A phylogenomic and metagenomic meta-analysis of bacterial diversity in the phyllosphere lifts a veil on Hyphomicrobiales dark matter

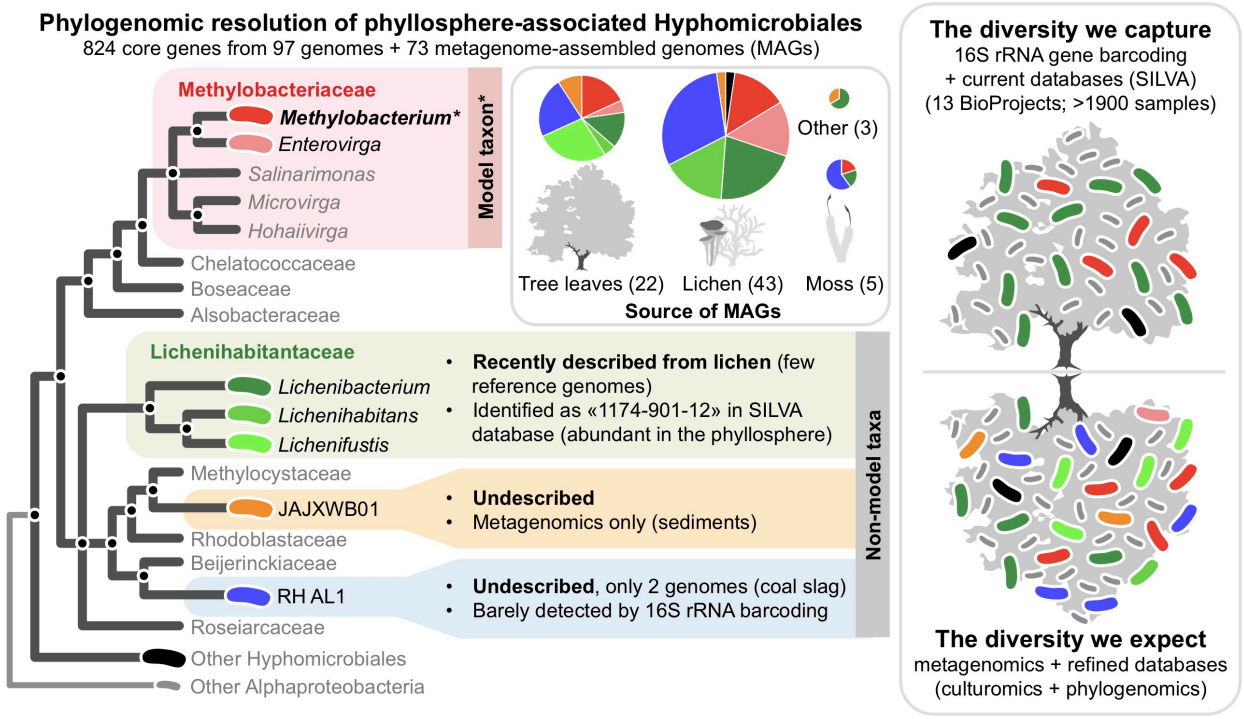
Authors: Jean-Baptiste Leducq[1]*, Louis-Philippe St-Amand[1], David Ross[2] and Steven W. Kembel[2]

Affiliations: 1 - Département de Phytologie (FSAA), IBIS, CRIV - Université Laval - Québec (QC) G1V 0A6 – Canada

2 - Département des sciences biologiques – Université du Québec à Montréal – Canada

Email address: (*corresponding author): jean-baptiste.leducq.1@fsaa.ulaval.ca

Graphical Abstract



16 **Highlights**

17 - Global meta-analysis reveals core Hyphomicrobiales in the phyllosphere
18 - *Lichenibacterium* and *Methylobacterium* dominate the phyllosphere and exposed
19 surfaces
20 - Novel phyllosphere clades RH-AL1 and JAJXWB01 were identified by phylogenomics
21 - 16S rRNA gene limits taxonomic resolution in the phyllosphere; metagenomics refines
22 taxonomy

23

24 **Keywords:** Phyllosphere, Hyphomicrobiales, *Methylobacterium*, *Lichenibacterium*,
25 metagenomics, phylogenomics.

26

Abstract - The phyllosphere, or above-ground part of plants, hosts diverse bacterial communities that play critical ecological roles and provide beneficial functions for the plant. Yet, the diversity of major plant-associated bacteria groups like Hyphomicrobiales (Alphaproteobacteria) remain largely uncharacterized in the phyllosphere, due to limitations of 16S rRNA metabarcoding and biases in reference databases toward model taxa and environments. Using a meta-analysis combining metabarcoding, metagenomics and phylogenomics, we explored the worldwide diversity of leaf-associated Hyphomicrobiales. We confirmed *Methylobacterium* was ubiquitous in the phyllosphere and revealed the dominance of two under-characterized Hyphomicrobiales taxa: *Lichenibacterium*, a recently described genus of lichen-associated bacteria previously identified as “1174-901-12” in taxonomic databases, and RH-AL1, an undescribed genus of bacteria related to Beijerinckiaceae, previously isolated from coal slag. Despite their abundance in the phyllosphere, *Lichenibacterium* and RH_AL1 could not be properly detected by classical 16S rRNA gene barcoding, underlining biases in reference databases and a lack of taxonomic resolution for this marker gene. As for *Methylobacterium,* a significant proportion of *Lichenibacterium* and RH-AL1 were also detected in association with lichens and in environments with harsh conditions like exposed surfaces, air and snow, suggesting airborne or waterborne dispersal and high resilience in harsh environments. Overall our study stresses the need to move toward metagenomics and culturomics to increase the representativeness of leaf-associated bacterial taxa in reference databases, and to improve our understanding of the evolutionary and functional mechanisms underpinning bacteria adaptations to living on plants.

INTRODUCTION

The phyllosphere, or the aerial parts of plants including leaves, has been intensively studied due to its crucial role in global carbon and nitrogen cycles. Recent advances in microbial ecology also revealed that, despite exposure to UV radiation, drought, and rapid environmental fluctuations, plant leaves host a large microbial diversity providing essential functions for plant health including nutrient acquisition, deterring pathogens, and facilitating resilience to stresses [40]. The phyllosphere microbiota can also modulate leaf surface chemical composition by releasing plant growth promoters such as vitamins and hormones, and by degrading toxic compounds such as pesticides [47]. Our knowledge about beneficial plant-associated bacterial diversity has been dominated by studies of the rhizosphere, and the phyllosphere was less well studied until recently, first because of technical biases, as classical microbiology often relied on unrealistic lab conditions developed for the optimal growth of a few model microbes, and second for biological reasons, because of the inherent characteristics of different plants tissues. Roots are less exposed to environmental variation and are more nutrient-rich habitats than leaves. Therefore, root-associated microbes appear to be more abundant, and microbial diversity and culture conditions are more predictable in the rhizosphere than in the phyllosphere. For instance, Hyphomicrobiales (previously Rhizobiales) is a major plant-associated bacterial order that includes the intensively studied nitrogen-fixing bacteria in root nodules [12,42], while the phyllosphere-associated Hyphomicrobiales are mostly known for *Methylobacterium* (Methylobacteriaceae), a ubiquitous group of bacteria metabolizing one-carbon compounds such as methanol that are released by plants during cell-wall synthesis [38]. As a result, only a tiny fraction of leaf-associated microbes have been studied and cultivated so far, stressing the need for more studies that consider the phyllosphere, and for culture conditions that maximize their isolation success, such as those developed for the *Arabidopsis* microbiome [6]. Since the advent of molecular biology, microbial taxonomy has been based on universal marker genes (i.e. the 16S rRNA gene in bacteria), allowing identification of major taxonomic groups and their

75 phylogenetic relationships [28]. Unfortunately, those markers do not give the resolution required
76 to confidently define lower taxonomic units (genus, species), introducing unwanted artefacts in
77 the phylogeny and classification of major groups of plant-associated microbes including
78 *Methylobacterium* [2]. Alternatively, the *rpoB* gene has higher resolution [27] and has been used
79 as a molecular marker to infer *Methylobacterium* diversity in the phyllosphere, with potential use
80 at a larger taxonomic scale in the Hyphomicrobiales [19]. Overall, metabarcoding has allowed
81 the quantification of microbial diversity without using cultivation, but is still limited by using a
82 single marker gene. Alternatively, shotgun metagenomics is a promising approach to infer the
83 taxonomy, function and diversity of phyllosphere bacterial communities. This approach is
84 however limited by the overall low bacterial abundance on the surface of plant leaves, and by
85 the presence of high concentrations of plant host DNA for endophytic organisms. Finally, whole
86 genome sequencing (WGS) of isolates, including plant-associated bacteria, coupled with
87 phylogenomics (whole-genome-based phylogeny) substantially improved the taxonomy of
88 Hyphomicrobiales in recent years [12,16]. Regardless of the method, taxonomic databases are
89 still biased toward model taxa like *Methylobacterium*, and we argue that major phyllosphere-
90 associated Hyphomicrobiales taxa are still omitted, underrepresented or misclassified in these
91 databases. For instance, *1174-901-12* is an uncultured Hyphomicrobiales taxon identified in the
92 16S rRNA gene SILVA database [32] that has been consistently detected worldwide on the
93 surface of plant leaves [3,15,18,19,22,35,43,45], where it typically represents 3 to 22% of total
94 bacteria abundance. Little is known about the taxonomy and the ecological role of *1174-901-12*,
95 although its detection in lichens [34], in aerosols [24] and on various exposed surfaces
96 [4,25,33,39] suggests a broader ecological spectrum than the phyllosphere. According to the
97 SILVA database (v138.1), *1174-901-12* is member of the Beijerinckiaceae, a catch-all family of
98 phyllosphere-associated Hyphomicrobiales also including *Methyllocella*, *Methylocystis*,
99 *Methylobacterium* and *Roseiarcus*, for instance. According to the List of Prokaryotic names with
100 Standing in Nomenclature (LPSN), these genera belong to distinct families, with only

101 *Methyllocella* being a “true” Beijerinckiaceae member. The underrepresentation and
102 misclassification of major phyllosphere-associated Hyphomicrobiales such as *Methylobacterium*
103 and *1174-901-12* illustrates the limits of the 16S rRNA gene in describing the composition of
104 microbial communities in the phyllosphere. It also stresses the need to move towards a
105 metagenomic approach to enrich databases with genomes from uncultured taxa, and to a
106 phylogenomic-based approach to clarify their taxonomy.
107
108 Using a meta-analysis combining metabarcoding, metagenomics and phylogenomics, we
109 explored worldwide the diversity of leaf-associated Hyphomicrobiales. We confirmed the ubiquity
110 of *Methylobacterium* in the phyllosphere and revealed the dominance of two under-characterized
111 Hyphomicrobiales taxa: *Lichenibacterium*, a recently described genus of lichen-associated
112 bacteria formerly identified as *1174-901-12*, and RH-AL1, an undescribed genus of bacteria
113 related to Beijerinckiaceae, previously isolated from coal slag. The fact that *Lichenibacterium*
114 and RH_AL1 were not properly detected by classical 16S rRNA gene barcoding despite their
115 abundance in the phyllosphere illustrates biases in reference databases and the lack of
116 resolution for this marker gene. Like for *Methylobacterium,* a significant proportion of
117 *Lichenibacterium* and RH-AL1 were also detected in lichen and in environments with harsh
118 conditions like exposed surfaces, air and snow, suggesting airborne or waterborne dispersal and
119 high resilience in stressful environments. Overall, our study stresses the biases of classical tools
120 and reference database, as well as the power of metagenomics and phylogenomic to reveal
121 uncultured leaf-associated bacteria. Future works should also use such tools, as well as
122 culturomics, to increase the representation of these taxa in reference databases, and to address
123 evolutionary and functional mechanisms underpinning bacterial adaptation to life on plants.
124
125 **METHODS**
126

