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 contribute to ecosystem processes in boreal forests. Yet, how host-specific characteristics and the environment drive the composition and metabolic potential of moss microbiomes is still poorly understood. In this study, we use shotgun metagenomics to identify the taxonomy and metabolic potential of the bacteria of four moss species of the boreal forests of Northern Québec, Canada. To 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 to a lesser extent soil pH and soil temperature, drive moss-associated bacterial community composition and diversity. In the brown gametophyte section, two *Stigonema* spp. showed partial pathway completeness for photosynthesis and nitrogen fixation, while all brown-associated Hyphomicrobiales had complete assimilatory nitrate reduction pathways and many nearly complete carbon fixation pathways. Several brown-associated species showed partial to complete pathways for coenzyme M and F420 biosynthesis, important for methane metabolism. In addition, green-associated Hyphomicrobiales (*Methylobacteria* spp.) displayed potential for the anoxygenic photosystem II pathway. Overall, our findings demonstrate how host-specific characteristics and environmental factors shape the composition and metabolic potential of moss bacteria, highlighting their roles in carbon fixation, nitrogen cycling, and methane metabolism in boreal forests.

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 sequestration 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 soil temperature have been 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, closely related moss species are likely to

harbor more similar microbial taxa than more distantly related ones [18, 21, 22]. 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 species 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 forest [27]. Importantly, these taxa are key candidates for remediation and restoration efforts as they often thrive in degraded landscapes [28]. Thus, there is an important knowledge gap on moss-microbe interactions beyond common moss genera. Considering the urgency to develop conservation strategies for bryophytes, it is crucial to deepen our understanding of moss-microbe interactions and their functional roles in unaltered boreal ecosystems by investigating understudied moss species, their phylogenetic relationships, and local environmental factors.

Beyond moss species identity, physiological differences within a 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, separated along their vertical gradient in two distinct sections: a living “green” section and an underlying, decaying “brown” section [29]. These sections can differ in microbial diversity, as well as in the composition and functional potential of their microbial communities. [30, 31]. The green section can harbor photosynthetically active microbes [32, 33] and is associated with N-fixing cyanobacteria [34]; it is often the focus of studies on moss-bacteria associations. However, 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 [33, 35]. Notably, green and brown sections of the moss gametophyte are exposed to distinct microenvironmental conditions [36], whereby 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 [39]. WMS is therefore particularly useful in characterizing understudied microbiomes, such as that of boreal mosses. The discovery of novel bacterial species, along with an improved understanding of their metabolic potential in different gametophyte sections and environmental conditions, is important for predicting how these interactions will shift as climate change and human activities alter boreal ecosystems.

Here, we measure the influence of moss identity, gametophyte section, and abiotic environment on moss bacterial communities and their 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) characterize the diversity, community composition, and metabolic potential of moss-associated bacteria; and (2) quantify the relative importance of moss species, gametophyte section, soil pH, and soil 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 soil temperature

First, we investigated the influence of moss genus, moss species (nested within moss genus), gametophyte section, soil pH, and soil temperature on bacterial community composition and diversity (Table 1a). We then narrowed our focus on the three *Polytrichum* spp. (excluding *Dicranum undulatum*; 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). Notably, when restricting the analysis to the three *Polytrichum* spp., the effect of moss species on bacterial community composition remained significant (R^2 = 19.7%, p-value = 0.001; Table 1b). Thus, our results highlight that differences in bacterial community composition are driven by both moss genus and, within these genera, by moss species. These results are in line with previous studies in which moss phylogeny was shown to be an important driver of bacterial community composition [4, 8, 25]. This finding illustrates a potential co-evolutionary relationship between mosses and some of the bacteria members within the community [17, 40]. 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 *D. undulatum* samples were 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 [41] and micro-environmental conditions (e.g., soil pH, moisture, temperature) [36]. In addition, we observed a significant, albeit small, interaction between gametophyte

section 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 *D. undulatum* 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, the interaction of soil pH and temperature affected 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 (bacterial Shannon diversity) being higher for *Polytrichum* than for *D. undulatum* (χ^2 = 8.6, p-value = 0.003; 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 the green section was consistently lower than in the 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 the green section or by (ii) the proximity of the brown section to the soil, which is known to be extremely rich in microorganisms. As such, the 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 and may contain less chlorophyll than the above active, green section [42]. Finally, the underlying

brown section tends to have higher moisture contents due to the boundary layer effect (i.e., the physical phenomenon in which wind velocity is reduced); an effect that minimizes moisture loss [43]. Among abiotic factors, soil temperature was the only factor showing a significant association with moss bacterial alpha diversity, evident in both models, with and without *D. undulatum* ($\chi^2 = 6.3$, 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, these results show that moss genus, gametophyte section, soil pH, and soil temperature are important drivers of moss bacterial community composition and diversity, but that moss species only affects the former. Our work further demonstrates that green and brown sections within a moss gametophyte differ in bacterial community composition (Fig. 2d) and alpha diversity, emphasizing the relationship between moss characteristics and the associated bacterial communities.

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

By assembling metagenomes, we discovered 110 potentially novel bacterial species whose genomes were 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 medium to high quality (82.46 ± 11.68 % completeness, 1.97 ± 1.26 % contamination). 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 bacterial species spanning 23 orders across nine phyla. 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 reference genome at the species level. These nMAGs 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, including the number of contigs and proportion of GC content, are reported in Supplementary Data 1 and summarized in Fig. 3.

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 Fig. S1). This increase 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. As detailed in Methods, all results presented here (including the above section on diversity) rely on community composition estimation performed using the improved reference database.

Of note, we modified the GTDB taxonomy used to annotate our genomes according to the latest nomenclature in the *List of Prokaryotic names with Standing in Nomenclature* (LPSN) [44]. Specifically, GTDB classifies Hyphomicrobiales (LPSN nomenclature) as Rhizobiales, so we corrected our annotations to use Hyphomicrobiales throughout. GTDB also uses the order Cyanobacteriales, which encompasses species of many different orders (e.g., Oscillatoriales, Synechococcales, Chroococcales, etc.) according to LPSN. Since every member of the class Cyanobacteria found in our samples belonged to the Nostocaceae family, we renamed the Cyanobacteriales order to Nostocales, following LPSN.

We recovered seven MAGs belonging to the order Baltobacterales [45] from the phylum Eremiobacterota [46] (Fig. 3). Despite the presence of 71 Baltobacterales species-representative genomes in the GTDB R214 database (28th April 2023), all seven recovered genomes were identified as nMAGs. Many Eremiobacterota species were recently described by metagenome assembly from a number of boreal ecosystems such as Alaskan mosses [18], Canadian acidic springs [47], and Swedish peatlands [48]. Baltobacterales were highly prevalent in our samples, detected in 109/131 (~83%) samples, and two *Nyctobacter* MAGs were of high quality (QS = 0.95 and 0.91 respectively; Supplementary Data 1). While there is little information on the ecological traits of Eremiobacterota, several genera within this group were shown to have the capacity for C fixation [39]. Our study supports this finding, as each Eremiobacterota we detected showed partial completeness in most prokaryotic carbon fixation pathway considered (Supplementary Fig. S2). Pessi *et al.* [50] detected *nif* homologues in the Eremiobacterota species Candidatus *Lamibacter sapmiensis*. However, we only found partial completeness of an assimilatory nitrate reduction module (M00531) in Candidatus *Velthaea versatilis*, whose nitrogen reduction capacity has already been well described [49]. Interestingly, all Eremiobacterota nMAGs except *Tumulicola* BrAA6 can encode either partial or complete pathways for anoxygenic photosynthesis, supporting findings by Ward *et al.* [45]. Additionally, we detected partial completeness of the methane oxidation pathway across all detected species within this phylum, supporting the findings of Holland-Moritz *et al.* [51], who identified only two cases of the methane oxidation marker gene *pmoA*. Moreover, Ray *et al*. [52] suggested Eremiobacterota might use atmospheric chemosynthesis by assimilating formaldehyde through the serine pathway. Our results support this hypothesis, as all Eremiobacterota we detected exhibited partial completeness of this pathway (see Supplementary Fig. S2). Our study thus contributes novel Eremiobacterota

genomes of potentially undiscovered species and information about their metabolic potential, and further supports existing knowledge about moss-associated members of this clade.

Bacterial species associated with moss species

In order to identify differentially abundant taxa across groups of samples, we used ANCOM-BC2 [53], 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* as a reference group. *D. undulatum* was selected because of its contrasting bacterial community composition, which served as a baseline for evaluating differential abundant bacteria among 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 Fig. S3). Of these 28 bacterial species, 24 were nMAGs, underlining the contribution of the assembly process to the depth of our analyses.

We also performed a differential abundance test at the order level in a pairwise manner (i.e., without a reference host). Interestingly, we found that the order Baltobacterales was more abundant in *P. commune* by 2.7 to 3.3 $\log_2$-folds compared to all other moss species (Supplementary Fig. S3). Previous metagenomic studies showed that Baltobacterales encode genes necessary for hopanoid synthesis (a biochemical process in which hopanoid, steroid-like compounds, are produced) [49]. Hopanoid molecules were shown to contribute to bacterial tolerance to low pH soils [54], temperature stress [49], microbial defense [56], and are linked to N-fixation [57, 58]. Notably, *P. commune* is known for its antibacterial properties [60] 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 could provide an adaptive advantage for colonization and persistence of bacteria in acidic and anti-microbial habitats as found in *P. commune* mosses.

