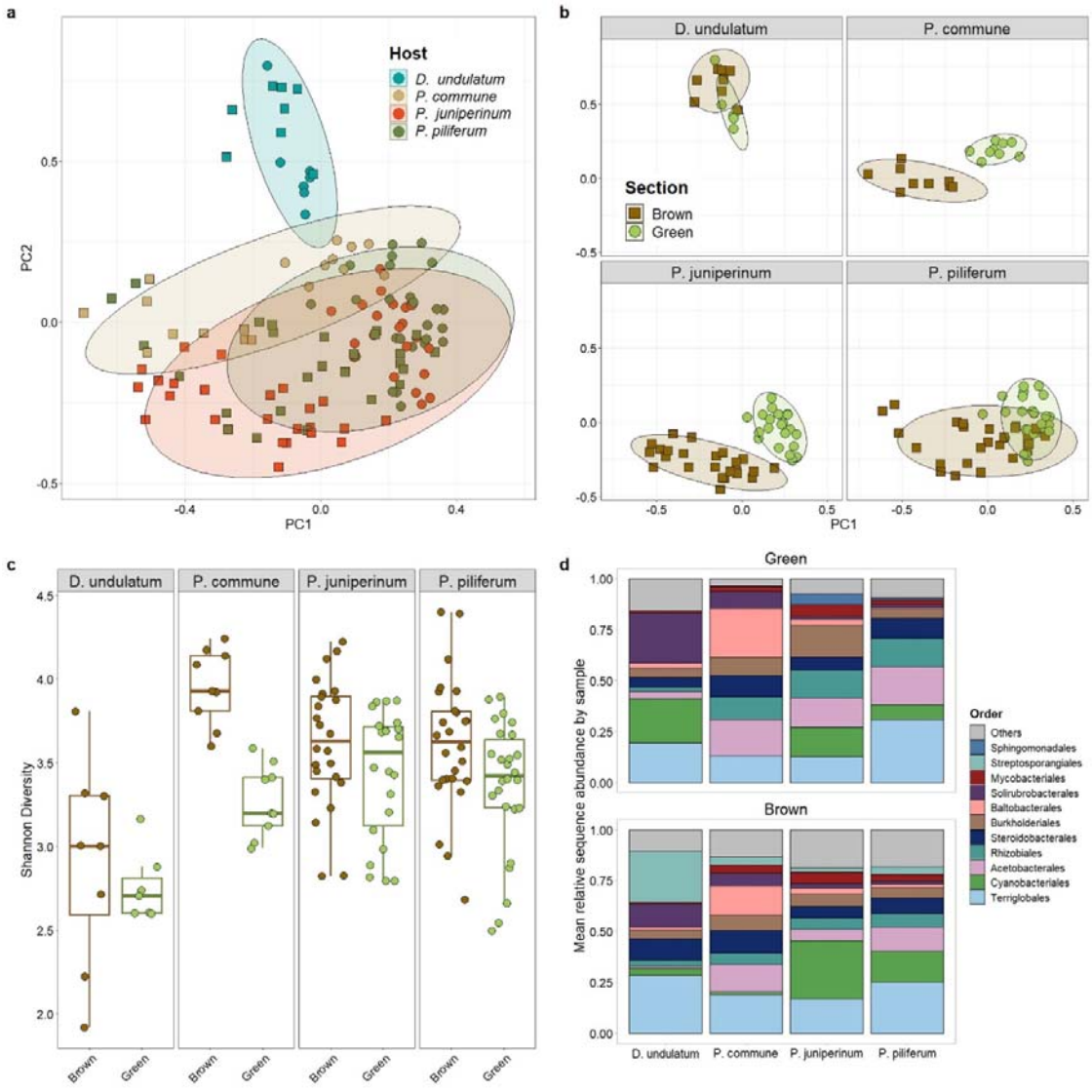


Figure 2. (a, b) Principal coordinate analysis (PCoA) of bacterial communities based on Bray-Curtis dissimilarity index, calculated using vst-transformed abundance table. Points represent moss samples, and ellipses indicate 95% confidence interval around moss samples from the same (a) moss species or (b) gametophyte section. (c) Differences in alpha diversity (Shannon index) of moss bacterial communities faceted by moss species: *P. juniperinum* (n_B = 24, n_G = 22), *P. commune* (n_B = 9, n_G = 9), *P. piliferum* (n_B = 26, n_G = 26), and *D. undulatum* (n_B = 8, n_G = 7). (d) Bacterial taxonomic profiles between green and brown gametophyte sections and among moss species. Bar charts are faceted by moss species, and bar colours represent different microbial taxa.