127 ***Metagenome and genome collection of representative Hyphomicrobiales taxa from the***
128 ***phyllosphere, moss and lichen***
129
130 We assembled a collection of Hyphomicrobiales (syn. Rhizobiales) genomes and metagenome-
131 assembled genomes (MAGs) covering the diversity observed at the surface of tree leaves. We
132 used 22 MAGs from the leaves of 25 tree species occurring in the Canadian temperate forest,
133 that were previously identified as Rhizobiales with the GTDB database (unpublished data;
134 bioproject PRJNA1291777; Dataset 1). According to the GTDB-based taxonomic classification,
135 these 22 leaf-associated MAGs were assigned to family Beijerinckiaceae, in genera
136 *Methylobacterium* (n= 4), *Enterovirga* (n=1), *Lichenihabitans* (n= 8), RH-AL1 (n= 5) JAJXWB01
137 (n= 2) or unidentified at the genus level (n=2; Dataset 1). From here, we use exclusively the
138 LPSN as a reference for the taxonomy of Hyphomicrobiales. According to the LPSN,
139 *Methylobacterium* and *Enterovirga* are part of Methylobacteriaceae, a well-characterized
140 phyllosphere-associated family [2]. *Lichenihabitans* [26] is part of Lichenihabitantaceae, a
141 recently described lichen-associated family of bacteria, also including genera *Lichenifustis* [30],
142 *and Lichenibacterium* [29]. RH_AL1 is represented by two strains isolated from coal slag [44].
143 JAJXWB01 was identified by metagenomics in mine drainage sediments (unpublished data;
144 bioproject PRJNA666025; Dataset 1) and has no reference genome in NCBI.
145 Accordingly, we retrieved 97 reference genomes from NCBI, spanning the taxonomic and
146 phylogenic range of Hyphomicrobiales taxa identified in leaf-associated MAGs
147 (Methylobacteriaceae, Lichenihabitantaceae). For Methylobacteriaceae, we maximized the
148 genomic diversity of reference genomes by selecting the best assembly from each lineage
149 previously identified in *Methylobacterium* (n=23; [2]), *Enterovirga* (n=3), *Microvirga* (n=2) and in
150 a newly described genera, *Hohaiivirga* (n=1; [41]; Dataset 1). Because other leaf-associated
151 taxa had fewer representatives in genomic databases, we retrieved all available genomes for
152 Lichenihabitantaceae (n=8), and RH-AL1 [44] (n=2; Dataset 1). According to the most recent

153 phylogenomic-based study on Hyphomicrobiales [12], the aforementioned taxa form a
154 monophyletic group within Hyphomicrobiales, which we will refer as the phyllosphere-associated
155 Hyphomicrobiales, using Rhabdaerophilaceae as the closest known outgroup, for which we
156 retrieved the reference genome (Dataset 1). To mitigate biases in phylogenomic reconstruction
157 of the whole group, we also included NCBI RefSeq genomes from families that were not
158 identified in leaf-associated MAGs, but were also members of this monophyletic group:
159 Methylocystaceae (n=12), Beijerinckiaceae (n=10), Roseiarcaceae (n=1), Rhodoblastaceae
160 (n=3), Chelatococcaceae (n=11), Alsobacteraceae (n=3), Boseaceae (n=14) and
161 Salinarimonadaceae (n=3; Dataset 1).
162 We also retrieved any available MAGs in NCBI databases that could be assigned to the
163 phyllosphere-associated Hyphomicrobiales families. Most identified MAGs came from a
164 worldwide survey of lichen-associated microbiota (n=43; [37]) and from a study on moss-
165 associated microbiota from the Canadian boreal forest (n=5; [17]). Three additional
166 phyllosphere-associated Hyphomicrobiales MAGs were identified in glacier sediments [9], in
167 mining soil [23] and in mine drainage sediments (JAJXWB01; unpublished data; Bioproject
168 PRJNA666025; Dataset 1).
169
170 ***Phylogenomic reconstruction of plant-associated Hyphomicrobiales***
171
172 We reconstructed the phylogenetic tree of phyllosphere-associated Hyphomicrobiales as
173 described previously for Methylobacteriaceae [20] with the following modifications. Briefly, we
174 annotated *de novo* the 97 genomes and 73 MAGs using RAST [5] and identified core genes (i.e.
175 genes present in a single copy in all genomes), excluding hypothetical proteins. Because our
176 analysis included highly fragmented MAGs, we used more relaxed criteria to define the core
177 genome. First, we considered any gene present on average in 2 copies or less in at least 80% of
178 genome assemblies as candidate core genes (MAGs excluded). Second, we estimated the real

179 copy number of each candidate core gene in each assembly. As the proxy for the copy number,
180 we normalized the total size of gene copies observed for a given gene in each assembly by the
181 average size of all copies observed across all assemblies. We considered any normalized value
182 lower than 0.7 (missing or too fragmented copy) or above 1.3 (likely duplication) as missing data.
183 Third, we only considered candidate core genes with missing data in less than 30% of
184 assemblies as core genes. We performed sequence alignment for each core gene individually,
185 with the R package *msa* [8], using nucleotide sequence for rna-encoding genes and amino-acid
186 sequence for protein-encoding genes. We computed a maximum-likelihood phylogenetic tree for
187 each core gene independently using RAxML (GTRGAMMA model, 1,000 replicated trees [36]).
188 We reconstructed the phyllosphere-associated Hyphomicrobiales consensus tree using
189 ASTRAL-III, a coalescent-based method combining ML trees determined for each core gene
190 independently, accounting for incomplete lineage sorting and horizontal gene transfers [46].
191
192 **Identification of *1174-901-12* as *Lichenihabitantaceae***
193
194 We performed a preliminary phylogenetic analysis of 16S rRNA sequences used to define *1174-*
195 *901-12* in the SILVA database, version 138.1, in order to validate that this taxon corresponded to
196 *Lichenibacterium*, as identified in the SILVA database version 138.2. This included the 24 SILVA
197 database sequences (that we numbered from 1 to 24), all complete 16S rRNA gene sequences
198 available for the family Lichenihabitantaceae in NCBI databases, and two sequences from
199 *Roseiarcus* as outgroup. The Maximum-Likelihood tree was inferred with RAxML (1,000
200 bootstraps, GTRGAMMA).
201
202 ***Meta-analysis of barcoding-based phyllosphere studies***
203

204 We selected previously published metabarcoding data based on the 16S rRNA gene from the
205 tree phyllosphere and potentially related microbial environments and sources (other plants,
206 mosses, lichen, exposed surfaces, air, snow). We restricted our survey to 13 recent studies
207 (Dataset 2) for which BioProjects and related metadata were publicly available. For each
208 BioProject, we processed 16S rRNA gene reads to obtain an ASV (Amplicon Sequence
209 Variants) abundance table per sample, using the package *dada2* in R [10] as described
210 previously [19], with the following modifications. According to sequence quality profiles specific
211 to each BioProject, 3' ends of forward and reverse reads from each sequence were trimmed
212 (option truncLen in function *filterAndTrim*) at specific lengths (Dataset 2). Additionally, 5' ends of
213 both reads were trimmed 20bp to remove the primer part (option trimLeft in function
214 *filterAndTrim*). To correct for the great variation in read number among samples and BioProjects,
215 we performed rarefaction curves to determine a conservative number of sequences to conserve
216 per sample. Accordingly, we randomly selected 3,000 sequences per sample, hence excluding
217 samples with fewer than 3,000 sequences (Dataset 2). We pooled together unique ASV
218 sequences found among all BioProjects after rarefaction (N=57,100 unique sequences) and
219 inferred their taxonomy in two steps. We first classified all ASVs to the genus level with the
220 SILVA 138.2 database [32] using the *assignTaxonomy* function (R package dada2). To verify
221 SILVA v 138.2 accuracy in predicting phyllosphere-associated Hyphomicrobiales ASVs, we
222 extracted 16S rRNA gene sequences from complete phyllosphere-associated Hyphomicrobiales
223 reference genomes (see section *Phylogenomic reconstruction of phyllosphere-associated*
224 *Hyphomicrobiales*) and classified them as we did for ASVs. We also performed a phylogenetic
225 analysis of 16S rRNA sequences extracted from reference genomes for a comparison with our
226 phylogenomic reconstruction. The Maximum-Likelihood tree was inferred with RAxML (1,000
227 bootstraps, GTRGAMMA). For the second step, we refined ASV classification restricted to
228 phyllosphere-associated Hyphomicrobiales sequences (i.e. identified as Beijerinckiaceae,
229 *Alsobacter* or *Phreatobacter* with SILVA v138.2 database) using the *assignTaxonomy* function

230 with 16s rRNA gene sequences we extracted from complete phyllosphere-associated
231 Hyphomicrobiales reference genomes as references. We subdivided Bioprojects from different
232 climatic regions, or with samples from different substrates, in different categories. We subdivided
233 Bioproject PRJNA100102 in three categories: Temperate forest, Subtropical forest and Tropical
234 forest [43]. We subdivided Bioproject PRJNA290145 in three categories: Bark samples, Lichen
235 samples and Moss Samples [4]. For each BioProject or category within BioProjects (n=17), we
236 estimated the diversity of bacteria taxon as the relative abundance of sequences assigned to
237 each taxon, focusing on Hyphomicrobiales dominant families and nested genera.
238
239 **Comparison of 16s rRNA and *rpoB*–based barcoding for the identification of**
240 **phyllosphere-associated Hyphomicrobiales taxa.**
241
242 To assess the performance of the *rpoB* gene as a phylogenetic and taxonomic marker for
243 phyllosphere-associated *Hyphomicrobiales*, we first retrieved complete *rpoB* gene sequences
244 from reference genomes and MAGs with known taxonomy. *rpoB* sequences were extracted from
245 RAST output files, aligned, and their phylogeny was assessed with RAxML as previously
246 described for core genes. To compare *rpoB*- and 16S rRNA gene–based taxonomic resolution,
247 we used amplicon data we previously generated [19], including 184 phyllosphere samples
248 collected from two temperate forests in Quebec, Canada: *Station Biologique des Laurentides*
249 (SBL, n = 99) and *Mont St-Hilaire* (MSH, n = 85). All samples were previously amplified with
250 primers optimized for Methylobacteriaceae and targeting a hypervariable region of the *rpoB*
251 gene [19]. We used *rpoB* ASV abundance tables and taxonomy from our previous study as a
252 starting point. We reclassified ASVs using *rpoB* sequences from genomes and MAGs as
253 references, excluding sequences that did not completely overlapped ASV sequences. Because
254 reference sequences only included phyllosphere-associated Hyphomicrobiales, we only
255 performed reclassification for ASVs previously identified as Rhizobiales (former