Two *Stigonema* spp. (Nostocales) were differentially abundant in *P. juniperinum* and *P. piliferum* (relative to *D. undulatum*), but not in *P. commune* (Fig. 4a). One major difference between *Polytrichum* species is the habitat in which *P. commune* was growing in. We found that *P. commune* grew in a significantly more shaded environment than the other *Polytrichum* species (Supplementary Data S2), and *Stigonema* may be more sensitive to these differences in shading. Indeed, Jean *et al.* (2020) found that Nostocaceae abundance was negatively associated with canopy cover [8]. Furthermore, the aforementioned anti-microbial properties of *P. commune* could select against *Stigonema* spp. growth. Finally, *Stigonema*, known for having an alternative nitrogenase, is also common in boreal lichens (*Stereocaulon*) growing near *Polytrichum* and *D. undulatum*. [57]. The proximity of *Stigonema* to these moss species, combined with its unique functional adaptation, such as its alternative nitrogenase, may influence the abundance or composition of the bacterial communities associated with these mosses. In general, all Acetobacterales and Terriglobales species were positively associated with *Polytrichum* species (Fig. 4a). Of note, Terriglobales, one of the most overall relatively abundant species in our dataset (mean 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 [62].

Three species from the orders Solirubrobacterales, Steroidobacterales, and Streptosporangiales were found to be much more abundant in *D. undulatum* compared to *Polytrichum* hosts (3.20 to 4.63 $\log_2$-folds, Fig. 4a). The most closely related species from these orders were incidentally all species identified through metagenome assembly and were described in soil and roots in non-boreal ecosystems [52, 63]. A potential reason for the differences in bacterial communities between *Dicranum* and *Polytrichum* genera could be due to host environmental preferences (*D. undulatum* preferring moist soils), as well as moss N and phosphorus (P) content [64]. Although our results show that moss-associated bacterial

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 species associated with brown and green gametophyte sections

We identified 44 species as differentially abundant between green and brown gametophyte sections, 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).

Bacterial species with the highest $\log_2$-fold change in brown gametophyte sections included three nMAGs from the genus *Edaphobacter*, uncultured species from the order Solirubrobacterales (*VAYN01*) and Terriglobales (*CAIQPK01*), as well as three uncultured species from the order Streptosporangiales (*Bog-532 sp019240215* and two *Trebonia* nMAGs; 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 [65, 66]. Although Solirubrobacterales are poorly studied, researchers have detected members of this order in diverse ecosystems including agricultural lands [67] and Antarctic soils [68]. Several other brown-associated taxa, such as *Rudaea* spp., *CAIMXF01*, *Trebonia*, and *Acidisarcina*, are also mainly associated with soils [69–71]. Finally, two *Stigonema* species (cyanobacteria of the order Nostocales) were differentially abundant in brown gametophytes (Fig. 4b). In line with our findings, Renaudin *et al*. [10] found that cyanobacteria biomass increased towards the bottom of moss gametophyte, peaking at approximately 2-3 cm from the apex. In their study on the bacterial communities of two *Racomitrium* moss species, Klarenberg *et al.* [72] demonstrated that there is little

transmission of bacteria from soil to moss gametophytes [69]. Meanwhile, Escolástico-Ortiz *et al*. [73] showed overlap between soil bacterial communities and the moss *Racomitrium lanuginosum*. However, the authors also revealed the presence of a core microbiome for moss gametophytes and soil, with each harbouring distinct, non-overlapping core bacterial communities. While we did not characterize the bacterial community of the surrounding soil, our finding of a number of soil species in the brown gametophyte section suggests this habitat may be colonized 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 change in absolute abundance ($\log_2$-fold > 1.29; Fig. 4b) compared to the brown gametophyte section. Hay *et al*. [75] 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 certain species of Hyphomicrobiales (Rhizobiales) to be green-associated, specifically from the genera *Methylobacteria*, *Lichenihabitans*, and *RH-AL1*, as well as all species of the genus *Bradyrhizobium* (Fig 4b). *Methylobacteria*, for example, were found in the core microbiome of boreal shrubs [76] and were shown to colonize non-boreal mosses (e.g., *Didymodon tectorum* and *Funaria hygrometrica* [77, 78]). Similarly, *Lichenihabitans minor* and *RH-AL1* spp. were shown to harbor genes associated with methanol consumption [74]. Although *Lichenihabitans minor* tends to be associated with lichens, the detection of this species in our samples is not surprising, given the prevalence of lichens in the habitat of boreal mosses [79]. *Lichenicola* and *Lichenicoccus* species, eponymously, can associate with lichens. For example, *Lichenicola cladoniae* was previously found in Antarctic lichens [80]. Overall, through our analysis, we observed the association of several species from genera such as *CAHJXF01*, *CAHJXG01*, *Methylobacteria*, *Lichenihabitans*, and *RH-AL1* spp. with the 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 gametophyte section-associated bacteria by analyzing which pathways their genomes have the potential to express. We focused on four pathway groups relevant to 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 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 at least partially encoding these pathways. Moreover, two *Stigonema* spp. stood out because of their partially complete N fixation (nifHDK + vnfDKGH) pathway (*Stigonema sp903903925* = 75% and *Stigonema* BrZ1.MAG = 50% completeness) (Fig. 5; Supplementary Data S3). This supports a previous finding in feathermosses, which showed *Stigonema* spp. to be important contributors to N fixation [81, 82]. *Stigonema* spp. also showed partial pathway completeness in all methane modules relevant to methanogenesis, and not methane oxidation. Although methanogenesis has not been shown to occur in any organisms except archaea [83], partial completeness of several methanogenesis pathways alludes to the potential for certain cyanobacteria to contribute to non-methanogenesis methane production (Fig. 5; Supplementary Data S3). Overall, our findings suggest that Nostocales species associated with the brown gametophyte section are photosynthetically active and have the potential for nitrogen fixation [85, 86].

Second, we detected complete assimilatory nitrate reduction pathways in both green- and brown-associated Hyphomicrobiales (Rhizobiales), in line with earlier findings on nitrogen metabolic capacities of this clade (Fig. 5; Supplementary Data S3). Hyphomicrobiales taxa were previously shown to be N-fixers for nodule-producing leguminous plants [87] and Sphagnum mosses [17]. Interestingly, we found that nitrogen metabolism pathways were not significantly more complete in bacteria associated with either green or brown gametophyte sections. This suggests that the potential of the bacterial community to encode these pathways is not influenced by section type (Supplementary Table S1). Furthermore, most methane metabolic pathways were also partially complete in both green and brown-associated Hyphomicrobiales genomes (Fig. 5; Supplementary Data S3). For example, the methanogenesis acetate pathway was significantly more complete in brown-associated species of *Bradyrhizobium* (p-value = 0.021; Fig. 5; Supplementary Table S1). This aligns with the results of a previous study on *Sphagnum* mosses, which found that methanotrophy is primarily carried out by Hyphomicrobiales species [88]. Additionally, two of the three green-associated Methylobacteria showed potential to fully encode the anoxygenic photosystem II pathway, whereas the third one (one of the nMAGs) can only encodes 50% of the required genes (Fig. 5; Supplementary Data S3). Of note, the only photosynthesis pathway whose completeness was significantly associated with the green-associated species was also the anoxygenic photosystem II (p-value = 0.008 Supplementary Table S1). This is in line with recent findings that aerobic anoxygenic photosynthetic bacteria are present among plant phyllosphere microbial communities [89], including certain species of Methylobacteria [90]. Next, not a single green-associated Hyphomicrobiales encodes any gene of the phosphate acetyltransferase-acetate kinase pathway (carbon fixation), whereas among brown-associated Hyphomicrobiales, six *Bradyrhizobium* spp. encode 100% of the required genes (Fig. 5; Supplementary Data S3). In line with this finding, the phosphate

acetyltransferase-acetate kinase pathway was also the only pathway significantly more complete in differentially brown-associated species (p-value = 0.012; Supplementary Table S1). This pathway is part of a central metabolism of bacteria [91]; its high completeness in *Bradyrhizobia* indicates the metabolic potential for C fixation [92]. To sum up, our findings suggest that green- and brown-associated Hyphomicrobiales 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 for denitrification activity, 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 we did not find methane oxidation pathways in these species, they still showed completeness in several methane metabolism pathways, such as formaldehyde assimilation (Fig. 5; Supplementary Data S3). These findings suggest the distinct metabolic capabilities of Burkholderiales in different sections of the gametophyte [93, 94]. Curiously, pathways for coenzyme M (p-value = 0.015) and F420 biosynthesis (p-value = 0.169) were partially to fully complete in several brown-associated species (Fig. 5; Supplementary Table S1 and Supplementary Data S3). F420 is a costly co-enzyme to synthesize; nevertheless, Ney *et al*. [95] demonstrated 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 [96], which are present in boreal forest soils [97]. More interestingly, alkenes are known to be produced by bacteria associated with lichens [98]. In our sampling sites, lichens frequently grew alongside mosses, indicating that some of the bacterial taxa may be shared between the groups. Our results suggest that the presence of coenzyme M in

bacteria associated with brown gametophyte sections may be attributed to their utilization of alkenes commonly found in lichens.

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 demonstrated that moss genus, moss species, gametophyte section, and, to a lesser extent, soil pH and soil temperature are significant drivers of bacterial community composition and diversity (Tables 1-2; Fig. 2). Through metagenomic assembly, we identified 110 potentially novel bacterial species (Fig. 3) and characterized distinct bacterial communities associated with four different moss species (Fig. 4). We found that bacterial communities in green and brown gametophyte sections occupy unique functional niches (Fig. 5), with partial partial overlap. Our results show that the brown gametophyte section, despite their senescent nature, can support bacterial communities that contribute to ecological functions such as nitrogen fixation, carbon fixation, and methane metabolism, while green-associated bacteria exhibit potential for anoxygenic photosynthesis.

To build on these findings, several key areas for future research emerge. First, validating the specific metabolic pathways of the 110 potentially novel bacterial species and their contributions to processes like nitrogen fixation and carbon sequestration is important. For example, experimental studies and transcriptomic studies could help to identify which genes are actively transcribed under different environmental conditions. Second, exploring the interactions between moss-associated bacteria, fungi, and other microbial members could provide more insights into their joint roles in ecosystem functioning. Gaining a clear

understanding of moss metabolites could further elucidate the potential co-evolutionary or symbiotic relationships between non-cyanobacterial microbial communities and mosses, revealing how metabolites influence microbial interactions and ecosystem processes. To summarize, our results contribute to our understanding of how moss-bacteria interactions contribute to ecosystem functions, such as nutrient cycling and carbon sequestration, emphasizing the importance of both green and brown gametophyte section-associated bacteria in these processes.