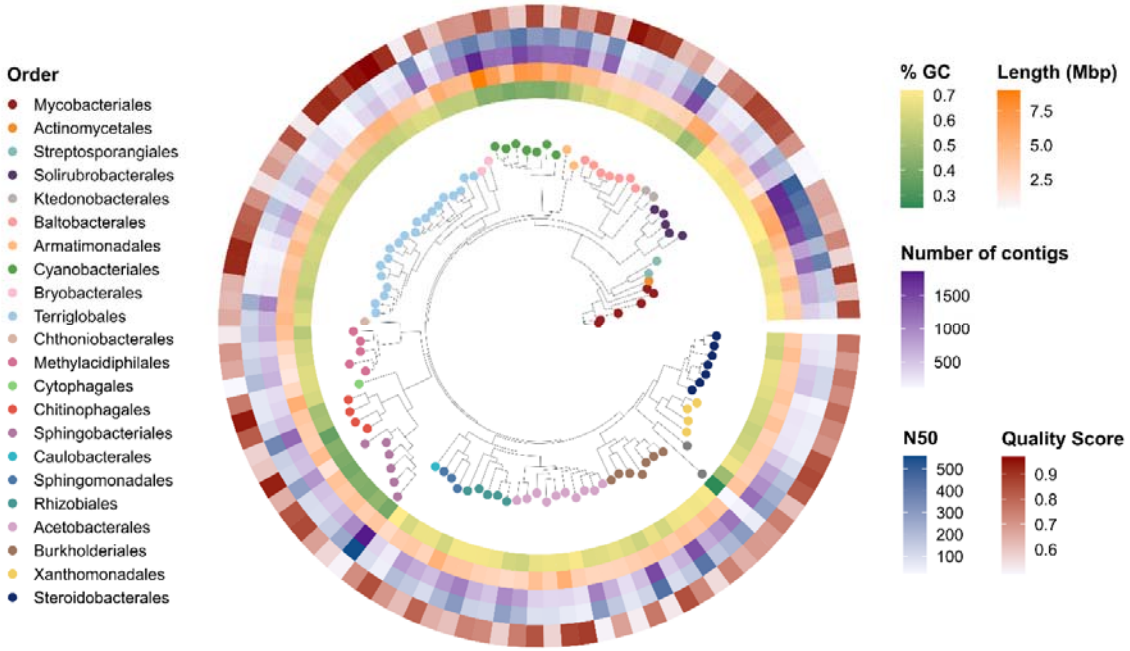
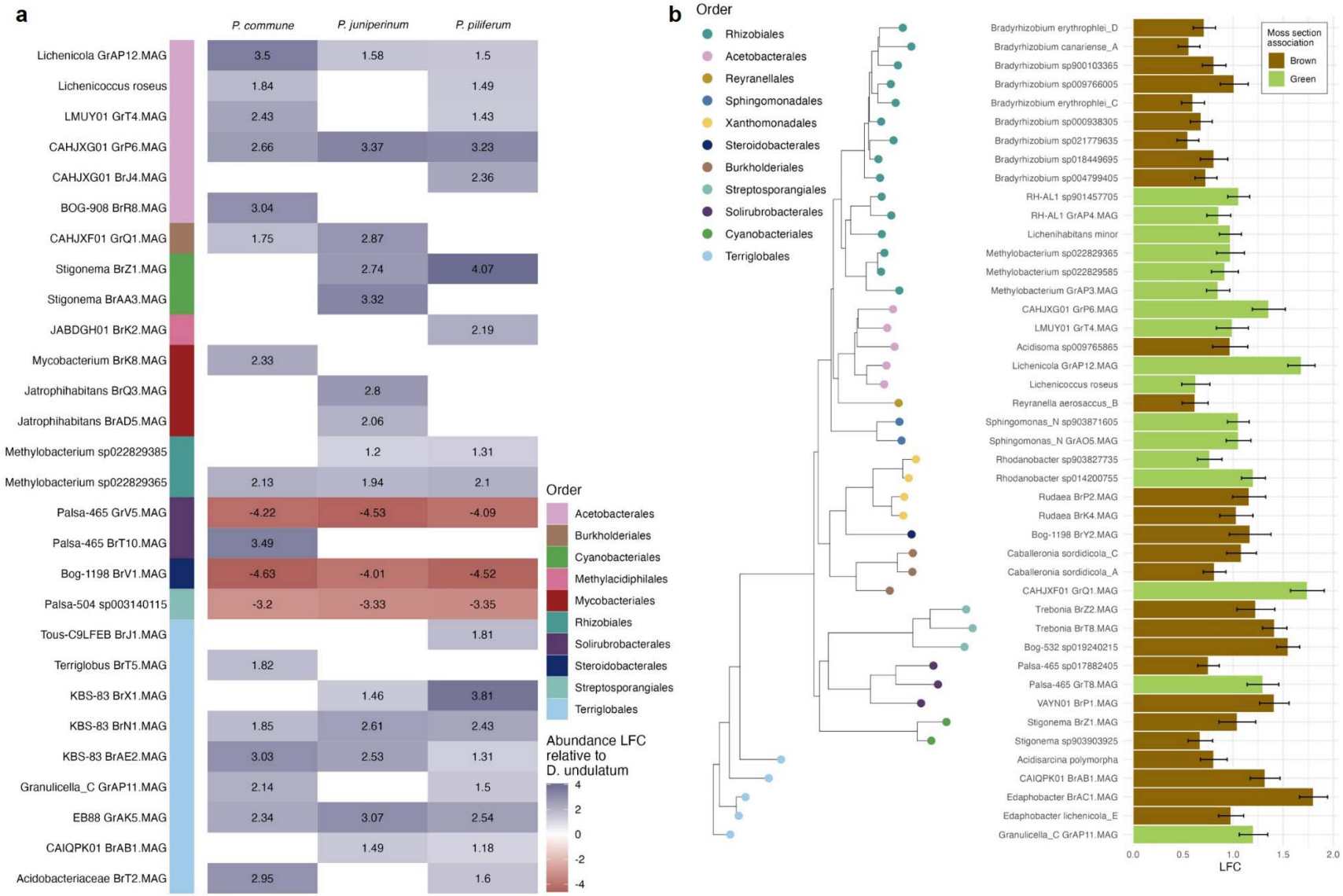
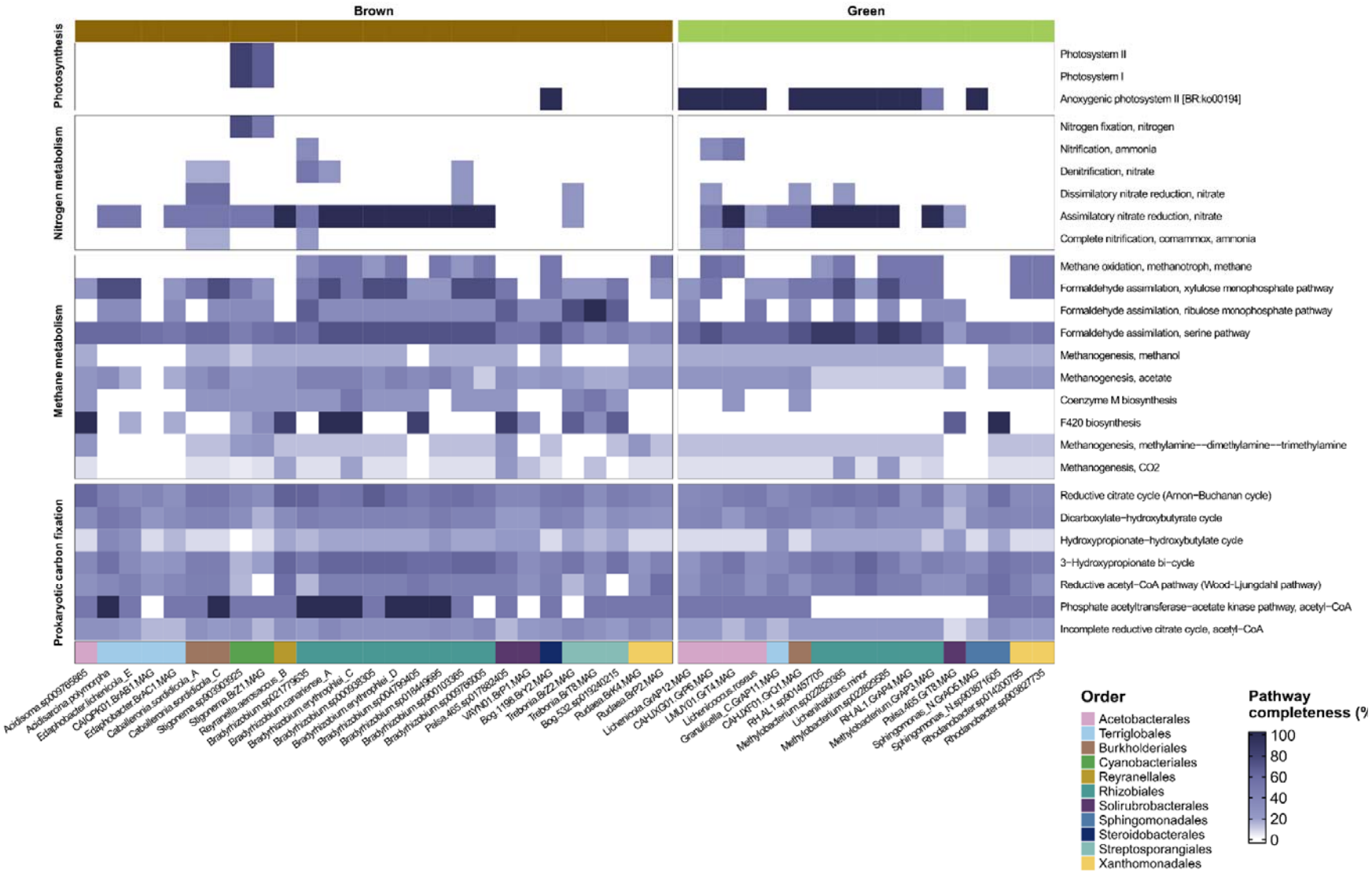


Figure 3. Novel species-level metagenome-assembled genomes obtained with co-assembly. From inner to outer layers: proportion of guanine and cytidine in assembly nucleotides (%□GC), genome size (length, Mbp), number of contigs in each MAG, length of the shortest of the largest contigs required to cover 50% of the total assembly length (N50) and quality score (QC). The inner section shows the phylogenetic tree of 110 metagenome-assembled genomes constructed using *PhyloPhlAn* and built phylogenetic trees using *ggtree* packages, allowing a cross-validation with their assigned taxonomy in the GTDB database. The different colored dots on the branches of the tree correspond to the colors of the bacterial orders listed in the legend.