256 Hyphomicrobiales), hence excluding Caulobacterales, Sphingomonadales and Rhodospirillales.
257 Within Rhizobiales, we excluded ASVs previously assigned to other families
258 (Aurantimonadaceae, Bradyrhizobiaceae, Brucellaceae, Hyphomicrobiaceae,
259 Phyllobacteriaceae and Rhizobiaceae). Relative abundance per sample was thereafter
260 recalculated for major Hyphomicrobiales taxa. We evaluated the congruence in relative
261 abundance estimations between *rpoB*- and 16S rRNA gene–based barcoding. We compared the
262 relative abundance of major phyllosphere-associated Hyphomicrobiales genera in 41 leaf
263 samples for which metabarcoding was performed with both 16S rRNA and *rpoB* genes. To
264 evaluate the congruence, we used a linear model on relative abundances after a Hellinger
265 transformation to correct for rare taxa [13] and compared the observed estimate (slope) of the
266 model to its expected distribution by realizing 10,000 random permutations of relative
267 abundance values among leaf samples. Because relative abundances observed for a particular
268 taxa could be site-dependent in this dataset [19], we constrained permutations within sites (SBL
269 and MSH). We calculated p-values using the following formula: p=(b+1)/(m+1) where b was the
270 number of expected values higher than the observed value and m, the number of permutations
271 [31].

272 **RESULTS**

273 ***Phyllosphere-associated Hyphomicrobiales MAGs mostly belong to Methylobacteriaceae,***
274 ***Lichenihabitantaceae and two undescribed taxa***
275

276 Phyllosphere-associated Hyphomicrobiales (syn. Rhizobiales) MAGs were previously identified
277 in the phyllosphere of Canadian temperate forests (n=22; unpublished data; bioproject
278 PRJNA1291777), in mosses from the Canadian tundra (n=5; [17]), in lichens from a worldwide
279 survey (n=43; [37]), in glacier sediments (n=1; [9]) and in mining soils (n=2; [9,23]; Dataset 1).

280 We refined the taxonomic classification of phyllosphere-associated Hyphomicrobiales MAGs
281 using a phylogenomic-based approach. Combining the 73 MAGs and 97 reference
282 Hyphomicrobiales genomes covering the known phylogenomic diversity of phyllosphere-
283 associated Hyphomicrobiales [12] (Figure 1A; Figure S2; Dataset 1), we identified 10,726 unique
284 genes (hypothetical protein, repeat and mobile elements excluded). We considered 1,046 genes
285 that were present in at least 80% of reference genomes (MAGs excluded) as candidate core
286 genes. After excluding genes that were duplicated or missing data in more than 30% of
287 assemblies (genomes and MAGs combined), we identified 824 core genes (37 rna and 787
288 protein-encoding genes; Dataset 3) for which we could retrieve the DNA sequence in at least
289 70% of genomes and MAGs, and thus infer a phylogenetic tree (RAxML, 1,000 replicated tree,
290 GTRGAMMA model). Our consensus phyllosphere-associated Hyphomicrobiales coalescent
291 tree combining the 824 core gene phylogenetic trees (with ASTRAL-III) was largely congruent
292 with a previous phylogenomic-based classification of Hyphomicrobiales [12], except for the
293 genus *Salinarimonas*, previously classified in Salinarimonadaceae, but embedded within the
294 family Methylobacteriaceae according to our consensus tree. Our tree also provided a few
295 additions for the genera *Lichenifustis*, *Methylocella* and *Hohaiirga*, confirming they belong to the
296 families Lichenihabitantaceae, Beijerinckiaceae and Methylobacteriaceae, respectively. Our
297 consensus tree supported the classification of most phyllosphere-associated Hyphomicrobiales
298 MAGs into five monophyletic groups (branch support =1; Figure 1B, Figure S2). These groups
299 corresponded to the families Methylobacteriaceae (n=18), Lichenihabitantaceae (n=27),
300 Rhabdaerophilaceae (n=1) and to two undescribed taxa that we named according to the
301 previous GTDB-based classification of phyllosphere-associated Hyphomicrobiales MAGs
302 (unpublished data; bioproject PRJNA1291777); namely: *RH-AL1* (n=21) and *JAJXWB01* (n=3).
303 *RH-AL1* included only two reference genomes obtained from strains previously isolated from
304 coal slag [44]. According to our consensus tree, it was a sister but distinct phylogenomic group
305 from Beijerinckiaceae. *JAJXWB01* included no reference genome and was named according to

306 a MAG identified in mine drainage sediments (unpublished data). According to our consensus
307 tree, it was a sister but distinct phylogenomic group from Methylocystaceae and
308 Rhodoblastaceae. No phyllosphere-associated Hyphomicrobiales MAG from related
309 Hyphomicrobiales families could be identified (Alsobacteraceae, Boseaceae, Chelatococcaceae,
310 Roseiarcaceae, Rhodoblastaceae, Methylocystaceae and Beijerinckiaceae). In
311 Methylobacteriaceae, phyllosphere-associated Hyphomicrobiales MAGs from the genera
312 *Enterovirga* (n=7) and *Methylobacterium* (n=11) were identified, but not from *Salinarimonas*,
313 *Microvirga* or *Hohaiivirga*. In Lichenihabitantaceae, phyllosphere-associated Hyphomicrobiales
314 MAGs were previously all assigned to Lichenihabitans according to the GTDB-based
315 classification, but spanned all described genera according to our consensus tree, namely
316 *Lichenihabitans* (n=8), *Lichenifustis* (n=6) and *Lichenibacterium* (n=13).

317 **Limits and improvements in the classification of phyllosphere-associated**
318 **Hyphomicrobiales taxa with 16S rRNA metabarcoding**

319 A maximum-likelihood tree inferred from 24 16S rRNA sequences used to define *1174-901-12* in
320 the SILVA database, version 138.1, and identified as *Lichenibacterium* in version 138.2,
321 confirmed that this taxon corresponded to *Lichenibacterium* and to *Lichenifustis* (Figure S1). In
322 the tree, *1174-901-12* formed a monophyletic and strongly supported group with 16S rRNA
323 genera *Lichenibacterium* and *Lichenifustis*, but not with their related counterpart *Lichenihabitans*.

324 We evaluated the power of the 16S rRNA gene as a marker to classify phyllosphere-associated
325 Hyphomicrobiales taxa. We extracted 92 complete 16S rRNA sequences from reference
326 phyllosphere-associated Hyphomicrobiales genomes (for which we validated the taxonomy
327 using a phylogenomic approach; Figure 1A) that could be further used as a reference to refine
328 phyllosphere-associated Hyphomicrobiales ASV taxonomy. Due to incomplete assemblies, no
329 complete 16S rRNA gene sequence could be retrieved from most MAGs, but was recovered for

330 JAJXWB01. A maximum-likelihood tree built from complete 16S rRNA gene sequences shows
331 the limited power of this gene as a marker to infer the evolutionary history and taxonomy of
332 phyllosphere-associated Hyphomicrobiales taxa (Figure 2A). In this tree, the families
333 Methylobacteriaceae and Chelatococcaceae were not monophyletic, but deeply intermingled
334 together. Within Beijerinckiaceae, the genera *Methylocystis* and *Methylocapsa* were not
335 mophophyletic. According to the 16S rRNA gene tree, RH_AL1 and JAJXWB01 deeply
336 branched within families Beijerinckiaceae and Methylocystaceae, respectively (Figure 1A), while
337 these two undescribed groups formed sister but distinct clades from the two families in our
338 phylogenomic tree (Figure 2A). We classified the complete 16S rRNA gene sequences using the
339 SILVA 138.2 database and *assignTaxonomy* R function (Dataset 4). All phyllosphere-associated
340 Hyphomicrobiales families and relatives (Boseaceae, Methylobacteriaceae, Chelatococcaceae,
341 Roseiarcaceae, Rhodoblastaceae, Methylocystaceae, Beijerinckiaceae and
342 Lichenihabitantaceae) were wrongly grouped by SILVA 138.2 in the Beijerinckaceae family,
343 except for the two most basal families Alsobacteraceae and Rhabdaerophillaceae, which were
344 identified as "Hyphomicrobiales *Incertae Sedis*" (Dataset 4). Most phyllosphere-associated
345 Hyphomicrobiales genera were correctly assigned by SILVA 138.2, with few exceptions, mostly
346 regarding genera for which taxonomy was updated after the database was released. RH_AL1
347 was classified as *Methylocella* and JAJXWB01 as *Methylocystis*. *Lichenifustis* and
348 *Lichenihabitans* were both identified as *Lichenibacterium*. *Hohaiivirga* was not identified at the
349 genus level and some *Methylobacterium* sequences were identified as *Methylorubrum* (Dataset
350 4).

351 We further explored the worldwide distribution and abundance of phyllosphere-associated
352 Hyphomicrobiales taxa in the phyllosphere and related environments, using metabarcoding data
353 from 13 studies based on the 16S rRNA gene (Dataset 2; Figure 2B). We identified 57,100
354 unique amplicon sequence variants (ASV) across 1998 samples. We first used SILVA 138.2 as

355 a reference database to classify those ASVs at the order level. We identified 4,637 ASVs
356 classified as Hyphomicrobiales, among which 3,089 were classified in phyllosphere-associated
357 taxa, mostly Beijerinckiaceae (n=3,073). Because of SILVA databases limitations (Dataset 4),
358 we used 92 complete 16 rRNA gene sequences extracted from complete phyllosphere-
359 associated Hyphomicrobiales genomes (n=91) and MAG (n=1) as a database to reclassify the
360 3,089 ASVs at the family and genus level, according to the phylogenomic consensus tree
361 (Figure 1A), with two exceptions. For family prediction, and according to the 16S rRNA gene
362 phylogeny (Figure 2A), we classified JAJXWB01 within Methylocystaceae and RH_AL1 within
363 Beijerinckiaceae (Dataset 4). Classification with SILVA, version 138.2 and with sequences
364 extracted from reference genomes resulted in the same genus for 52.22% of ASVs (Dataset 4).
365 Because most genus were classified within Beijerinckiaceae in SILVA database, we could not
366 directly compare both methods for family classification. We thus deduced families in SILVA
367 classification from the genus (e.g. by classifying *Methylobacterium* ASVs in Methylobacteriaceae
368 instead of Beijerinckiaceae) and found that it matched family classification with sequences
369 extracted from reference genomes for 81.57% of ASVs. Family could not be assigned with at
370 least one method for 17.03% of ASVs. Among 1,338 ASVs identified as *Lichenibacterium* with
371 SILVA, v 138.2 (43.31% of ASVs), we confirmed that 1,213 belonged to Lichenihabitantaceae
372 using sequences from reference genomes as a database, and refined their classification in
373 *Lichenibacterium* (701 ASVs), *Lichenifustis* (312), *Lichenihabitans* (121) and unknown
374 Lichenihabitantaceae (79; Dataset 4).