Methods

Study area

This study was conducted within the Eeyou Istchee (James Bay) region of Northern Québec (Fig. 1a; N51.7040721°, W76.0521982°). The sampling area is part of Québec's boreal forest biome and is dominated by closed-crown forest stands of Jack pine (*Pinus banksiana*) and black spruce (*Picea mariana*) [99]. The monthly average annual temperature of this region is 0.3□, while mean annual precipitation is 741 mm [100]. The soils are predominantly acidic and nutrient poor [101].

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 native 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 sampled directly (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 [102]. 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* ($N_{brown} = 24$, $N_{green} = 22$), *P. commune* ($N_{brown} = 9$, $N_{green} = 9$), *P. piliferum* ($N_{brown} = 26$, $N_{green} = 26$), and *D. undulatum* ($N_{brown} = 8$, $N_{green} = 7$) from five sites (see Supplementary Table S2 for sample breakdown).

<u>DNA extraction and whole metagenome sequencing (WMS)</u>

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 2.3 mM BioSpec 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).

Bioinformatic analyses

Samples preprocessing (quality control), assembly, MAG generation and annotation, as well as novelty analyses, were performed on high-performance computer with the aid of SLURM workload management for parallel processing. The pipeline for these steps is written in bash. In the next steps, the R programming language was used for diversity and differential

abundance analyses, as well as production of figures and tables. R packages are listed in Supplementary Table S4, along with all bioinformatic software (version and non-default options, when applicable) outlined in the following methods. The complete workflow is provided in Supplementary Fig. S4.

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 0.12.0, keeping reads with a minimum length of 50 bp and an average quality score of 30 over a 4-nucleotide sliding window. 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 *Physcomitrella patens*, as well as the human genome GRCh38 to remove human contamination, using Bowtie 2.4.4 (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 [103]. 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 [104]. In this study, we pooled samples (i.e., concatenated the *.fastq* metagenome files) taken from the same moss species, gametophyte section, and microsite 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 [105]. First, paired-end sequences were assembled using metaSPAdes 3.15.4 and re-assembled unassembled reads using MEGAHIT 1.2.9. Contigs were then binned with three binners (MetaBAT 2.12.1, CONCOCT 1.1.0, and MaxBin 2.2.4). Afterwards, these three sets of bins were hybridized with the MetaWRAP 1.3 *bin_refinement* module to produce a final set of refined bins for each co-assembly. Bin quality was then assessed using CheckM 1.0.18 and bins with completeness < 50% and contamination > 10% were excluded [106], 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 3.4.0 [107]. dRep uses a two-step clustering strategy, whereby bins are first clustered using Mash 2.3 [108]; then, within each cluster, bins are dereplicated at a 95% average nucleotide identity (ANI) using MUMmer 3.23 [109]. GUNC 1.0.5 [110] 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 [111]. 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 0.2.1 [112] 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 4.7.0 [113]. 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 2.6.3 [114] and passing these to the MicrobeAnnotator 2.0.5 [115] 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 3.0.3 [116] and phylogenetic trees built using ggtree 3.10.0 [117] and ggtreeExtra 1.12.0, allowing a cross-validation with their attributed GTDB taxonomy. nMAGs statistics (length, L50, number of contigs and GC contents) were assessed using bbmap 38.86.

Taxonomy characterization and abundance estimation of bacterial community

Sample community composition and relative abundance at the species level was estimated using Sourmash gather [113] with a *k*-mer size of 31, using the GTDB representative species genome index [111] 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*-mer-

based community profiling tools generate a *sequence* abundance table; these abundance values are not normalized for genome length (as opposed to marker-gene-based tools), making them distinct from taxonomic abundance [118]. 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 R214 reference index. The GenBank 2022-03 reference index (comprising 1.22 M microbial genomes) was also compared, and the containment increase is reported for both scenarios (Supplementary Fig. S1). Finally, the sequence abundance matrix was built into a phyloseq (version 1.46.0) object using the raw counts; this object was then used as a starting point for all subsequent analyses.

Diversity analyses

All statistical analyses and plots were made using the open-access software R 4.2.1 [119]. For community composition (multivariate) analyses, uneven sequencing depth and heteroskedasticity were addressed by applying variance stabilizing transformations (VST, DESeq2 1.42 [120]) to the abundance matrix. This matrix was then used to estimate pairwise Bray-Curtis dissimilarity between samples [121], and to compute principal coordinates using the *capscale* function in the vegan package (version 2.6-4) [122]. Finally, the *adonis2* function from the vegan package [122] was used to run PERMANOVAs, 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 [123].

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 [40, 124]. Linear models were run on Shannon index using function *lmer* in the lme4 1.1-35.1 [125] package, then the *anova* function in the car 3.1-2 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* from the DHARMa 0.4.6 package were used to assess model fit [126].

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 used.

Differential abundance analyses (DAA)

To identify species associated with moss species and gametophyte section, DAA were performed using ANCOMBC 2.4.0 package [53]. This approach was chosen because it takes into account the compositionality of next-generation sequencing data (i.e., sequencing data is commonly expressed in terms of relative abundances, meaning they sum up to an arbitrary constant and reside in a simplex mathematical space) [127], controls for false discovery rate by accounting for multiple testing and multiple group comparisons, and corrects for both 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-10) and moss species (SoilpH ~ Host, R^2 = 0.62, p-value < 1e-10), and location is significantly associated with moss species (χ^2 = 179.1, p-value < 1e-10). For both approaches, we only considered species present in at least 10% (n = 13) sample [128]. The Holm procedure was used to account for multiple comparisons. We only reported results that

passed ANCOMBC's sensitivity test to pseudo-count addition and had an adjusted p-value < 0.01.

Pathway completeness analysis

Pathway completeness is a value from 0 to 100% representing 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) [129]. Pathways for which completeness was ≤ 20% in more than 80% of the 44 DA species considered were excluded, yielding a subset of 23 pathways. Since pathway completeness correlates with genome completeness [130], 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 can include both boundaries (0% and 100%). P-values were corrected using the Holm method and only pathways with adjusted p-values < 0.05 were reported as significant. The heatmap in Fig. 5 was generated using the ComplexHeatmap 2.18.0package and shows a slightly different set of pathways compared to Supplementary Table S1, as certain pathways that were excluded for the statistical test on the basis of completeness-prevalence (as mentioned above) were still of interest for the discussion.

Environmental DNA contamination assessment

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) using BLAST+ 2.12.0. 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 Fig. S5-6). Therefore, the effective sequencing depth of bacterial species, even after adding our nMAGs, is lower than the sample sequencing depth. Of note, as the focus of this work being on bacteria, we designed the metagenome assembly pipeline to focus on metagenomic bins of bacterial origin, implicitly excluding eukaryotes that might have been assembled.

Furthermore, since no reference genome for the studied mosses was available at the time the assembly and analyses were performed, we assessed *a posteriori* the potential of host DNA to have contaminated our MAGs. To this end, reads from pooled sample were mapped to every MAG separately using Bowtie2 *(--very-sensitive*). Reads aligning with MAGs were then mapped to the recently published genome of *P. commune* (GCA_950295325.1). With one exception, the proportion of sample reads mapping to both, any MAG and *P. commune*'s genome, was extremely small, at 0.02 ± 0.07%. The exceptional MAG, *Green_N.bin.3*, had 3.7% of reads mapping to it also mapping to *P. commune*'s genome. This MAG was annotated as a *Buchnera* species (Supplementary Data 1); a phylogenetically close species, *Buchnera aphidicola*, has a particularly small genome (618 kbp), which could explain why the proportion of its genome conserved with a plant could be higher than for a bacteria with a more common genome size. Nevertheless, we tested our differential abundance analyses anew by excluding this MAG from the dataset, and the results presented in this work remained completely unaltered, as this MAG had not been found to be differentially abundant across gametophyte section or host mosses.

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 the beautiful depictions of mosses presented in Fig. 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 all bioinformatic and statistical analyses, along with abundance tables and metadata used for analysis, were made public in an online repository (https://github.com/jorondo1/borealMoss). Raw sequencing reads, assemblies, and MAGs were deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB76464.

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.

10. Renaudin M, Blasi C, Bradley RL, Bellenger J-P. New insights into the drivers of moss-associated nitrogen fixation and cyanobacterial biomass in the eastern Canadian boreal forest. *Journal of Ecology* 2022; **110**: 1403–1418.

11. Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Pean C, Chen Y, et al. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. 2021. Intergovernmental Panel on Climate Change, Switzerland.

12. Martin M, Grondin P, Lambert M-C, Bergeron Y, Morin H. Compared to Wildfire, Management Practices Reduced Old-Growth Forest Diversity and Functionality in Primary Boreal Landscapes of Eastern Canada. *Front For Glob Change* 2021; **4**.

13. Yin X, Martineau C, Fenton NJ. How big is the footprint? Quantifying offsite effects of mines on boreal plant communities. *Global Ecology and Conservation* 2023; **41**: e02372.

14. Jonsson M, Kardol P, Gundale MJ, Bansal S, Nilsson M-C, Metcalfe DB, et al. Direct and Indirect Drivers of Moss Community Structure, Function, and Associated Microfauna Across a Successional Gradient. *Ecosystems* 2015; **18**: 154–169.

15. Royles J, Griffiths H. Invited review: climate change impacts in polar regions: lessons from Antarctic moss bank archives. *Global Change Biology* 2015; **21**: 1041–1057.

16. Klarenberg IJ, Keuschnig C, Colmenares AJR, 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* 2021; **234**: 2044–2056.

17. Bragina A, Berg C, Cardinale M, Shcherbakov A, Chebotar V, Berg G. Sphagnum mosses harbour highly specific bacterial diversity during their whole lifecycle. *ISME J* 2012; **6**: 802–813.

