An aerobic link between lignin degradation and C_1 metabolism: growth on methoxylated aromatic compounds by members of the genus *Methylobacterium*

Jessica A. Lee[12345], Sergey Stolyar[123], and Christopher J. Marx[123]

Running title

Methylobacterium growth on methoxylated aromatics

Affiliations

[1]Department of Biological Sciences, University of Idaho, Moscow, ID
[2]Center for Modeling Complex Interactions, University of Idaho, Moscow, ID
[3]Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID
[4]Global Viral, San Francisco, CA
[5]Department of Biology, San Francisco State University, San Francisco, CA

Corresponding authors

Jessica A. Lee: jessica.audrey.lee@gmail.com
Christopher J. Marx: cmarx@uidaho.edu

Abstract

Microorganisms faces many barriers in the degradation of the polycyclic aromatic polymer lignin, one of which is an abundance of methoxy substituents. Demethoxylation of lignin-derived aromatic monomers in aerobic environments releases formaldehyde, a potent cellular toxin that organisms must eliminate in order to further degrade the aromatic ring. Here we provide the first comprehensive description of the ecology and evolution of the catabolism of methoxylated aromatics in the genus *Methylobacterium*, a plant-associated genus of methylotrophs capable of using formaldehyde for growth. Using comparative genomics, we found that the capacity for aromatic catabolism is ancestral to two clades, but has also been acquired horizontally by other members of the genus. Through laboratory growth assays, we demonstrated that several *Methylobacterium* strains can grow on *p*-hydroxybenzoate, protocatechuate, vanillate, and ferulate; furthermore, whereas non-methylotrophs excrete formaldehyde as a byproduct during growth on vanillate, *Methylobacterium* do not. Finally, we surveyed published metagenome data to find that vanillate-degrading *Methylobacterium* can be found in many soil and rhizosphere ecosystems but is disproportionately prominent in the phyllosphere, and the most highly represented clade in the environment (the root-nodulating species *M. nodulans*) is one with few cultured representatives.

Introduction

Microbial processes for degrading lignin and lignin-derived aromatics are of intense interest in a diversity of fields, ranging from bioenergy engineering—in which the recalcitrance of lignin is a major hurdle in the processing of plant biomass [1]—to the fossil fuel industry—which seeks to understand microbial transformations of coal [2]. Lignin, which comprises approximately 20% of the carbon fixed by photosynthesis on land [3], is an exceptionally complex, irregular, polycyclic aromatic polymer, in which many of the constituent aromatic rings are heavily substituted with methoxy ($-OCH_3$) groups [4].

In the aerobic microbial degradation of lignin-derived aromatic monomers such as vanillate, the degradation of the aromatic ring proceeds by the protocatechuate branch of the beta-ketoadipate pathway. The gene cluster encoding this pathway is widely distributed among soil microorganisms, and has a complex evolutionary history resulting in diverse patterns of gene organization and regulation [5–7]. In the case of methoxylated aromatics, ring cleavage must be preceded by the removal of the methoxy group; in most aerobic organisms, this occurs via vanillate monooxygenase, a Rieske [2Fe-2S] enzyme that demethylates vanillate to generate protocatechuate and formaldehyde [8–12] (Fig. 1). Formaldehyde is a potent electrophile, and toxic to microorganisms due to its reactivity with DNA and proteins [13]. Elimination of this toxin is therefore essential to lignin degradation. Multiple studies have demonstrated experimentally, through either engineered disruption or constitutive expression of formaldehyde oxidation capacity in *Bradyrhizobium diazoefficiens* [11], *Pseudomonas putida* [14], *Burkholderia cepacia* [9], and *Corynebacterium glutamicum* [15], that demethoxylation of vanillate is dependent on the activity of an functional formaldehyde detoxification system, and formaldehyde removal may be the rate-limiting step to the degradation of lignin-derived methoxylated aromatics.

Detoxification pathways in lignin-degrading organisms are diverse, and in some organisms include more than one mechanism. In some cases formaldehyde is oxidized

to CO_2 via formate, such as in the thiol-dependent systems in *C. glutamicum* and *B. japonicum* or the zinc-dependent dehydrogenase in *P. putida* [16]; in others it may be incorporated into biomass via the RuMP pathway, as in *B. cepacia* [17]. Furthermore, it has recently been discovered that methanogens of the genus *Methermicoccus* are capable of using the methoxy groups of coal-derived aromatic compounds for the production of methane through a novel metabolic pathway ("methoxydotrophic methanogenesis"), illuminating a previously-unknown link between coal degradation and methane production via C_1 metabolism [18].