375 ***Lichenihabitantaceae and Methylobacteriaceae are dominant* phyllosphere-associated**
376 **Hyphomicrobiales *taxa worldwide according to 16S rRNA metabarcoding***

377 Worldwide, Pseudomonadota was the most abundant bacteria phylum in the phyllosphere of
378 vascular, root-forming plants (excluding mosses and epiphytic plants), representing on average
379 56.1% of sequences identified as bacteria, or bacteria diversity [range: 25.1-87.1%] among

380 Bioproject (Dataset 5; Figure 2C) and mostly consisted of Alphaproteobacteria (42.3 [16.7-78.7]
381 % of Bacteria diversity). On average, Hyphomicrobiales dominated Alphaproteobacteria diversity
382 in the phyllosphere (21.6 [6.8-40.6]% of total bacteria diversity) followed by Sphingomonadales
383 (10.6 [2.2-21.0]%) and Acetobacterales (6.8[0.6-15.9%]). Hyphomicrobiales diversity in the
384 phyllosphere was dominated by Lichenihabitantaceae (10.0 [2.0-19.7]% of total bacteria
385 diversity) and by Methylobacteriaceae (5.8[0.7-11.8]% of total bacteria diversity). Other
386 Hyphomicrobiales families, including Rhizobiaceae (2.2 [0.1-14.0]% of total bacteria abundance)
387 and Beijerinckiaceae+RH_AL1 (1.7 [0.0-5.0]% of total bacteria diversity), could be occasionally
388 detected at the surface of tree leaves, but barely Methylocystaceae+JAJXWB01 (0.3 [0.0-0.9]%
389 of total bacteria diversity). In the phyllosphere, ASVs identified as Methylobacteriaceae mostly
390 corresponded to *Methylobacterium* (5.3 [0.6-11.0]% of total bacteria diversity) and ASVs
391 identified as Lichenihabitantaceae mostly corresponded to *Lichenibacterium* (8.7 [2.0-17.5]% of
392 total bacteria diversity) and *Lichenifustis* (1.1[0.0-3.7]%). ASVs classified as RH_AL1 were
393 barely observed (0.4[0.0-2.0]% of total bacteria diversity) and JAJXWB01 was almost never
394 detected (< 0.1%). On average, Hyphomicrobiales represented a smaller proportion of bacterial
395 diversity in lichen habitats (8.1[7.8-8.4]% of total bacteria diversity) and moss habitats (2.9[2.4-
396 3.5]% of total bacteria diversity) compared to the phyllosphere of vascular plants (21.6 [6.8-
397 40.6]% of total bacteria diversity). In lichen, the Hyphomicrobiales were still dominated by
398 Lichenihabitantaceae lichen (3.6[2.8-4.4]% of total bacteria diversity). Notable proportions of
399 Lichenihabitantaceae were also detected in snow surface (10.6% of total bacteria abundance),
400 the surface of tree bark (1.6%), in epiphytic plants (1.9%), and in aerosols (1.0%).
401 Methylobacteriaceae was one of the dominant bacterial families detected at the surface of
402 photovoltaic panels (30.2% of total bacteria abundance) and was also detected in aerosols
403 (3.4%), on epiphyte plants (2.1%) and on snow surface (1.6%).

404 **Comparison of 16s rRNA and *rpoB*–based barcoding for the identification of**
405 **phyllosphere-associated Hyphomicrobiales taxa.**
406
407 We explored the potential of the *rpoB* gene, which we previously adapted to metabarcoding to
408 evaluate the diversity of *Methylobacterium* in the phyllosphere, as a broader taxonomic marker
409 for Hyphomicrobiales [19]. We extracted 102 complete *rpoB* sequences from reference
410 phyllosphere-associated Hyphomicrobiales genomes (n=95) and MAGs (n=7), which could be
411 further used as a reference to refine the taxonomy of phyllosphere-associated Hyphomicrobiales
412 ASVs. A maximum-likelihood phylogenetic tree built from complete *rpoB* gene nucleotide
413 sequences shows that this gene is a reliable marker to infer the evolutionary history and
414 taxonomy of phyllosphere-associated Hyphomicrobiales taxa (Figure S2A). In this tree, all
415 families predicted by the phylogenomic-based consensus tree (Figure 1A) were monophyletic
416 (branch support > 0.90). Within phyllosphere-associated Hyphomicrobiales families
417 (Lichenihabitantaceae, Methylobacteriaceae), all genera predicted by the phylogenomic-based
418 consensus tree (Figure 1A) were monophyletic (branch support > 0.81).
419
420 We evaluated the abundance of Hyphomicrobiales taxa from *rpoB* metabarcoding previously
421 developed to target Methylobacteriaceae and performed on 184 leaf samples from two
422 temperate forests in Quebec [19]. ASVs were mostly assigned to Hyphomicrobiales (average
423 per sample; 95.6%; range per forest: 93.8-97.7%) and Caulobacterales (3.9[1.7-5.8]%; Dataset
424 6). We refined previous taxonomic assignments only for Hyphomicrobiales ASVs, using *rpoB*
425 sequences extracted from reference genomes and MAGs as a reference. In Hyphomicrobiales,
426 ASVs were mostly classified in three families: Methylobacteriaceae (24.6[12.2-39.0%]) -
427 dominated by *Methylobacterium* (23.4[11.1-37.7]%), RH_AL1 (19.8[13.1-25.5]%) and
428 Lichenihabitantaceae (36.7[31.1-41.5]%) - dominated by *Lichenibacterium* (31.9[23.0-39.6]%)
429 and *Lichenifustis* (4.3[1.1-8.0]%). *Enterovirga* (Methylobacteriaceae) and JAJXWB01 could only

430 be detected at very low abundances (0.2[0.1-0.3]% and 0.1[0.0-0.3]%, respectively) (Figure
431 S2B). We compared the relative abundance of Phyllosphere-associated Hyphomicrobiales
432 genera in 41 leaf samples for which metabarcoding was performed with both 16S rRNA and
433 *rpoB* genes, namely *Methylobacterium*, *Enterovirga*, *Lichenifustis*, *Lichenibacterium,* RH_AL1
434 and JAJXWB01 (Figure S3C). For these taxa, we observed a significant correlation between
435 relative abundances estimated by 16S rRNA gene and *rpoB* gene barcoding (R^2= 0.529,
436 p<0.001; Figure S3C). This correlation was not significantly different than expected after 10,000
437 permutations of relative abundances among leaf samples, within or among sites ($p > 0.05$;
438 Figure S3D), indicating that it was mostly explained by the contrast between highly frequent and
439 rare taxa (*Lichenibacterium* and *Methylobacterium* vs *Lichenifustis* and *Enterovirga*), consistently
440 observed among leaf samples, and by differences between forests (SBL vs MSH).
441

442 **DISCUSSION**

443 To clarify the distribution and composition of phyllosphere-associated Hyphomicrobiales, we
444 examined over 1,900 phyllosphere 16S rRNA gene datasets from 13 studies spanning multiple
445 continents. This meta-analysis confirmed that *Hyphomicrobiales* constitute a significant fraction
446 of phyllosphere bacterial communities, with the families *Methylobacteriaceae* and
447 *Lichenihabitantaceae* being dominant across environments. Notably, these groups also co-
448 occurred in related aerial habitats such as lichen and moss surfaces, snow, and aerosols,
449 suggesting they may form part of a larger aeroterrestrial biome. The identification of
450 Lichenihabitantaceae, which was solely isolated and characterized in lichen [29], as a dominant
451 taxon of the phyllosphere of trees, has profound implications for understanding microbial
452 colonization dynamics in both the phyllosphere and extreme environments. The ecological
453 distribution of Lichenihabitantaceae on exposed surfaces, leaves and potentially in snow,
454 suggests its resilience to environmental stresses, marking it as a pioneer organism capable of

455 colonizing newly exposed surfaces. This resilience aligns with potential pioneering capacities of
456 *Lichenibacterium*, enabling it to establish communities on surfaces exposed to UV and nutrient
457 scarcity [29]. For instance, metagenomics from lichens confirmed the abundance of
458 *Lichenibacterium* and suggested that these bacteria may play a key role in providing essential
459 nutrients to their fungal symbiotic partners in harsh environments [1]. Furthermore,
460 *Lichenibacterium*'s dominance as an early leaf colonizer implies that it could facilitate
461 subsequent microbial community development, potentially influencing plant health and
462 ecosystem functioning.

463 Our analyses also highlighted the limitations of the 16S rRNA gene for resolving phyllosphere-
464 associated Hyphomicrobiales diversity. Key families such as *Methylobacteriaceae* appeared
465 polyphyletic based on 16S sequences, and clade-level patterns were often ambiguous or
466 inconsistent across classification schemes, illustrating the low resolution of the 16S rRNA gene
467 in inferring taxonomy and evolutionary history in some bacterial taxa [2,7,14]. We also
468 determined that Lichenihabitantaceae, especially the genera *Lichenifustis* and *Lichenibacterium*,
469 corresponded to *1174-901-12*, an undescribed Hyphomicrobiales taxon frequently associated
470 with the phyllosphere [3,15,19,22,35,43,45]. Despite its abundance, the potential ecological role
471 of *1174-901-12* on the surface of plant leaves remains poorly understood, likely due to the lack
472 of data about this taxon in many reference databases and limitations of marker gene resolution.
473 These issues can lead to overgeneralization or misinterpretation of community structure and
474 ecological patterns in environmental studies.

475 To overcome these limitations, we applied a genome-resolved phylogenomic approach using
476 core single-copy genes to analyze a representative set of 73 phyllosphere-associated
477 Hyphomicrobiales MAGs alongside 97 reference *Hyphomicrobiales* genomes. This strategy
478 provided a clearer taxonomic framework, revealing that the majority of phyllosphere-associated
479 Hyphomicrobiales MAGs clustered into four robust clades: *Methylobacteriaceae*,

480 *Lichenihabitantaceae*, and two previously undescribed but well-supported lineages—RH-AL1
481 and JAJXWB01.