18. Holland-Moritz H, Stuart JEM, Lewis LR, Miller SN, Mack MC, Ponciano JM, et al. The bacterial communities of Alaskan mosses and their contributions to N2-fixation. *Microbiome* 2021; **9**: 53.

19. McLaren JR, Buckeridge KM, van de Weg MJ, Shaver GR, Schimel JP, Gough L. Shrub encroachment in Arctic tundra: Betula nana effects on above- and belowground litter decomposition. *Ecology* 2017; **98**: 1361–1376.

20. Cao W, Xiong Y, Zhao D, Tan H, Qu J. Bryophytes and the symbiotic microorganisms, the pioneers of vegetation restoration in karst rocky desertification areas in southwestern China. *Appl Microbiol Biotechnol* 2020; **104**: 873–891.

21. Tian Y, Li YH. Comparative analysis of bacteria associated with different mosses by 16S rRNA and 16S rDNA sequencing. *J Basic Microbiol* 2017; **57**: 57–67.

22. Wang Y, Xue D, Chen X, Qiu Q, Chen H. Structure and Functions of Endophytic Bacterial Communities Associated with Sphagnum Mosses and Their Drivers in Two Different Nutrient Types of Peatlands. *Microb Ecol* 2024; **87**: 47.

23. Holland-Moritz H, Stuart JEM, Lewis LR, Miller SN, Mack MC, Ponciano JM, et al. The bacterial communities of Alaskan mosses and their contributions to N2-fixation. *Microbiome* 2021; **9**: 1–14.

24. Elzen E van den, Bengtsson F, Fritz C, Rydin H, Lamers LPM. Variation in symbiotic N2 fixation rates among Sphagnum mosses. *PLOS ONE* 2020; **15**: e0228383.

25. Kostka JE, Weston DJ, Glass JB, Lilleskov EA, Shaw AJ, Turetsky MR. The Sphagnum microbiome: new insights from an ancient plant lineage. *New Phytologist* 2016; **211**: 57–64.

26. Stuart JEM, Holland-Moritz H, Jean M, Miller SN, Ponciano JM, McDaniel SF, et al. The relationship of C and N stable isotopes to high-latitude moss-associated N2 fixation. *Oecologia* 2021; **197**: 283–295.

27. Macdonald SE, Fenniak TE. Understory plant communities of boreal mixedwood forests in western Canada: Natural patterns and response to variable-retention harvesting. *Forest Ecology and Management* 2007; **242**: 34–48.

28. Liu C, Pouliot K, Roy S, Rochefort L. Moss regeneration for lithium mine waste rock revegetation in Québec, Canada. *Botany* 2024; **102**: 26–40.

29. Lindo Z, Gonzalez A. The bryosphere: an integral and influential component of the Earth's biosphere. *Ecosystems* 2010; **13**: 612–627.

30. Arróniz-Crespo M, Bougoure J, Murphy DV, Cutler NA, Souza-Egipsy V, Chaput DL, et al. Revealing the transfer pathways of cyanobacterial-fixed N into the boreal forest through the feather-moss microbiome. *Front Plant Sci* 2022; **13**: 1036258.

31. Chen K-H, Liao H-L, Bellenger J-P, Lutzoni F. Differential gene expression associated with fungal trophic shifts along the senescence gradient of the moss Dicranum scoparium. *Environmental Microbiology* 2019; **21**: 2273–2289.

32. Solheim B, Zielke M. Associations Between Cyanobacteria and Mosses. In: Rai AN, Bergman B, Rasmussen U (eds). *Cyanobacteria in Symbiosis*. 2002. Springer Netherlands, Dordrecht, pp 137–152.

33. Liu X, Rousk K. The moss traits that rule cyanobacterial colonization. *Annals of Botany* 2022; **129**: 147–160.

34. Renaudin M, Blasi C, Bradley RL, Bellenger J-P. New insights into the drivers of moss-associated nitrogen fixation and cyanobacterial biomass in the eastern Canadian boreal forest. *Journal of Ecology* 2022; **110**: 1403–1418.

35. Rousk K. Biotic and abiotic controls of nitrogen fixation in cyanobacteria–moss associations. *New Phytologist* 2022; **235**: 1330–1335.

36. Mohanasundaram B, Pandey S. Effect of environmental signals on growth and development in mosses. *Journal of Experimental Botany* 2022; **73**: 4514–4527.

37. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 2017; **35**: 833–844.

38. Thomas AM, Segata N. Multiple levels of the unknown in microbiome research. *BMC Biology* 2019; **17**: 48.

39. Carter MM, Olm MR, Merrill BD, Dahan D, Tripathi S, Spencer SP, et al. Ultra-deep sequencing of Hadza hunter-gatherers recovers vanishing gut microbes. *Cell* 2023; **186**: 3111-3124.e13.

40. Wicaksono WA, Cernava T, Berg C, Berg G. Bog ecosystems as a playground for plant–microbe coevolution: bryophytes and vascular plants harbour functionally adapted bacteria. *Microbiome* 2021; **9**: 170.

41. Budke JM, Bernard EC, Gray DJ, Huttunen S, Piechulla B, Trigiano RN. Introduction to the Special Issue on Bryophytes. *Critical Reviews in Plant Sciences* 2018; **37**: 102–112.

42. Martin CE, Churchill SP. Chlorophyll concentrations and a/b ratios in mosses collected from exposed and shaded habitats in Kansas. *Journal of Bryology* 1982.

43. Halpern CB, Dovčiak M, Urgenson LS, Evans SA. Substrates mediate responses of forest bryophytes to a gradient in overstory retention. *Canadian Journal of Forest Research* 2014; **44**: 855–866.

44. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 2020; **70**: 5607–5612.

45. Ward LM, Cardona T, Holland-Moritz H. Evolutionary Implications of Anoxygenic Phototrophy in the Bacterial Phylum Candidatus Eremiobacterota (WPS-2). *Front Microbiol* 2019; **10**.

46. Ji M, Greening C, Vanwonterghem I, Carere CR, Bay SK, Steen JA, et al. Atmospheric trace gases support primary production in Antarctic desert surface soil. *Nature* 2017; **552**: 400–403.

47. Sheremet A, Jones GM, Jarett J, Bowers RM, Bedard I, Culham C, et al. Ecological and genomic analyses of candidate phylum WPS-2 bacteria in an unvegetated soil. *Environmental Microbiology* 2020; **22**: 3143–3157.

48. Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, Zayed AAF, et al. Genome-centric view of carbon processing in thawing permafrost. *Nature* 2018; **560**: 49–54.

49. Ji M, Williams TJ, Montgomery K, Wong HL, Zaugg J, Berengut JF, et al. Candidatus Eremiobacterota, a metabolically and phylogenetically diverse terrestrial phylum with acid-tolerant adaptations. *ISME J* 2021; **15**: 2692–2707.

50. Pessi IS, Delmont TO, Zehr JP, Hultman J. Discovery of Eremiobacterota with nifH homologues in tundra soil. *Environ Microbiol Rep* 2024; **16**: e13277.

51. 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.

52. Ray AE, Zaugg J, Benaud N, Chelliah DS, Bay S, Wong HL, et al. Atmospheric chemosynthesis is phylogenetically and geographically widespread and contributes significantly to carbon fixation throughout cold deserts. *ISME J* 2022; **16**: 2547–2560.

53. Lin H, Peddada SD. Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. *Nat Methods* 2024; **21**: 83–91.

54. Brenac L, Baidoo EEK, Keasling JD, Budin I. Distinct functional roles for hopanoid composition in the chemical tolerance of Zymomonas mobilis. *Molecular Microbiology* 2019; **112**: 1564–1575.

55. Kulkarni G, Busset N, Molinaro A, Gargani D, Chaintreuil C, Silipo A, et al. Specific Hopanoid Classes Differentially Affect Free-Living and Symbiotic States of Bradyrhizobium diazoefficiens. *mBio* 2015; **6**: 10.1128/mbio.01251-15.

56. Schmerk CL, Bernards MA, Valvano MA. Hopanoid Production Is Required for Low-pH Tolerance, Antimicrobial Resistance, and Motility in Burkholderia cenocepacia. *Journal of Bacteriology* 2011; **193**: 6712–6723.

57. Dobritsa SV, Potter D, Gookin TE, Berry AM. Hopanoid lipids in Frankia: identification of squalene-hopene cyclase gene sequences. *Can J Microbiol* 2001; **47**: 535–540.

58. Cornejo-Castillo FM, Zehr JP. Hopanoid lipids may facilitate aerobic nitrogen fixation in the ocean. *Proc Natl Acad Sci U S A* 2019; **116**: 18269–18271.

59. Kiss L. Chapter 3 Intracellular mycoparasites in action: Interactions between powdery mildew fungi and *Ampelomyces*. *British Mycological Society Symposia Series*. 2008. Elsevier, pp 37–52.

60. Klavina L, Springe G, Nikolajeva V, Martsinkevich I, Nakurte I, Dzabijeva D, et al. Chemical Composition Analysis, Antimicrobial Activity and Cytotoxicity Screening of Moss Extracts (Moss Phytochemistry). *Molecules* 2015; **20**: 17221–17243.

61. Villarreal JC, Renaudin M, Beaulieu-Laliberté A, Bellenger JP. Stigonema associated with boreal Stereocaulon possesses the alternative vanadium nitrogenase. *The Lichenologist* 2021; **53**: 215–220.

62. Ortiz-Rivero J, Garrido-Benavent I, Heiðmarsson S, de los Ríos A. Moss and Liverwort Covers Structure Soil Bacterial and Fungal Communities Differently in the Icelandic Highlands. *Microb Ecol* 2023; **86**: 1893–1908.

63. Taş N, Conejo N, Salmon VG. Metagenomes and Metagenome-Assembled Genome Sequences from Nitrogen-Fixing Alder Nodules. *Microbiol Resour Announc* 2023; **12**: e01266-22.