a
P. commune
P. juniperinum
P. piliferum
Lichenicola GrAP12.MAG 3.5 1.58 1.5
Lichenicoccus roseus 1.84 1.49
LMUY01 GrT4.MAG 2.43 1.43
CAHJXG01 GrP6.MAG 2.66 3.37 3.23
CAHJXG01 BrJ4.MAG 2.36
BOG-908 BrR8.MAG 3.04
CAHJXF01 GrQ1.MAG 1.75 2.87
Stigonema BrZ1.MAG 2.74 4.07
Stigonema BrAA3.MAG 3.32
JABDGH01 BrK2.MAG 2.19
Mycobacterium BrK8.MAG 2.33
Jatrophihabitans BrQ3.MAG 2.8
Jatrophihabitans BrAD5.MAG 2.06
Methylobacterium sp022829385 1.2 1.31
Methylobacterium sp022829365 2.13 1.94 2.1
Palsa-465 GrV5.MAG -4.22 -4.53 -4.09
Palsa-465 BrT10.MAG 3.49
Bog-1198 BrV1.MAG -4.63 -4.01 -4.52
Palsa-504 sp003140115 -3.2 -3.33 -3.35
Tous-C9LFEB BrJ1.MAG 1.81
Terriglobus BrT5.MAG 1.82
KBS-83 BrX1.MAG 1.46 3.81
KBS-83 BrN1.MAG 1.85 2.61 2.43
KBS-83 BrAE2.MAG 3.03 2.53 1.31
Granulicella_C GrAP11.MAG 2.14 1.5
EB88 GrAK5.MAG 2.34 3.07 2.54
CAIQPK01 BrAB1.MAG 1.49 1.18
Acidobacteriaceae BrT2.MAG 2.95 1.6
Order
Acetobacterales
Burkholderiales
Cyanobacteriales
Methylacidiphilales
Mycobacteriales
Rhizobiales
Solirubrobacterales
Steroidobacterales
Streptosporangiales
Terriglobales
Abundance LFC relative to D. undulatum
4
2
0
-2
-4
b
Order
Rhizobiales
Acetobacterales
Reyranellales
Sphingomonadales
Xanthomonadales
Steroidobacterales
Burkholderiales
Streptosporangiales
Solirubrobacterales
Cyanobacteriales
Terriglobales
Moss section association
Brown
Green
Bradyrhizobium erythrophlei_D
Bradyrhizobium canariense_A
Bradyrhizobium sp900103365
Bradyrhizobium sp009766005
Bradyrhizobium erythrophlei_C
Bradyrhizobium sp000938305
Bradyrhizobium sp021779635
Bradyrhizobium sp018449695
Bradyrhizobium sp004799405
RH-AL1 sp901457705
RH-AL1 GrAP4.MAG
Lichenihabitans minor
Methylobacterium sp022829365
Methylobacterium sp022829585
Methylobacterium GrAP3.MAG
CAHJXG01 GrP6.MAG
LMUY01 GrT4.MAG
Acidisoma sp009765865
Lichenicola GrAP12.MAG
Lichenicoccus roseus
Reyranella aerosaccus_B
Sphingomonas_N sp903871605
Sphingomonas_N GrAO5.MAG
Rhodanobacter sp903827735
Rhodanobacter sp014200755
Rudaea BrP2.MAG
Rudaea BrK4.MAG
Bog-1198 BrY2.MAG
Caballeronia sordidicola_C
Caballeronia sordidicola_A
CAHJXF01 GrQ1.MAG
Trebonia BrZ2.MAG
Trebonia BrT8.MAG
Bog-532 sp019240215
Palsa-465 sp017882405
Palsa-465 GrT8.MAG
VAYN01 BrP1.MAG
Stigonema BrZ1.MAG
Stigonema sp903903925
Acidisarcina polymorpha
CAIQPK01 BrAB1.MAG
Edaphobacter BrAC1.MAG
Edaphobacter lichenicola_E
Granulicella_C GrAP11.MAG
0.0
0.5
1.0
1.5
2.0
LFC

1091

Figure 4. Differential abundance of bacterial species between moss species and between green and brown gametophyte sections. LFC: $\log_2$-fold change in absolute abundances**,** with all values showing significant differential abundance ($p < 0.01$). (a) Differentially abundant bacterial species in moss species relative to *D. undulatum*. LFC are relative to *D. undulatum*. Positive LFC (blue) indicates enrichment in specified moss species relative to *D. undulatum*, whereas negative LFC (red) indicates enrichment in *D. undulatum* relative to specified moss species**.** (b) Bacterial species that are differentially abundant in either green or brown sections of moss gametophyte. Bar colours indicate in which section (green or brown) the species was differentially enriched, with LFC calculated in relation to the reference group. Blank cells correspond to species for which no significant differences in abundance were detected, and therefore no LFC values reported.