482 RH-AL1 is represented by two reference isolates from coal slag and is related to
483 *Beijerinckiaceae* [44], but forms a distinct and phylogenetically diverse lineage, including MAGs
484 from the phyllosphere, lichens and from mosses. JAJXWB01 appears completely uncultured and
485 forms a related but distinct lineage from *Methylocystaceae*, including MAGs form lichen and from
486 the phyllosphere. Interestingly, RH-AL1 reference isolates, like many members of the
487 Beijerinckiaceae and Methylocystaceae families, are methylotrophic bacteria, a trait shared with
488 *Methylobacterium* and allowing leaf-associated bacteria to use methanol released by plant as
489 carbon source [38]. However, no MAG-assembled genome was placed within Methylocystaceae
490 or Beijerinckiaceae, suggesting a relatively narrow and structured set of phyllosphere-associated
491 Hyphomicrobiales lineages adapted to phyllosphere environments. A similar narrow ecological
492 distribution was observed for Methylobacteriaceae, with phyllosphere-associated
493 Hyphomicrobiales MAGs only identified within *Enterovirga* and *Methylobacterium*, but not in
494 other genera. *Enterovirga* was traditionally associated with animal feces or aerosols [11,21] , but
495 its detection in the phyllosphere and in lichens suggest a larger ecological niche. In
496 *Methylobacterium*, MAGs were only detected in traditionally leaf-associated phylogroups A, B
497 and D, but not in the most ancestral group C, confirming secondary leaf-association during the
498 evolution of this genus [2,20]. The occurrence of phyllosphere-associated traits should be further
499 investigated in phyllosphere-associated Hyphomicrobiales genera, for instance by screening the
500 occurrence of methylotrophic pathways in phyllosphere-associated Hyphomicrobiales genomes
501 and MAGs. The ability for these bacteria to effectively metabolize one-carbon compounds like
502 methanol should be next validated by culturomics and phenotypic screening, in particular within
503 taxa traditionally not associated with the phyllosphere like Lichenihabitantaceae and *Enterovirga*,
504 or poorly represented in genomic databases and strain collections like RH_AL1 and JAJXWB01.

505 To complement the phylogenomic analysis of phyllosphere-associated Hyphomicrobiales, we
506 evaluated the utility of the alternative marker gene *rpo*B that we previously used to improve
507 resolution in *Methylobacterium* diversity characterization in the phyllosphere [19]. Unlike the 16S
508 rRNA gene, *rpoB*-based phylogenies recovered phyllosphere-associated Hyphomicrobiales
509 family- and genus-level relationships with much greater accuracy and congruence with genome-
510 based placements. In particular, using *rpoB* amplicon sequencing revealed the strong
511 prevalence of *Methylobacterium* and *Lichenibacterium*, supporting their dominance inferred from
512 metagenomic data. Furthermore, the RH-AL1 clade, which was underrepresented in 16S
513 datasets, accounted for a large fraction of *Hyphomicrobiales* reads in *rpoB*-based profiles, of
514 roughly the same abundance as *Methylobacterium* and *Lichenibacterium*, highlighting the rpoB
515 marker's superior sensitivity and specificity for uncovering cryptic diversity in phyllosphere-
516 associated Hyphomicrobiales.

517 Our study provides a revised and integrative view of phyllosphere-associated *Hyphomicrobiales*,
518 revealing both the global dominance of known families such as *Methylobacteriaceae* and
519 *Lichenihabitantaceae*, and the existence of undescribed but recurrent clades. We show that
520 relying solely on the 16S rRNA gene obscures key phylogenetic and ecological patterns of
521 bacterial communities in the phyllosphere, particularly due to its limited resolution and lack of
522 data in reference databases. By integrating *rpoB* markers and genome-resolved phylogenetics,
523 we propose a more robust framework to track phyllosphere-associated Hyphomicrobiales
524 diversity and guide future ecological and evolutionary studies. Looking forward, the cultivation
525 and characterization of the novel RH-AL1 and JAJXWB01 lineages, as well as a better
526 delineation of the genera within *Lichenihabitantaceae*, are important research priorities.
527 Combining long-read metagenomics, transcriptomics, and targeted functional assays will further
528 allow us to link genome evolution to ecological adaptation, helping to uncover the roles these
529 widespread Hyphomicrobiales play in the phyllosphere and beyond.

Funding information – JBL and SWK were supported by NSERC Discovery Grants RGPIN-2024-05825 and RGPIN-2025-05608 respectively.

Acknowledgments: NIngeborg Klarenberg and Wisnu Adi Wicaksono for providing Metadata; Marta Alonso Garcia and Zihui Wang for discussion.

Data availability: Code and data, including Datasets 1-6, are publicly available on Github repository: JBLED/Hyphomicrobiales-phyllosphere

539 **Figure 1: Phyllosphere-associated Hyphomicrobiales MAGs mostly belong to**

540 **Methylobacteriaceae, Lichenihabitantaceae and to RH_AL1, an undescribed taxon related**

541 **to Beijerinckiaceae.** A) Phylogenomic tree of 96 reference genomes (NCBI ref_seq genomes)

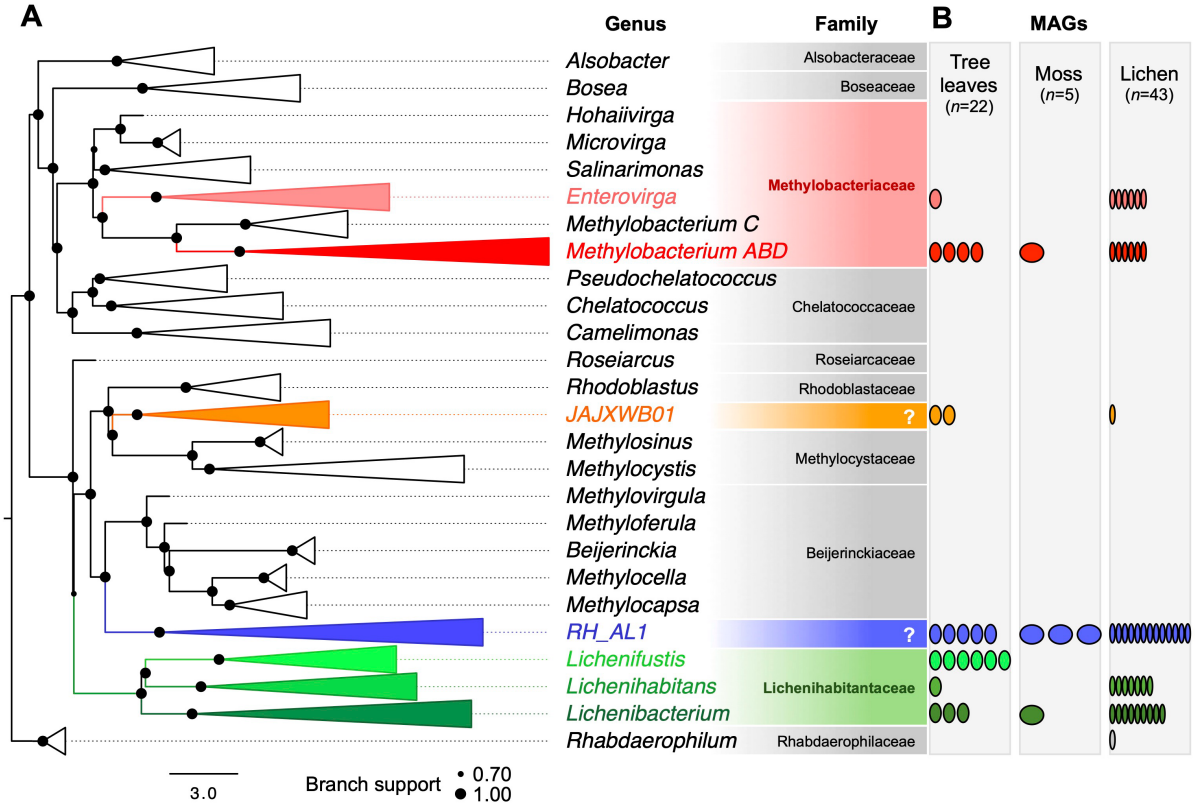
542 and 73 MAGs (tree leaves, moss and lichen). The consensus evolutionary tree was inferred from

543 824 core gene phylogenies combined with ASTRAL-III (See Figure S2 for detailed tree). The

544 tree was rooted on *Rhabdaerophilum*. Scale represents coalescent time. Branch support

545 represents local posterior probability. B) Distribution of tree leaf-, moss- and lichen-associated

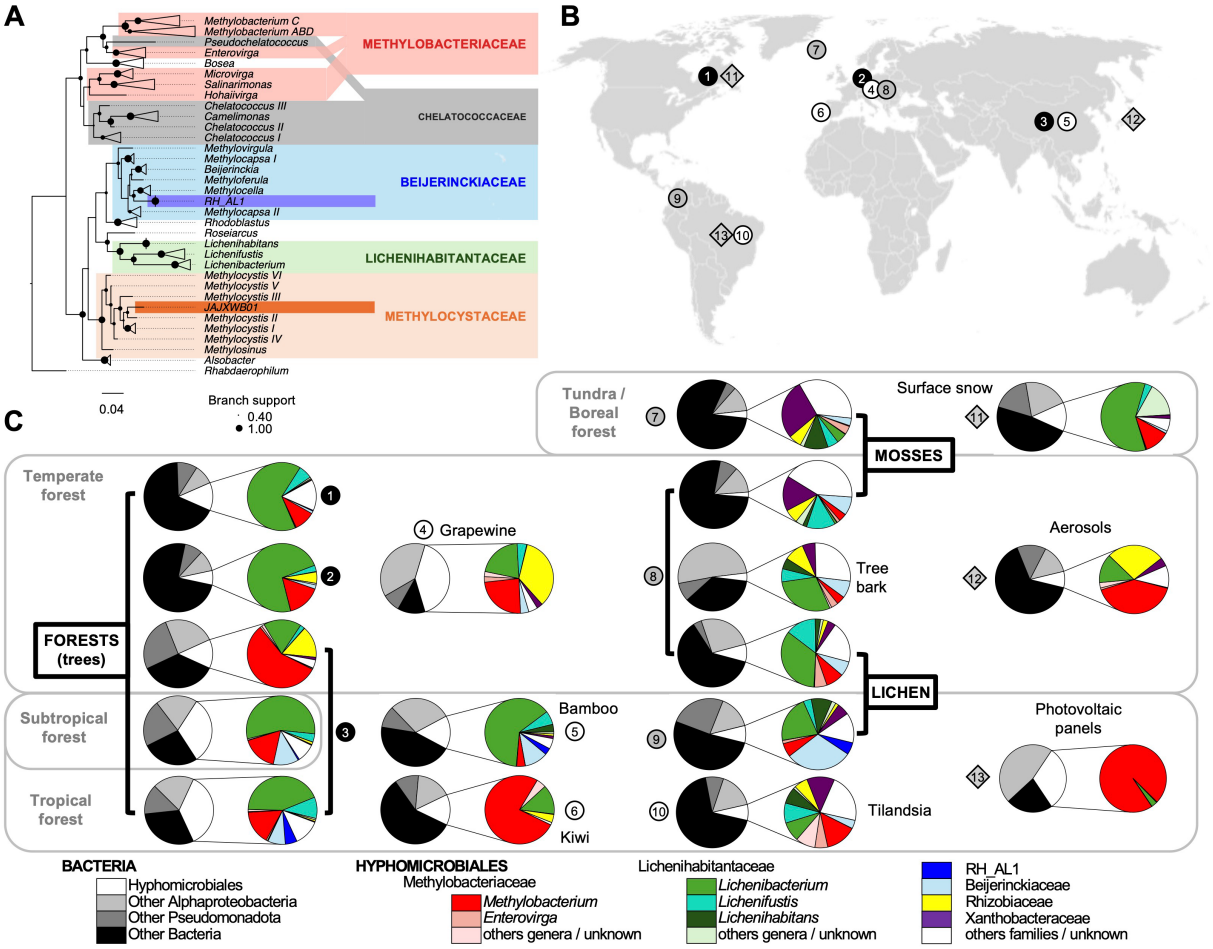
546 MAGs per Hyphomicrobiales genus. Each oval represents a MAG.