64. Gordon C, Wynn JM, Woodin SJ. Impacts of increased nitrogen supply on high Arctic heath: the importance of bryophytes and phosphorus availability. *New Phytologist* 2001; **149**: 461–471.

65. Belova SE, Suzina NE, Rijpstra WIC, Sinninghe Damsté JS, Dedysh SN. Edaphobacter lichenicola sp. nov., a member of the family Acidobacteriaceae from lichen-dominated forested tundra. *International Journal of Systematic and Evolutionary Microbiology* 2018; **68**: 1265–1270.

66. Koch IH, Gich F, Dunfield PF, Overmann J. Edaphobacter modestus gen. nov., sp. nov., and Edaphobacter aggregans sp. nov., acidobacteria isolated from alpine and forest soils. *International Journal of Systematic and Evolutionary Microbiology* 2008; **58**: 1114–1122.

67. Shange RS, Ankumah RO, Ibekwe AM, Zabawa R, Dowd SE. Distinct Soil Bacterial Communities Revealed under a Diversely Managed Agroecosystem. *PLOS ONE* 2012; **7**: e40338.

68. Chong CW, Convey P, Pearce DA, Tan IKP. Assessment of soil bacterial communities on Alexander Island (in the maritime and continental Antarctic transitional zone). *Polar Biol* 2012; **35**: 387–399.

69. Rapoport D, Sagova-Mareckova M, Sedláček I, Provaznik J, Králová S, Pavlinic D, et al. Trebonia kvetii gen. nov., sp. nov., an acidophilic actinobacterium, and proposal of the new actinobacterial family Treboniaceae fam. nov. *Int J Syst Evol Microbiol* 2020; **70**: 5106–5114.

70. Weon H-Y, Yoo S-H, Kim Y-J, Lee C-M, Kim B-Y, Jeon Y-A, et al. Rudaea cellulosilytica gen. nov., sp. nov., isolated from soil. *International Journal of Systematic and Evolutionary Microbiology* 2009; **59**: 2308–2312.

71. Mieszkin S, Trouche B, Ancousture J, Raouf Y, Uroz S, Alain K. Scleromatobacter humisilvae gen. nov., sp. nov., a novel bacterium isolated from oak forest soil. *International Journal of Systematic and Evolutionary Microbiology* 2023; **73**: 005793.

72. Klarenberg IJ, Keuschnig C, Salazar A, Benning LG, Vilhelmsson O. Moss and underlying soil bacterial community structures are linked to moss functional traits. *Ecosphere* 2023; **14**: e4447.

73. Escolástico-Ortiz DA, Blasi C, Bellenger J-P, Derome N, Villarreal-A JC. Differentially abundant bacteria drive the N2-fixation of a widespread moss in the forest-tundra transition zone. *Symbiosis* 2023; **90**: 193–211.

74. Tagirdzhanova G, Saary P, Cameron E, Garber A, Díaz Escandón D, Goyette S, et al. Evidence for a core set of microbial lichen symbionts from a global survey of metagenome. 2023.

75. Hay MC, Mitchell AC, Soares AR, Debbonaire AR, Mogrovejo DC, Els N, et al. Metagenome-assembled genomes from High Arctic glaciers highlight the vulnerability of glacier-associated microbiota and their activities to habitat loss. *Microb Genom* 2023; **9**: 001131.

76. Bragina A, Berg C, Berg G. The core microbiome bonds the Alpine bog vegetation to a transkingdom metacommunity. *Molecular Ecology* 2015; **24**: 4795–4807.

77. Schauer S, Kutschera U. A novel growth-promoting microbe, Methylobacterium funariae sp. nov., isolated from the leaf surface of a common moss. *Plant Signaling & Behavior* 2011; **6**: 510–515.

78. Gao Y-Q, Yu J-L, Ren D-M, Zhao D-P, Zhang D-J. Endophytic Methylobacterium in Tissue Culture of the Moss Didymodon tectorum. *anbf* 2022; **59**: 117–122.

79. Noh H-J, Park Y, Hong SG, Lee YM. Diversity and Physiological Characteristics of Antarctic Lichens-Associated Bacteria. *Microorganisms* 2021; **9**: 607.

80. Noh H-J, Shin SC, Park Y, Choi A, Baek K, Hong SG, et al. Lichenicola cladoniae gen. nov., sp. nov., a member of the family Acetobacteraceae isolated from an Antarctic lichen. *International Journal of Systematic and Evolutionary Microbiology* 2020; **70**: 5918–5925.

81. Warshan D, Bay G, Nahar N, Wardle DA, Nilsson M-C, Rasmussen U. Seasonal variation in nifH abundance and expression of cyanobacterial communities associated with boreal feather mosses. *ISME J* 2016; **10**: 2198–2208.

82. Pederson ERA, Warshan D, Rasmussen U. Genome Sequencing of Pleurozium schreberi: The Assembled and Annotated Draft Genome of a Pleurocarpous Feather Moss. *G3 Genes|Genomes|Genetics* 2019; **9**: 2791–2797.

83. Liu L-Y, Xie G-J, Ding J, Liu B-F, Xing D-F, Ren N-Q, et al. Microbial methane emissions from the non-methanogenesis processes: A critical review. *Science of The Total Environment* 2022; **806**: 151362.

84. Bižić M, Klintzsch T, Ionescu D, Hindiyeh MY, Günthel M, Muro-Pastor AM, et al. Aquatic and terrestrial cyanobacteria produce methane. *Science Advances* 2020; **6**: eaax5343.

85. Renaudin M, Laforest-Lapointe I, Bellenger J-P. Unraveling global and diazotrophic bacteriomes of boreal forest floor feather mosses and their environmental drivers at the ecosystem and at the plant scale in North America. *Science of The Total Environment* 2022; **837**: 155761.

86. Gavazov KS, Soudzilovskaia NA, van Logtestijn RSP, Braster M, Cornelissen JHC. Isotopic analysis of cyanobacterial nitrogen fixation associated with subarctic lichen and bryophyte species. *Plant Soil* 2010; **333**: 507–517.

87. Garrido-Oter R, Nakano RT, Dombrowski N, Ma K-W, AgBiome Team, McHardy AC, et al. Modular Traits of the Rhizobiales Root Microbiota and Their Evolutionary Relationship with Symbiotic Rhizobia. *Cell Host Microbe* 2018; **24**: 155-167.e5.

88. Kolton M, Weston DJ, Mayali X, Weber PK, McFarlane KJ, Pett-Ridge J, et al. Defining the Sphagnum Core Microbiome across the North American Continent Reveals a Central Role for Diazotrophic Methanotrophs in the Nitrogen and Carbon Cycles of Boreal Peatland Ecosystems. *mBio* 2022; **13**: e03714-21.

89. Nissinen R, Franz O, Kovanen S, Mäkelä M, Kraft V, Ketola K, et al. Aerobic anoxygenic phototrophic bacteria are ubiquitous in phyllo- and endosphere microbiomes of boreal and subarctic plants. 2023. bioRxiv. , 2023.02.19.529139

90. Zervas A, Zeng Y, Madsen AM, Hansen LH. Genomics of Aerobic Photoheterotrophs in Wheat Phyllosphere Reveals Divergent Evolutionary Patterns of Photosynthetic Genes in Methylobacterium spp. *Genome Biology and Evolution* 2019; **11**: 2895–2908.

91. Ingram-Smith C, Martin SR, Smith KS. Acetate kinase: not just a bacterial enzyme. *Trends in Microbiology* 2006; **14**: 249–253.

92. Smith AR, Kieft B, Mueller R, Fisk MR, Mason OU, Popa R, et al. Carbon fixation and energy metabolisms of a subseafloor olivine biofilm. *ISME J* 2019; **13**: 1737–1749.

93. Estrada-De Los Santos P, Bustillos-Cristales R, Caballero-Mellado J. Burkholderia, a Genus Rich in Plant-Associated Nitrogen Fixers with Wide Environmental and Geographic Distribution. *Applied and Environmental Microbiology* 2001; **67**: 2790–2798.

94. Li J, Yu C, Liu Z, Wang Y, Wang F. Microplastic accelerate the phosphorus-related metabolism of bacteria to promote the decomposition of methylphosphonate to methane. *Sci Total Environ* 2023; **858**: 160020.

95. Ney B, Ahmed FH, Carere CR, Biswas A, Warden AC, Morales SE, et al. The methanogenic redox cofactor F420 is widely synthesized by aerobic soil bacteria. *ISME J* 2017; **11**: 125–137.

96. Wu H-H, Pun MD, Wise CE, Streit BR, Mus F, Berim A, et al. The pathway for coenzyme M biosynthesis in bacteria. *Proceedings of the National Academy of Sciences* 2022; **119**: e2207190119.

97. Yang Y, Cheng S, Fang H, Guo Y, Li Y, Zhou Y, et al. Linkages between the molecular composition of dissolved organic matter and soil microbial community in a boreal forest during freeze–thaw cycles. *Front Microbiol* 2023; **13**.

98. Ikeda MA, Nakamura H, Sawada K. Aliphatic hydrocarbons in the lichen class Lecanoromycetes and their potential use as chemotaxonomic indicators and biomarkers. *Organic Geochemistry* 2023; **179**: 104588.

99. Moussaoui L, Leduc A, Girona MM, Bélisle AC, Lafleur B, Fenton NJ, et al. Success Factors for Experimental Partial Harvesting in Unmanaged Boreal Forest: 10-Year Stand Yield Results. *Forests* 2020; **11**: 1199.

100. Environment and Climate Change Canada Historical Climate Data. Monthly Climate Summaries. 2021.

101. Collin A, Messier C, Kembel SW, Bélanger N. Can sugar maple establish into the boreal forest? Insights from seedlings under various canopies in southern Quebec. *Ecosphere* 2018; **9**: e02022.

102. Carter MR, Gregorich EG (eds). Soil Sampling and Methods of Analysis, 2nd ed. 2007. CRC Press, Boca Raton.

103. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 2014; **15**: 121–132.

104. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2017; **2**: 1533–1542.

105. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 2018; **6**: 158.

106. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 2017; **35**: 725–731.

107. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* 2017; **11**: 2864–2868.

108. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology* 2016; **17**: 132.

109. Cirkoska E, Kishore C, Bose SC, Mubarakali A, Reshmi TR, Marina N. A Study on Suffix Trees and Their Applications in Genome Sequences Using MUMmer. In: Smys S, Palanisamy R, Rocha Á, Beligiannis GN (eds). *Computer Networks and Inventive Communication Technologies*. 2021. Springer Nature, Singapore, pp 15–28.

110. Orakov A, Fullam A, Coelho LP, Khedkar S, Szklarczyk D, Mende DR, et al. GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biology* 2021; **22**: 178.

111. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2020; **36**: 1925–1927.

112. Shaw J, Yu YW. Fast and robust metagenomic sequence comparison through sparse chaining with skani. *Nat Methods* 2023; **20**: 1661–1665.

113. Brown CT, Irber L. sourmash: a library for MinHash sketching of DNA. *Journal of Open Source Software* 2016; **1**: 27.

114. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**: 119.

115. Ruiz-Perez CA, Conrad RE, Konstantinidis KT. MicrobeAnnotator: a user-friendly, comprehensive functional annotation pipeline for microbial genomes. *BMC Bioinformatics* 2021; **22**: 11.

116. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat Commun* 2020; **11**: 2500.

117. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 2017; **8**: 28–36.

118. Sun Z, Huang S, Zhang M, Zhu Q, Haiminen N, Carrieri AP, et al. Challenges in benchmarking metagenomic profilers. *Nat Methods* 2021; **18**: 618–626.

119. R Core Team. R: A language and environment for statistical computing. 2022.

120. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 2014; **15**: 550.

121. Tian W, Wang H, Xiang X, Wang R, Xu Y. Structural Variations of Bacterial Community Driven by Sphagnum Microhabitat Differentiation in a Subalpine Peatland. *Frontiers in Microbiology* 2019; **10**: 1–11.

122. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. vegan: Community Ecology Package. 2022.

123. Simpson GL. permute: Functions for Generating Restricted Permutations of Data. R package version 0.9-5. 2019.

124. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* 2013; **8**: e61217.

125. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 2015; **67**: 1–48.

126. Hartig F. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. 2020.

127. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front Microbiol* 2017; **8**.

128. Nearing JT, Douglas GM, Hayes MG, MacDonald J, Desai DK, Allward N, et al. Microbiome differential abundance methods produce different results across 38 datasets. *Nat Commun* 2022; **13**: 342.

129. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000; **28**: 27–30.

130. Eisenhofer R, Odriozola I, Alberdi A. Impact of microbial genome completeness on metagenomic functional inference. *ISME Communications* 2023; **3**: 12.

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).

		a) with *D. undulatum*			b) without *D. undulatum*		
	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 R^2 (%)		**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) with *D. undulatum*			b) without *D. undulatum.*		
	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 χ^2 (%)		**62**			**59**	

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 coordinates analysis (PCoA) of bacterial communities based on Bray-Curtis dissimilarity index, calculated using variance-stabilizing-

transformed abundance table. Each point represents one moss sample, and ellipses indicate a 95% bivariate confidence interval. Each ellipse colour represents a moss species (a) or a moss gametophyte section (b). (c) Differences in alpha diversity (Shannon diversity) of moss bacterial communities. Showed are four moss species: *P. juniperinum* (N_{brown} = 24, N_{green} = 22), *P. commune* (N_{brown} = 9, N_{green} = 9), *P. piliferum* (N_{brown} = 26, N_{green} = 26), and *D. undulatum* (N_{brown} = 8, N_{green} = 7). (d) Mean sequence relative abundance by bacterial orders between green (top panel) and brown (bottom panel) gametophyte sections across four moss species.

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 (QS). The inner section shows the phylogenetic tree of 110 metagenome-assembled genomes constructed using PhyloPhlAn and built phylogenetic trees using the ggtree package, allowing a cross-validation with their assigned taxonomy in the GTDB database. Two clades were renamed using NCBI taxonomy, as the GTDB nomenclature used illegitimate names: Nostocales (Cyanobacteriales) and Hyphomicrobiales (Rhizobiales). The different colored dots on the branches of the tree correspond to the colors of the bacterial orders listed in the legend.

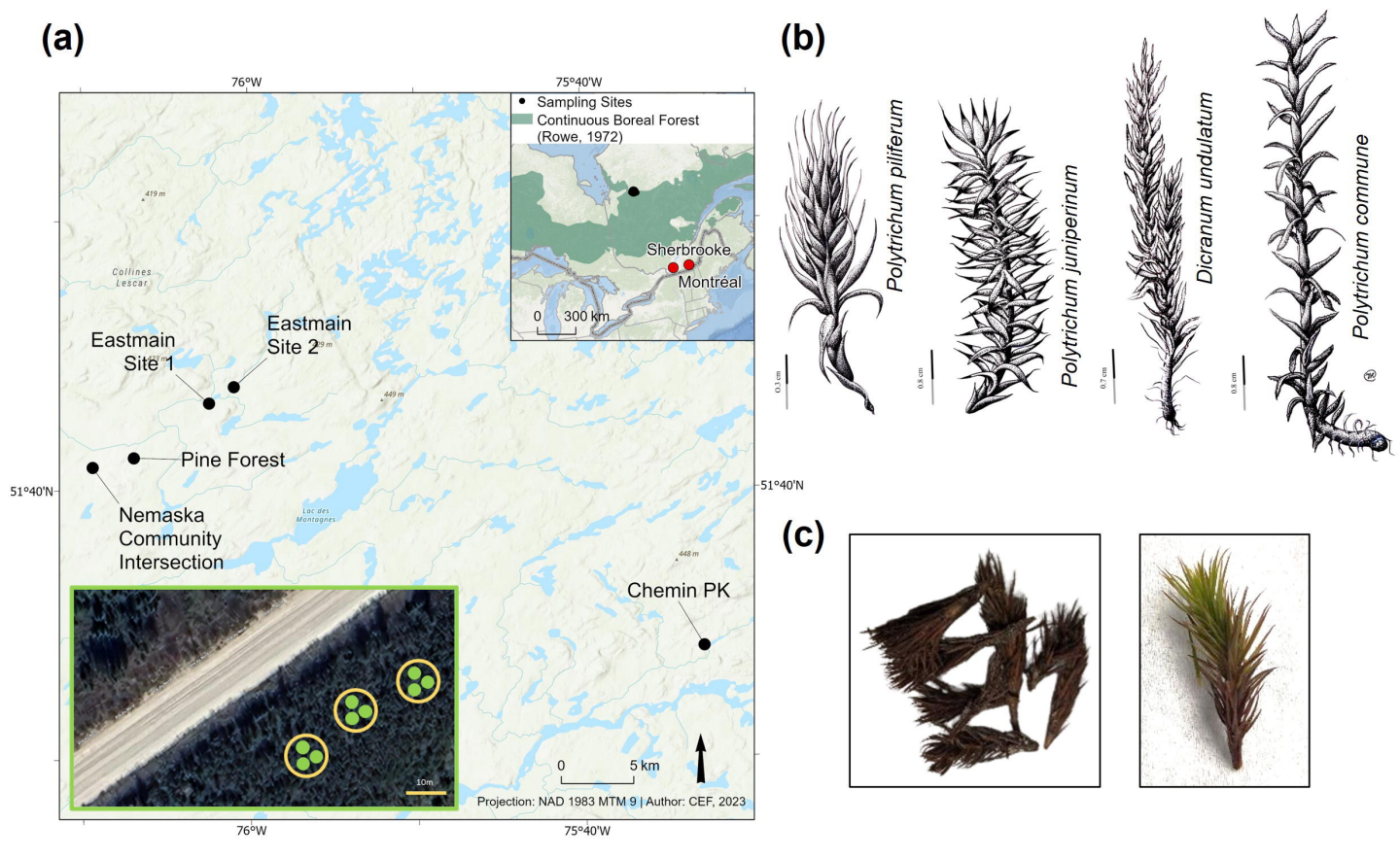
Figure 4. Differential abundance of bacterial species between moss species and between green and brown gametophyte sections.