Brown
Green
Photosynthesis
Nitrogen metabolism
Methane metabolism
Prokaryotic carbon fixation
Photosystem II
Photosystem I
Anoxygenic photosystem II [BR:ko00194]
Nitrogen fixation, nitrogen
Nitrification, ammonia
Denitrification, nitrate
Dissimilatory nitrate reduction, nitrate
Assimilatory nitrate reduction, nitrate
Complete nitrification, comammox, ammonia
Methane oxidation, methanotroph, methane
Formaldehyde assimilation, xylulose monophosphate pathway
Formaldehyde assimilation, ribulose monophosphate pathway
Formaldehyde assimilation, serine pathway
Methanogenesis, methanol
Methanogenesis, acetate
Coenzyme M biosynthesis
F420 biosynthesis
Methanogenesis, methylamine→dimethylamine→trimethylamine
Methanogenesis, CO2
Reductive citrate cycle (Arnon−Buchanan cycle)
Dicarboxylate−hydroxybutyrate cycle
Hydroxypropionate−hydroxybutylate cycle
3−Hydroxypropionate bi−cycle
Reductive acetyl−CoA pathway (Wood−Ljungdahl pathway)
Phosphate acetyltransferase−acetate kinase pathway, acetyl−CoA
Incomplete reductive citrate cycle, acetyl−CoA
Acidisoma.sp009785865
Acidisarcina.polymorpha
Edaphobacter.lichenicola_E
CAIQPK01.BrAB1.MAG
Edaphobacter.BrAC1.MAG
Caballeronia.sordidicola_A
Caballeronia.sordidicola_C
Stigonema.sp003003925
Stigmonema.BrZ1.MAG
Reyranella.aerosacus_B
Bradyrhizobium.sp021779535
Bradyrhizobium.canariense_A
Bradyrhizobium.erythrophlei_C
Bradyrhizobium.sp000338905
Bradyrhizobium.erythrophlei_D
Bradyrhizobium.sp004799405
Bradyrhizobium.sp018449595
Bradyrhizobium.sp000103365
Bradyrhizobium.sp009769005
Palsa.465.sp017682405
VAYN01.BrP1.MAG
Bog.1198.BrY2.MAG
Trebonia.BrZ2.MAG
Trebonia.BrTB.MAG
Bog.532.sp019248215
Rudaea.BrK4.MAG
Rudaea.BrP2.MAG
Lichenicola.GrAP12.MAG
CAHJXG01.GrP6.MAG
LMUY01.GrT4.MAG
Lichenicoccus.roseus
Granulicella_C.GrAP11.MAG
CAHJXF01.GrG1.MAG
RHAL1.sp001457705
Methylobacterium.sp022823385
Licheniphilatans.minor
Methylobacterium.sp022823535
RHAL1.GrAP4.MAG
Methylobacterium.GrAP3.MAG
Palsa.465.GrTB.MAG
Sphingomonas_N.GrAC5.MAG
Sphingomonas_N.sp003871605
Rhodanobacter.sp014200755
Rhodanobacter.sp003827735
Order
Acetobacterales
Terriglobales
Burkholderiales
Cyanobacteriales
Reyranellales
Rhizobiales
Solirubrobacterales
Sphingomonadales
Steroidobacterales
Streptosporangiales
Xanthomonadales
Pathway completeness (%
100
80
60
40
20
0

1099

Figure 5. Metabolic abilities of the differentially abundant species in brown and green gametophyte sections. The pathway completeness of selected metabolic pathway groups, such as photosynthesis, nitrogen and methane metabolism, and prokaryotic carbon fixation, is indicated for each genome (blue scale). This figure also shows the taxonomic rank (e.g., bacterial order) for each of the 44 differentially abundant genomes.