547

548

549 **Figure 2: Phyllosphere is dominated worldwide by Methylobacteriaceae and**
550 **Lichenihabitantaceae according to 16S rRNA gene metabarcoding.** A) Maximum-Likelyhood
551 tree (1,000 bootstrap, GTRGAMMA model) from complete 16S rRNA gene sequences extracted
552 from 96 phyllosphere-associated Hyphomicrobiales reference genomes (Fig 1). RH_AL1 is
553 embedded within Beijerinckiaceae and JAJXWBA01 in Methylocystaceae. Methylobacteriaceae,
554 Chelatococcaceae as well as some genera (*Chelatococcus*, *Methylocystis* and *Methylocapsa*)
555 appear polyphyletic. B) World map with BioProject origin (see **Dataset 2** for details). C) Bacterial
556 abundance according to 16S rRNA barcoding, focusing on phyllosphere-associated
557 Hyphomicrobiales families, in 13 BioProjects. Taxonomy from SILVA, refined with sequences
558 from phyllosphere-associated Hyphomicrobiales reference genomes (see Fig. 1). Bioprojects
559 were divided according to different climatic areas or substrates when applicable. For each
560 category, abundances are shown for main taxonomic subdivisions other than Hyphomicrobiales
561 (other Bacteria, Pseudomonadota and Alphaproteobacteria; left pie charts) and detailed for
562 major families within Hyphomicrobiales and for all genera within Methylobacteriaceae and
563 Lichenihabitantaceae (taxa with maximum relative abundance lower than 1% not showed,
564 including Methylocystaceae and JAJXWBA01; right pie charts).


A
Methylobacterium C
Methylobacterium ABD
Pseudochelatococcus
Enterovirga
Bosea
Microvirga
Salinarimonas
Hohaivirga
Chelatococcus III
Camelimonas
Chelatococcus II
Chelatococcus I
Methylovirgula
Methylocapsa I
Beijerinckia
Methyloferula
Methylocella
RH_AL1
Methylocapsa II
Rhodoblastus
Roseiarcus
Lichenihabitans
Lichenifustis
Lichenibacterium
Methylocystis VI
Methylocystis V
Methylocystis III
JAJXVB01
Methylocystis II
Methylocystis I
Methylocystis IV
Methylosinus
Alsobacter
Rhabdaerophilum
METHYLOBACTERIACEAE
CHELATOCOCCACEAE
BEIJERINCKIACEAE
LICHENIHABITANTACEAE
METHYLOCYSTACEAE
0.04
Branch support
0.40
1.00
B
C
Tundra / Boreal forest
Surface snow
MOSSES
Temperate forest
Grapewine
Tree bark
Aerosols
FORESTS (trees)
LICHEN
Subtropical forest
Bamboo
Photovoltaic panels
Tropical forest
Kiwi
Tilandsia
BACTERIA
Hyphomicrobiales
Other Alphaproteobacteria
Other Pseudomonadota
Other Bacteria
HYPHOMICROBIALES
Methylobacteriaceae
Methylobacterium
Enterovirga
others genera / unknown
Lichenihabitantaceae
Lichenibacterium
Lichenifustis
Lichenihabitans
others genera / unknown
RH_AL1
Beijerinckiaceae
Rhizobiaceae
Xanthobacteraceae
others families / unknown

565

566

567 **References**

568 [1] Ahmad, N., Ritz, M., Calchera, A., Otte, J., Schmitt, I., Brueck, T., Mehlmer, N. (2023)
569 Biosynthetic Potential of Hypogymnia Holobionts: Insights into Secondary Metabolite Pathways.
570 Journal of Fungi 9(5), 546, Doi: 10.3390/jof9050546.

571 [2] Alleman, A.B., Stolyar, S., Marx, C.J., Leducq, J.-B. (2025) Led astray by 16S rRNA:
572 phylogenomics reaffirms the monophyly of Methylobacterium and lack of support for
573 Methylorubrum as a genus. ISME J 19(1), wraf011, Doi: 10.1093/ismejo/wraf011.

574 [3] Ares, A., Pereira, J., Garcia, E., Costa, J., Tiago, I. (2021) The Leaf Bacterial Microbiota
575 of Female and Male Kiwifruit Plants in Distinct Seasons: Assessing the Impact of Pseudomonas
576 syringae pv. actinidiae. Phytobiomes Journal 5(3), 275–87, Doi: 10.1094/PBIOMES-09-20-0070-
577 R.

578 [4] Aschenbrenner, I.A., Cernava, T., Erlacher, A., Berg, G., Grube, M. (2017) Differential
579 sharing and distinct co-occurrence networks among spatially close bacterial microbiota of bark,
580 mosses and lichens. Molecular Ecology 26(10), 2826–38, Doi: 10.1111/mec.14070.

581 [5] Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K.,
582 Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L.,
583 Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C.,
584 Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O. (2008) The RAST Server: Rapid
585 Annotations using Subsystems Technology. BMC Genomics 9(1), 75, Doi: 10.1186/1471-2164-
586 9-75.

587 [6] Bai, Y., Müller, D.B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., Dombrowski,
588 N., Münch, P.C., Spaepen, S., Remus-Emsermann, M., Hüttel, B., McHardy, A.C., Vorholt, J.A.,
589 Schulze-Lefert, P. (2015) Functional overlap of the Arabidopsis leaf and root microbiota. Nature
590 528(7582), 364–9, Doi: 10.1038/nature16192.

591 [7] Bartoš, O., Chmel, M., Swierczková, I. (2024) The overlooked evolutionary dynamics of
592 16S rRNA revises its role as the “gold standard” for bacterial species identification. Sci Rep

593 14(1), 9067, Doi: 10.1038/s41598-024-59667-3.
594 [8] Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C., Hochreiter, S. (2015) msa: an R
595 package for multiple sequence alignment. Bioinformatics 31(24), 3997–9, Doi:
596 10.1093/bioinformatics/btv494.
597 [9] Busi, S.B., Bourquin, M., Fodelianakis, S., Michoud, G., Kohler, T.J., Peter, H.,
598 Pramateftaki, P., Styllas, M., Tolosano, M., De Staercke, V., Schön, M., de Nies, L., Marasco,
599 R., Daffonchio, D., Ezzat, L., Wilmes, P., Battin, T.J. (2022) Genomic and metabolic adaptations
600 of biofilms to ecological windows of opportunity in glacier-fed streams. Nat Commun 13(1), 2168,
601 Doi: 10.1038/s41467-022-29914-0.
602 [10] Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P.
603 (2016) DADA2: High resolution sample inference from Illumina amplicon data. Nat Methods
604 13(7), 581–3, Doi: 10.1038/nmeth.3869.
605 [11] Chen, X., Li, Q.-Y., Li, G.-D., Lei, H., Jiang, Y., Han, L., Huang, X.-S., Jiang, C.-L. (2017)
606 Enterovirga rhinocerotis gen. nov., sp. nov., isolated from Rhinoceros unicornis faeces. Antonie
607 Van Leeuwenhoek 110(4), 553–62, Doi: 10.1007/s10482-016-0823-1.
608 [12] diCenzo, G.C., Yang, Y., Young, J.P.W., Kuzmanović, N. (2024) Refining the taxonomy
609 of the order Hyphomicrobiales (Rhizobiales) based on whole genome comparisons of over 130
610 type strains. Int J Syst Evol Microbiol 74(4), 006328, Doi: 10.1099/ijsem.0.006328.
611 [13] Dixon, P. (2003) VEGAN, a package of R functions for community ecology. Journal of
612 Vegetation Science 14(6), 927–30, Doi: 10.1111/j.1654-1103.2003.tb02228.x.
613 [14] Hassler, H.B., Probert, B., Moore, C., Lawson, E., Jackson, R.W., Russell, B.T.,
614 Richards, V.P. (2022) Phylogenies of the 16S rRNA gene and its hypervariable regions lack
615 concordance with core genome phylogenies. Microbiome 10(1), 104, Doi: 10.1186/s40168-022-
616 01295-y.
617 [15] Herrmann, M., Geesink, P., Richter, R., Küsel, K. (2021) Canopy Position Has a Stronger
618 Effect than Tree Species Identity on Phyllosphere Bacterial Diversity in a Floodplain Hardwood

619 Forest. Microb Ecol 81(1), 157–68, Doi: 10.1007/s00248-020-01565-y.
620 [16] Hördt, A., López, M.G., Meier-Kolthoff, J.P., Schleuning, M., Weinhold, L.-M., Tindall,
621 B.J., Gronow, S., Kyrpides, N.C., Woyke, T., Göker, M. (2020) Analysis of 1,000+ Type-Strain
622 Genomes Substantially Improves Taxonomic Classification of Alphaproteobacteria. Frontiers in
623 Microbiology 11, 468, Doi: 10.3389/fmicb.2020.00468.
624 [17] Ishak, S., Rondeau-Leclaire, J., Faticov, M., Roy, S., Laforest-Lapointe, I. (2024) Boreal
625 moss-microbe interactions are revealed through metagenome assembly of novel bacterial
626 species. Sci Rep 14(1), 22168, Doi: 10.1038/s41598-024-73045-z.
627 [18] Klarenberg, I.J., Keuschnig, C., Russi Colmenares, A.J., Warshan, D., Jungblut, A.D.,
628 Jónsdóttir, I.S., Vilhelmsson, O. (2022) Long-term warming effects on the microbiome and nifH
629 gene abundance of a common moss species in sub-Arctic tundra. New Phytologist 234(6),
630 2044–56, Doi: 10.1111/nph.17837.
631 [19] Leducq, J.-B., Seyer-Lamontagne, É., Condrain-Morel, D., Bourret, G., Sneddon, D.,
632 Foster, J.A., Marx, C.J., Sullivan, J.M., Shapiro, B.J., Kembel, S.W. (2022) Fine-Scale
633 Adaptations to Environmental Variation and Growth Strategies Drive Phyllosphere
634 Methylobacterium Diversity. mBio 13(1), e03175-21, Doi: 10.1128/mbio.03175-21.
635 [20] Leducq, J.-B., Sneddon, D., Santos, M., Condrain-Morel, D., Bourret, G., Martinez-
636 Gomez, N.C., Lee, J.A., Foster, J.A., Stolyar, S., Shapiro, B.J., Kembel, S.W., Sullivan, J.M.,
637 Marx, C.J. (2022) Comprehensive Phylogenomics of Methylobacterium Reveals Four
638 Evolutionary Distinct Groups and Underappreciated Phyllosphere Diversity. Genome Biology
639 and Evolution 14(8), evac123, Doi: 10.1093/gbe/evac123.
640 [21] Lee, H., Chaudhary, D.K., Kim, D.-U. (2024) Enterovirga aerilata sp. nov. and Knoellia
641 koreensis sp. nov., isolated from an automobile air conditioning system. International Journal of
642 Systematic and Evolutionary Microbiology 74(5), 006383, Doi: 10.1099/ijsem.0.006383.
643 [22] Long, J., Luo, W., Xie, J., Yuan, Y., Wang, J., Kang, L., Li, Y., Zhang, Z., Hong, M. (2021)
644 Environmental Factors Influencing Phyllosphere Bacterial Communities in Giant Pandas' Staple