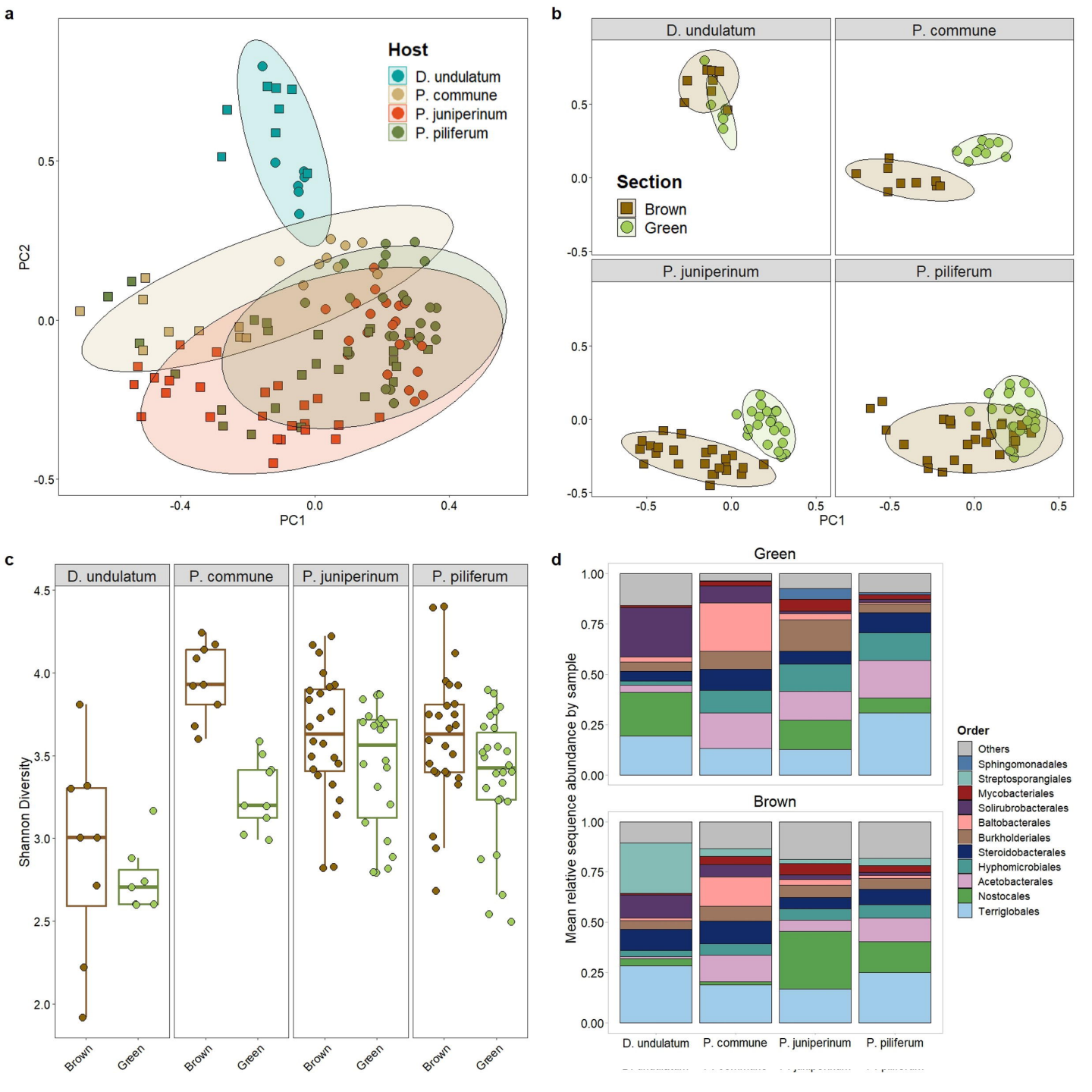
LFC: $\log_2$-fold change in absolute abundances, with all values showing significant differential abundance ($p < 0.01$). (a) Differentially abundant bacterial species in moss

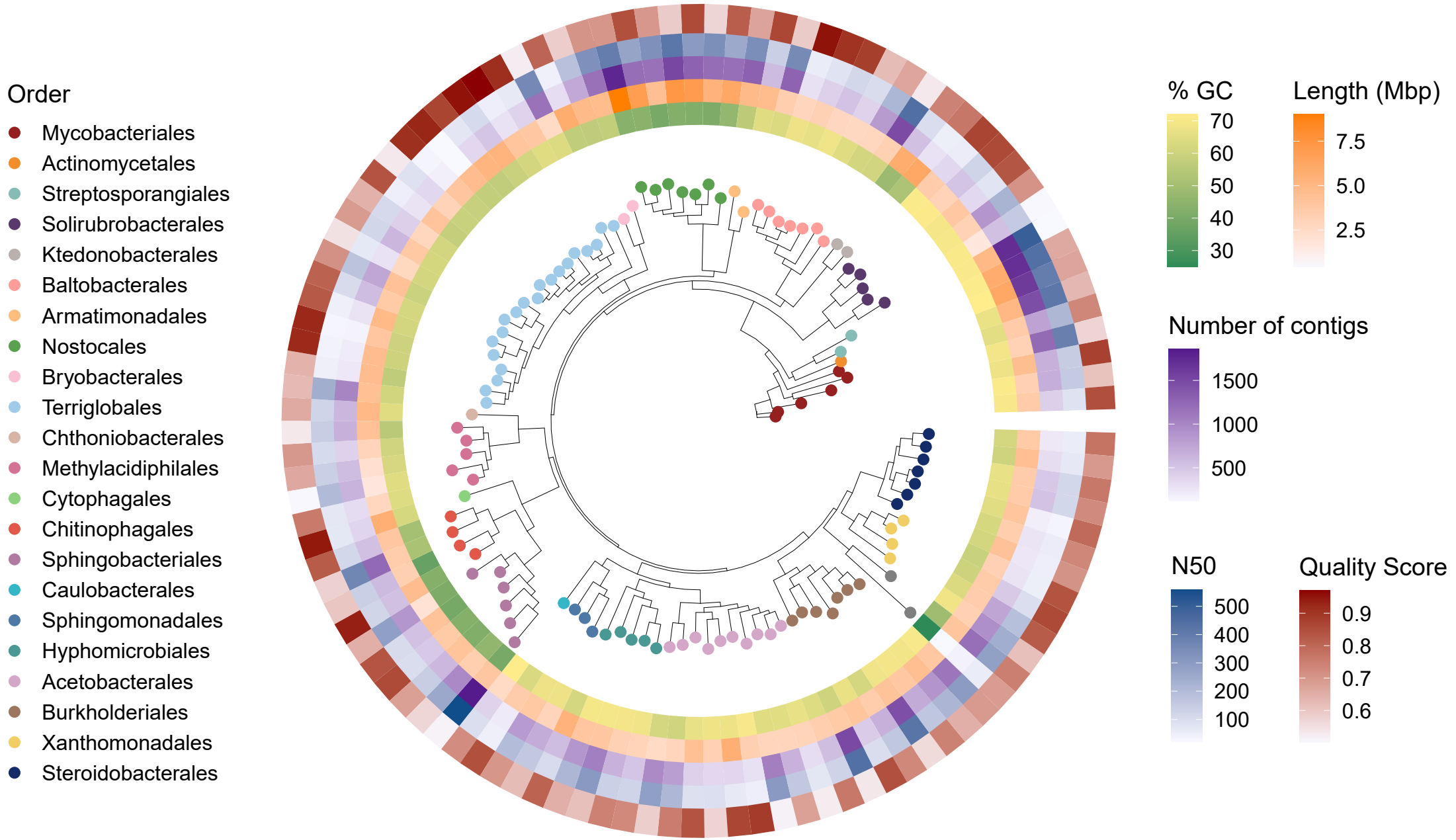
species relative to *D. undulatum*. LFC are relative to *D. undulatum*. Positive LFC (blue) indicates enrichment in specified moss species relative to *D. undulatum*, whereas negative LFC (red) indicates enrichment in *D. undulatum* relative to specified moss species. (b) Bacterial species differentially 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.

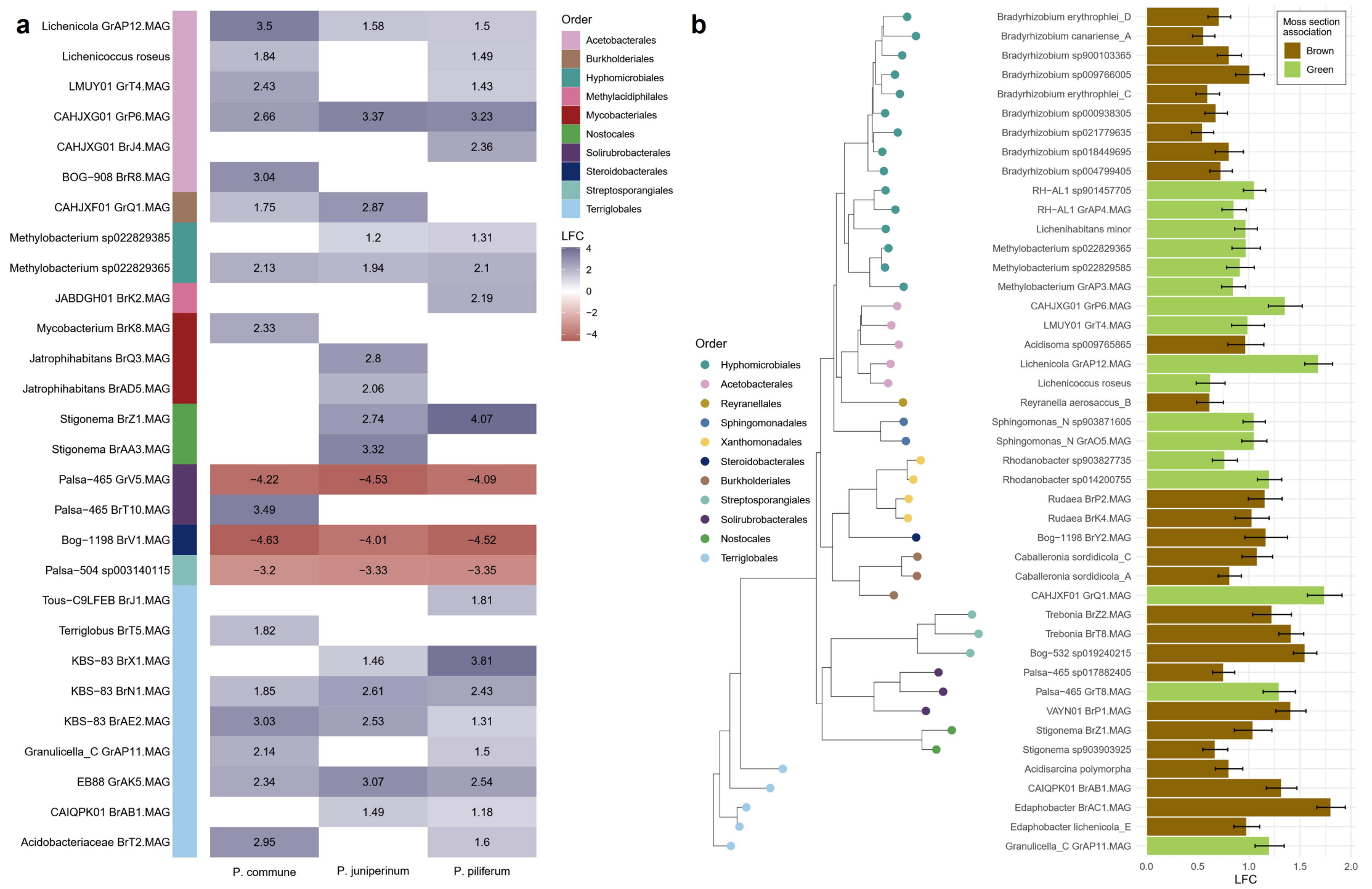
Figure 5. Metabolic abilities of the differentially abundant species in brown and green gametophyte sections.