645 Food Bamboos. Front. Microbiol. 12, Doi: 10.3389/fmicb.2021.748141.
646 [23] Ma, B., Lu, C., Wang, Y., Yu, J., Zhao, K., Xue, R., Ren, H., Lv, X., Pan, R., Zhang, J.,
647 Zhu, Y., Xu, J. (2023) A genomic catalogue of soil microbiomes boosts mining of biodiversity and
648 genetic resources. Nat Commun 14(1), 7318, Doi: 10.1038/s41467-023-43000-z.
649 [24] Maki, T., Hosaka, K., Lee, K.C., Kawabata, Y., Kajino, M., Uto, M., Kita, K., Igarashi, Y.
650 (2023) Vertical distribution of airborne microorganisms over forest environments: A potential
651 source of ice-nucleating bioaerosols. Atmospheric Environment 302, 119726, Doi:
652 10.1016/j.atmosenv.2023.119726.
653 [25] Moura, J.B., Delforno, T.P., do Prado, P.F., Duarte, I.C. (2021) Extremophilic taxa
654 predominate in a microbial community of photovoltaic panels in a tropical region. FEMS
655 Microbiology Letters 368(16), fnab105, Doi: 10.1093/femsle/fnab105.
656 [26] Noh, H.-J., Baek, K., Hwang, C.Y., Shin, S.C., Hong, S.G., Lee, Y.M. (2019)
657 Lichenihabitans psoromatis gen. nov., sp. nov., a member of a novel lineage
658 (Lichenihabitantaceae fam. nov.) within the order of Rhizobiales isolated from Antarctic lichen.
659 Int J Syst Evol Microbiol 69(12), 3837–42, Doi: 10.1099/ijsem.0.003695.
660 [27] Ogier, J.-C., Pagès, S., Galan, M., Barret, M., Gaudriault, S. (2019) rpoB, a promising
661 marker for analyzing the diversity of bacterial communities by amplicon sequencing. BMC
662 Microbiol 19(1), 171, Doi: 10.1186/s12866-019-1546-z.
663 [28] Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. Science
664 276(5313), 734–40, Doi: 10.1126/science.276.5313.734.
665 [29] Pankratov, T.A., Grouzdev, D.S., Patutina, E.O., Kolganova, T.V., Suzina, N.E.,
666 Berestovskaya, J.J. (2020) Lichenibacterium ramalinae gen. nov, sp. nov., Lichenibacterium
667 minor sp. nov., the first endophytic, beta-carotene producing bacterial representatives from
668 lichen thalli and the proposal of the new family Lichenibacteriaceae within the order Rhizobiales.
669 Antonie van Leeuwenhoek 113(4), 477–89, Doi: 10.1007/s10482-019-01357-6.
670 [30] Pankratov, T.A., Samylina, O.S., Tikhonova, E.N., Ianutsevich, E.A., Avtukh, A.N., Lee,

671 Y.M. (2023) A novel bacteriobiont of the Arctic lichen Flavocetraria nivalis, Lichenifustis
672 flavocetrariae gen. nov, sp. nov. demonstrating hydrolytic properties and containing a full set of
673 the Calvin-Benson-Bassham cycle genes. Arch Microbiol 205(6), 232, Doi: 10.1007/s00203-023-
674 03577-x.
675 [31] Phipson, B., Smyth, G.K. (2010) Permutation p-values should never be zero: calculating
676 exact p-values when permutations are randomly drawn. Statistical Applications in Genetics and
677 Molecular Biology 9(1), Doi: 10.2202/1544-6115.1585.
678 [32] Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J.,
679 Glöckner, F.O. (2013) The SILVA ribosomal RNA gene database project: improved data
680 processing and web-based tools. Nucleic Acids Res 41(D1), D590–6, Doi: 10.1093/nar/gks1219.
681 [33] Romani, M., Carrion, C., Fernandez, F., Intertaglia, L., Pecqueur, D., Lebaron, P., Lami,
682 R. (2019) High bacterial diversity in pioneer biofilms colonizing ceramic roof tiles. International
683 Biodeterioration & Biodegradation 144, 104745, Doi: 10.1016/j.ibiod.2019.104745.
684 [34] Sierra, M.A., Danko, D.C., Sandoval, T.A., Pishchany, G., Moncada, B., Kolter, R.,
685 Mason, C.E., Zambrano, M.M. (2020) The Microbiomes of Seven Lichen Genera Reveal Host
686 Specificity, a Reduced Core Community and Potential as Source of Antimicrobials. Front.
687 Microbiol. 11, Doi: 10.3389/fmicb.2020.00398.
688 [35] Siqueira, J.S., de Carvalho, L.A.L., Santos, C.H.B., Frezarin, E.T., Pinheiro, D.G.,
689 Nicodemo, D., Desoignies, N., Rigobelo, E.C. (2024) Influence of Growth Support on the
690 Diversity, Composition, and Functionality of Microbial Communities Associated with Tillandsia
691 recurvata. Microb Ecol 87(1), 129, Doi: 10.1007/s00248-024-02448-2.
692 [36] Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-
693 analysis of large phylogenies. Bioinformatics 30(9), 1312–3, Doi: 10.1093/bioinformatics/btu033.
694 [37] Tagirdzhanova, G., Saary, P., Cameron, E.S., Allen, C.C.G., Garber, A.I., Escandón,
695 D.D., Cook, A.T., Goyette, S., Nogerius, V.T., Passo, A., Mayrhofer, H., Holien, H., Tønsberg, T.,
696 Stein, L.Y., Finn, R.D., Spribille, T. (2024) Microbial occurrence and symbiont detection in a

global sample of lichen metagenomes. PLoS Biol 22(11), e3002862, Doi: 10.1371/journal.pbio.3002862.

[38] Tani, A., Masuda, S., Fujitani, Y., Iga, T., Haruna, Y., Kikuchi, S., Shuaile, W., Lv, H., Katayama, S., Yurimoto, H., Sakai, Y., Kato, J. (2023) Metabolism-linked methylotaxis sensors responsible for plant colonization in Methylobacterium aquaticum strain 22A. Front. Microbiol. 14, Doi: 10.3389/fmicb.2023.1258452.

[39] Touchette, D., Gostinčar, C., Whyte, L.G., Altshuler, I. (2023) Lichen-associated microbial members are prevalent in the snow microbiome of a sub-arctic alpine tundra. FEMS Microbiology Ecology 99(12), fiad151, Doi: 10.1093/femsec/fiad151.

[40] Vorholt, J.A. (2012) Microbial life in the phyllosphere. Nature Reviews Microbiology 10(12), 828–40, Doi: 10.1038/nrmicro2910.

[41] Wang, H.-C., Huang, M.-H., Guo, D.-Y., He, W., Wang, L., Fu, Z.-Y., Li, W.-J., Zhang, A.H., Zhang, D.-F. (2024) Hohaiivirga grylli gen. nov., sp. nov., a New Member of the Family Methylobacteriaceae, Isolated from Cricket (Gryllus chinensis). Curr Microbiol 81(11), 392, Doi: 10.1007/s00284-024-03922-3.

[42] Wang, S., Meade, A., Lam, H.-M., Luo, H. (2020) Evolutionary Timeline and Genomic Plasticity Underlying the Lifestyle Diversity in *Rhizobiales*. mSystems 5(4), e00438-20, /msystems/5/4/msys.00438-20.atom, Doi: 10.1128/mSystems.00438-20.

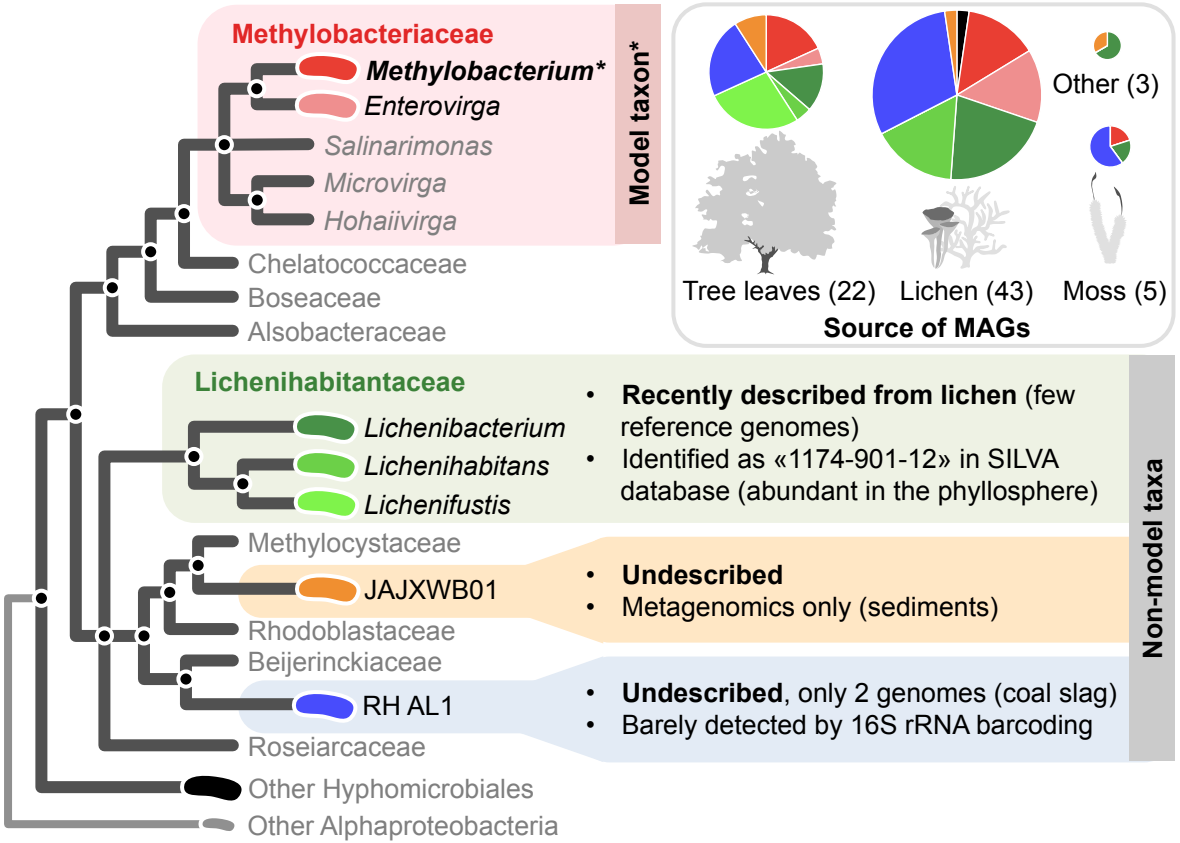
[43] Wang, Z., Jiang, Y., Zhang, M., Chu, C., Chen, Y., Fang, S., Jin, G., Jiang, M., Lian, J.-Y., Li, Y., Liu, Y., Ma, K., Mi, X., Qiao, X., Wang, X., Wang, X., Xu, H., Ye, W., Zhu, L., Zhu, Y., He, F., Kembel, S.W. (2023) Diversity and biogeography of plant phyllosphere bacteria are governed by latitude-dependent mechanisms. New Phytologist 240(4), 1534–47, Doi: 10.1111/nph.19235.

[44] Wegner, C.-E., Gorniak, L., Riedel, S., Westermann, M., Küsel, K. (2019) Lanthanide-Dependent Methylotrophs of the Family Beijerinckiaceae: Physiological and Genomic Insights. Appl Environ Microbiol 86(1), e01830-19, Doi: 10.1128/AEM.01830-19.

723 [45] Wicaksono, W.A., Morauf, C., Müller, H., Abdelfattah, A., Donat, C., Berg, G. (2023) The
724 mature phyllosphere microbiome of grapevine is associated with resistance against Plasmopara
725 viticola. Front. Microbiol. 14, Doi: 10.3389/fmicb.2023.1149307.
726 [46] Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S. (2018) ASTRAL-III: polynomial time
727 species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19(6), 153,
728 Doi: 10.1186/s12859-018-2129-y.
729 [47] Zhao, Q., Liu, W., Li, Y., Ke, M., Qu, Q., Yuan, W., Pan, X., Qian, H. (2020)
730 Enantioselective effects of imazethapyr residues on *Arabidopsis thaliana* metabolic profile and
731 phyllosphere microbial communities. Journal of Environmental Sciences 93, 57–65, Doi:
732 10.1016/j.jes.2020.03.009.
733

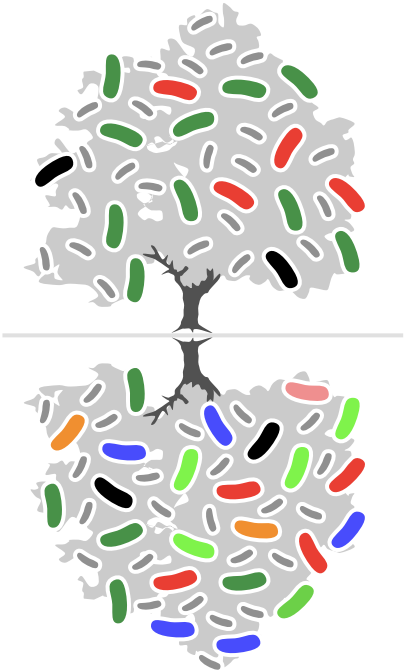
Phylogenomic resolution of phyllosphere-associated Hyphomicrobiales

824 core genes from 97 genomes + 73 metagenome-assembled genomes (MAGs)



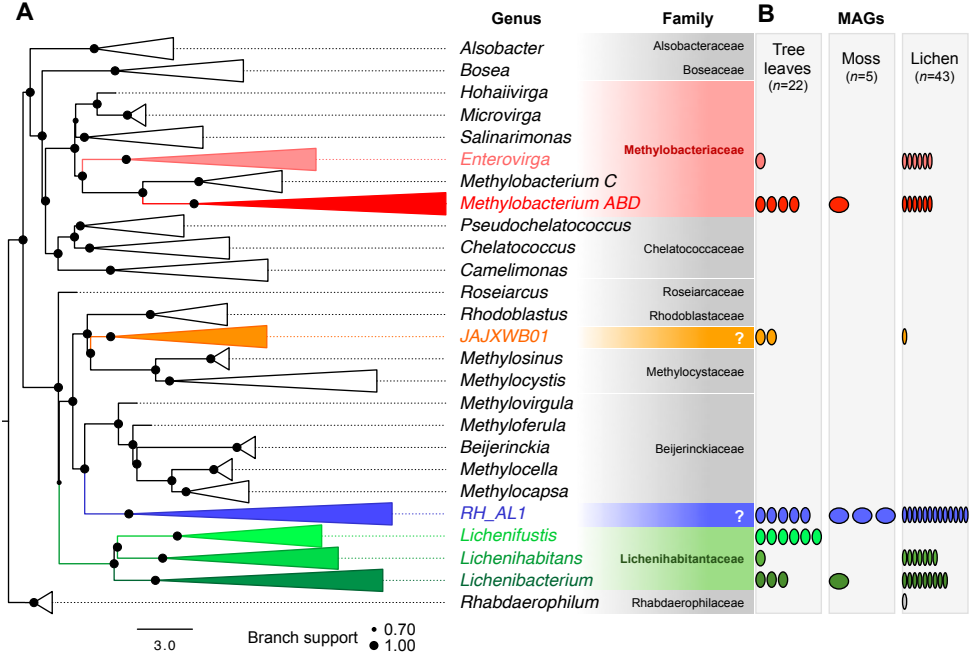
The diversity we capture

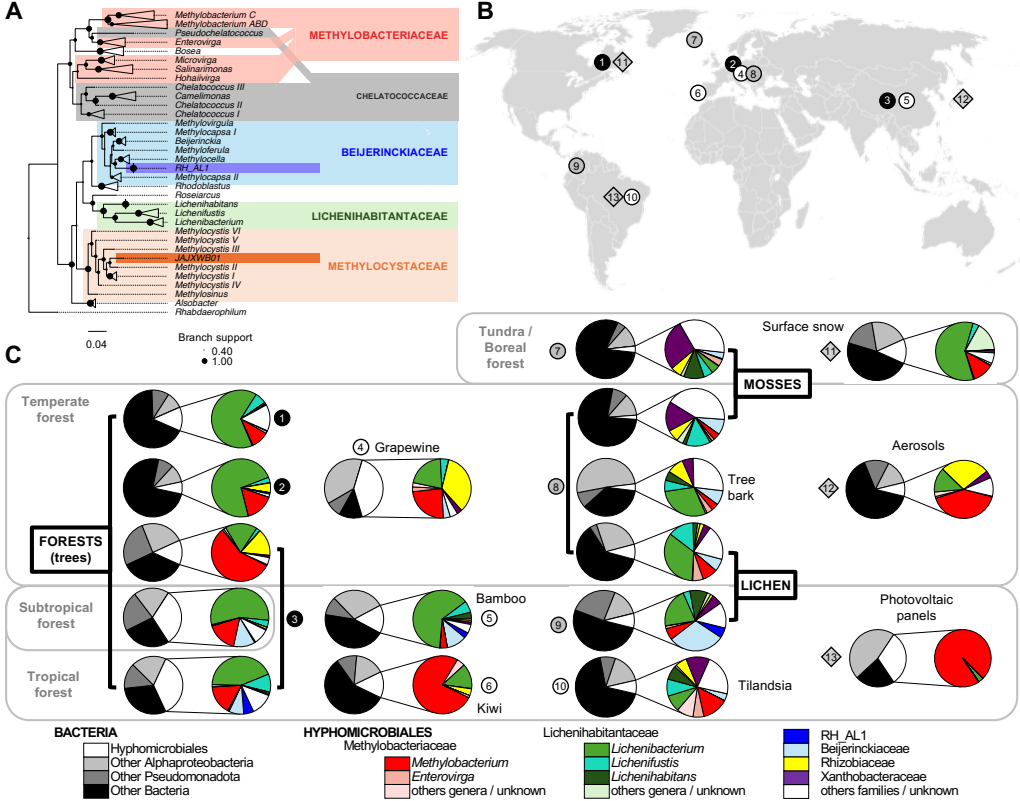
16S rRNA gene barcoding
+ current databases (SILVA)
(13 BioProjects; >1900 samples)



The diversity we expect

metagenomics + refined databases
(culturomics + phylogenomics)


A
Genus
Family
B
MAGs
Tree leaves (n=22)
Moss (n=5)
Lichen (n=43)
Alsobacter
Bosea
Hohaiivirga
Microvirga
Salinarimonas
Enterovirga
Methylobacterium C
Methylobacterium ABD
Pseudochelatococcus
Chelatococcus
Camelimonas
Roseiarcus
Rhodoblastus
JAJXWB01
Methylosinus
Methylocystis
Methylovirgula
Methyloferula
Beijerinckia
Methylocella
Methylocapsa
RH_AL1
Lichenifustis
Lichenihabitans
Lichenibacterium
Rhabdaerophilum
Alsobacteraceae
Boseaceae
Methylobacteriaceae
Chelatococcaceae
Roseiarcaceae
Rhodoblastaceae
?
Methylocystaceae
Beijerinckiaceae
?
Lichenihabitantaceae
Rhabdaerophilaceae
3.0
Branch support
0.70
1.00


A
Methylobacterium C
Methylobacterium ABD
Pseudochelatococcus
Enteroviriga
Bosea
Microviriga
Salinarimonas
Hohaiiviriga
Chelatococcus III
Camelimonas
Chelatococcus II
Chelatococcus I
Methylovirgula
Methylocapsa I
Beijerinckia
Methyloferula
Methylocella
RH_AL1
Methylocapsa II
Rhodoblastus
Roseiarcus
Lichenihabitans
Lichenifustis
Lichenibacterium
Methylocystis VI
Methylocystis V
Methylocystis III
JAJXWB01
Methylocystis II
Methylocystis I
Methylocystis IV
Methylosinus
Alsobacter
Rhabdaerophilum
METHYLOBACTERIACEAE
CHELATOCOCCACEAE
BEIJERINCKIACEAE
LICHENIHABITANTACEAE
METHYLOCYSTACEAE
0.04
Branch support
0.40
1.00
B
C
Tundra / Boreal forest
Surface snow
MOSSES
Temperate forest
Grapewine
Tree bark
Aerosols
FORESTS (trees)
LICHEN
Subtropical forest
Bamboo
Photovoltaic panels
Tropical forest
Kiwi
Tilandsia
BACTERIA
Hyphomicrobiales
Other Alphaproteobacteria
Other Pseudomonadota
Other Bacteria
HYPHOMICROBIALES
Methylobacteriaceae
Methylobacterium
Enteroviriga
others genera / unknown
Lichenihabitantaceae
Lichenibacterium
Lichenifustis
Lichenihabitans
others genera / unknown
RH_AL1
Beijerinckiaceae
Rhizobiaceae
Xanthobacteraceae
others families / unknown