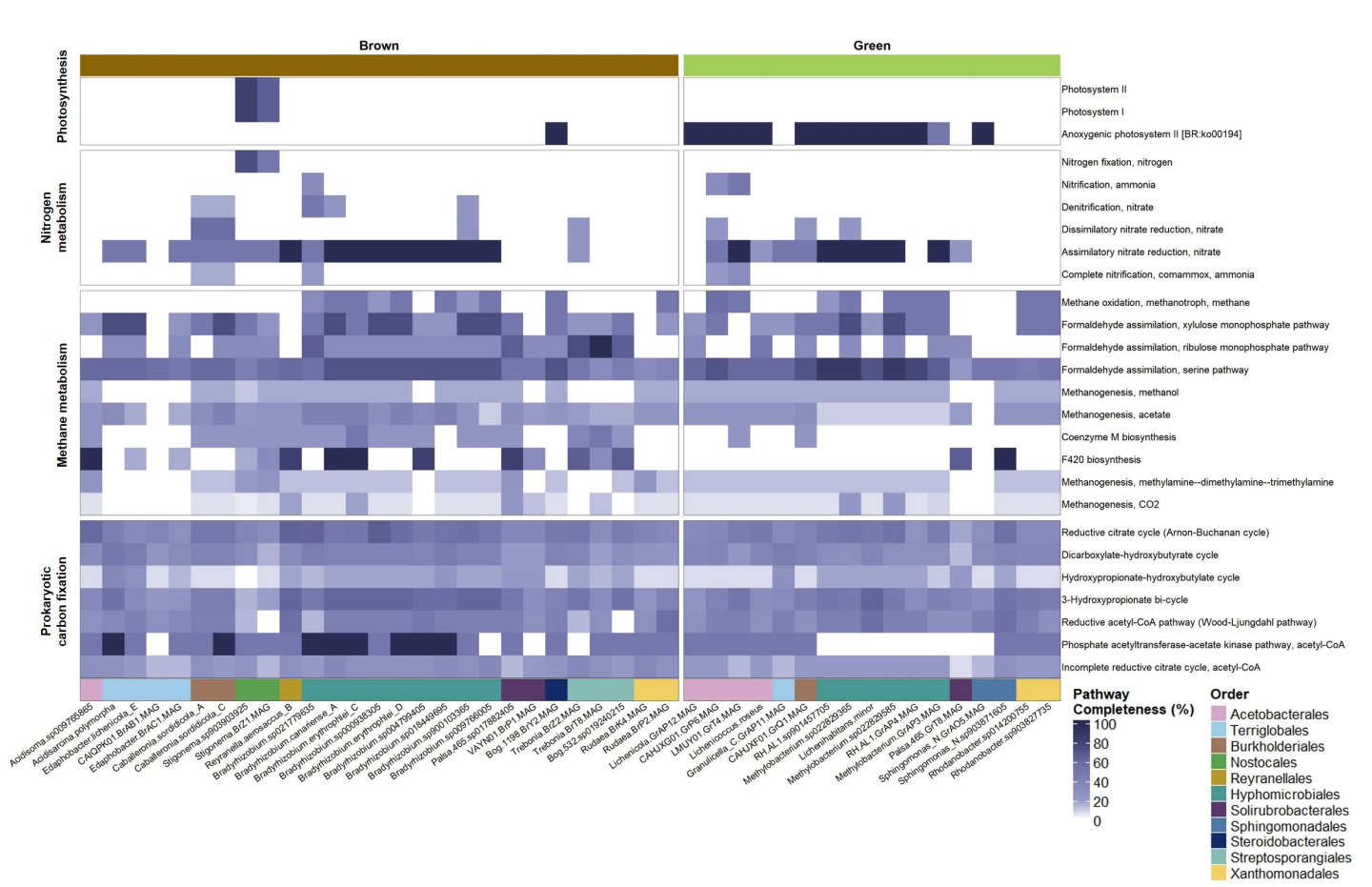
The pathway completeness of selected metabolic pathway groups, such as photosynthesis, nitrogen and methane metabolism, and prokaryotic carbon fixation, is indicated for each genome (blue scale). This figure also shows the taxonomic rank (e.g., bacterial order) for each of 44 differentially abundant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


(a)
76°W
75°40'W
51°40'N
Sampling Sites
Continuous Boreal Forest (Rowe, 1972)
Sherbrooke
Montréal
0
300 km
Collines Lescar
Eastmain Site 1
Eastmain Site 2
Pine Forest
Nemaska Community Intersection
Lac des Montagnes
Chemin PK
0
5 km
10m
Projection: NAD 1983 MTM 9 | Author: CEF, 2023
(b)
Polytrichum piliferum
Polytrichum juniperinum
Dicranum undulatum
Polytrichum commune
(c)


a
Host
D. undulatum
P. commune
P. juniperinum
P. piliferum
PC2
PC1
b
D. undulatum
P. commune
P. juniperinum
P. piliferum
Section
Brown
Green
PC1
c
D. undulatum
P. commune
P. juniperinum
P. piliferum
Shannon Diversity
Brown
Green
d
Green
Brown
Mean relative sequence abundance by sample
Order
Others
Sphingomonadales
Streptosporangiales
Mycobacteriales
Solirubrobacterales
Baltobacterales
Burkholderiales
Steroidobacterales
Hyphomicrobiales
Acetobacterales
Nostocales
Terriglobales
D. undulatum
P. commune
P. juniperinum
P. piliferum


Order
Mycobacteriales
Actinomycetales
Streptosporangiales
Solirubrobacterales
Ktedonobacterales
Baltobacterales
Armatimonadales
Nostocales
Bryobacterales
Terriglobales
Chthoniobacterales
Methylacidiphilales
Cytophagales
Chitinophagales
Sphingobacteriales
Caulobacterales
Sphingomonadales
Hyphomicrobiales
Acetobacterales
Burkholderiales
Xanthomonadales
Steroidobacterales
% GC
70
60
50
40
30
Length (Mbp)
7.5
5.0
2.5
Number of contigs
1500
1000
500
N50
500
400
300
200
100
Quality Score
0.9
0.8
0.7
0.6


a
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
Methylobacterium sp022829385 1.2 1.31
Methylobacterium sp022829365 2.13 1.94 2.1
JABDGH01 BrK2.MAG 2.19
Mycobacterium BrK8.MAG 2.33
Jatrophihabitans BrQ3.MAG 2.8
Jatrophihabitans BrAD5.MAG 2.06
Stigonema BrZ1.MAG 2.74 4.07
Stigonema BrAA3.MAG 3.32
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
P. commune
P. juniperinum
P. piliferum
Order
Acetobacterales
Burkholderiales
Hyphomicrobiales
Methylacidiphilales
Mycobacteriales
Nostocales
Solirubrobacterales
Steroidobacterales
Streptosporangiales
Terriglobales
LFC
4
2
0
−2
−4
b
Order
Hyphomicrobiales
Acetobacterales
Reyranellales
Sphingomonadales
Xanthomonadales
Steroidobacterales
Burkholderiales
Streptosporangiales
Solirubrobacterales
Nostocales
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


Brown
Green
Photosynthesis
Photosystem II
Photosystem I
Anoxygenic photosystem II [BR:ko00194]
Nitrogen metabolism
Nitrogen fixation, nitrogen
Nitrification, ammonia
Denitrification, nitrate
Dissimilatory nitrate reduction, nitrate
Assimilatory nitrate reduction, nitrate
Complete nitrification, comammox, ammonia
Methane metabolism
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
Prokaryotic carbon fixation
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.sp00976586 5
Acidipiscina.polymorpha
Edaphobacter.lichenicola_E
CAQPK01.BrAB1.MAG
Edaphobacter.BrAC1.MAG
Caballeronia.sordidicola_A
Caballeronia.sordidicola_C
Stigonema.sp003923925
Stigmonema.aeroaquae_B
Reyranella.BrZ1.MAG
Bradyrhizobium.sp021779635
Bradyrhizobium.canariense_A
Bradyrhizobium.erythrophei_C
Bradyrhizobium.sp000983305
Bradyrhizobium.erythrophei_D
Bradyrhizobium.sp004769405
Bradyrhizobium.sp018489695
Bradyrhizobium.sp900103365
Bradyrhizobium.sp009766005
Paliea.465.sp017862405
VAYN01.BrP1.MAG
Bog-1198.BrY2.MAG
Trebonia.B/Z2.MAG
Trebonia.B/T9.MAG
Bog.532.sp019240215
Rudaea.B/K4.MAG
Rudaea.B/P2.MAG
Lichenicola.GAP12.MAG
CAHJXG01.GrP6.MAG
LMUY01.GrT4.MAG
Lichenicoccus.roseus
Granulicella_C.GRAP11.MAG
CAHJXF01.GrQ1.MAG
RH.AL1.sp901457705
Methylobacterium.sp022829365
Lichenihabitans.minor
Methylobacterium.sp022830585
Methylobacterium.GrAP4.MAG
RH.AL1.GrAP3.MAG
Paliea.465_G/T9.MAG
Sphingomonas_N.GrAO5.MAG
Sphingomonas_N.sp903811605
Rhodanobacter.sp014200755
Rhodanobacter.sp903827735
Pathway Completeness (%)
100
80
60
40
20
0
Order
Acetobacterales
Terriglobales
Burkholderiales
Nostocales
Reyranellales
Hyphomicrobiales
Solirubrobacterales
Sphingomonadales
Steroidobacterales
Streptosporangiales
Xanthomonadales