Methylotrophs, organisms that use reduced C_1 compounds for growth, therefore make appealing candidates for efficient degradation of lignin-derived methyoxylated aromatics. Members of the genus *Methylobacterium*, and particularly the model organism *M. extorquens*, have long been studied for their metabolism of simple C_1 compounds such as methanol, formate, methylamine, and halogenated methanes [19]. Formaldehyde is a central intermediate in the metabolism of many of these substrates, raising the possibility that, if *Methylobacterium* were capable of demethoxylating aromatic compounds, they could also use the resulting formaldehyde as a growth substrate. It has previously been documented that two members of the genus are capable of growth on aromatic compounds [20–22]; however, the prevalence of this trait among *Methylobacterium* and the organization and evolutionary history of the genes involved has not been described. Furthermore, to our knowledge, it has not been shown whether these organisms are capable of methoxydotrophy: growth on methoxy groups from aromatic compounds.

Here we report novel findings on the ecology and evolution of methoxydotrophic growth and catabolism of aromatic compounds by members of the genus *Methylobacterium*. We explored the genomic capacity of *Methylobacterium* species by searching published genomes, verified growth and formaldehyde production/consumption on aromatic compounds in the laboratory, and surveyed published metagenome data to assess the distribution and prevalence of aromatic-degrading *Methylobacterium* in the environment. For a thorough description of aerobic microbial degradation of lignin-derived aromatic

compounds, we direct the reader to the excellent reviews published previously [5, 23]; a simplified diagram of the pathway and the four genes of interest to this study (*vanA*, *pcaG*, *pobA*, *ech*) is given in Figure 1. Because we were especially interested in organisms that could use the methoxy group of vanillate, our metagenomic analysis focused especially on *vanA* (KO:K03862), the gene encoding the alpha subunit of vanillate monooxygenase.

Materials and Methods

Phylogenetic analysis

Genetic potential of *Methylobacterium* species [24] (Table S1) for the degradation of aromatic compounds was assessed through analysis of genomes available on the IMG/M database [25]. We should note that Green and Ardley [26] recently proposed a restructuring of the genus *Methylobacterium*, including a renaming of the genus of the *M. extorquens* clade from *Methylobacterium* to *Methylorubrum*, and a suggestion to reclassify the *M. aquaticum* and *M. nodulans* clades into different genera entirely. However, because the present study is concerned with the four clades as a coherent group of evolutionarily related taxa that are still known as *Methylobacterium* by many research groups, and all are currently still listed within that genus by the List of Prokaryotic Names with Standing in Nomenclature [27], in this manuscript we continue to refer to all organisms as *Methylobacterium*.

Genes of interest were identified by KEGG Orthology ID: *vanA* (vanillate monooxygenase alpha subunit, KO:K03862); *pcaG* (protocatechuate 3,4-dioxygenase alpha subunit, KO:K00448); *pobA* (*p*-hydroxybenzoate 3-monooxygenase, KO:K00481); and *ech* (*trans*-feruloyl-CoA hydratase/vanillin synthase, KO:K18383; we searched also for IMG Term Object ID 8865, vanillin synthase). Known lignin-degrading organisms were added to the phylogenies as reference organisms, and *Prochlorococcus marinus* NATL2A and archaeal species were used as outgroups (Table S2). The genes used for the phylogeny in Figure 1 were 16S rRNA (the RNA component of the small ribosomal

subunit), *rpoB* (DNA-directed RNA polymerase subunit beta, KO:K03043), *atpD* (ATP Synthase F1 complex beta subunit, KO:K02112), and *recA* (recombination protein RecA, KO:K03553), all of which have been shown previously to be informative for phylogenetic analysis in *Methylobacterium* and closely related organisms [28–30].

All gene sequences were aligned using Clustal Omega v.1.2.3 [31], using four combined HMM/guide tree iterations and default settings for all other parameters. Phylogenetic trees were generated using IQ-Tree v.1.6.1 [32] on the CIPRES Science Gateway [33], with one partition, outgroups given in Table S2, ultrafast bootstrapping with 1000 replicates, and SH-like Single Branch test [34] with 1000 replicates; for all other parameters, the default settings were used.

Genomic context of aromatic catabolism genes

To assess synteny in the gene neighborhoods surrounding the *vanA* and *pcaH* genes in *Methylobacterium* species, we aligned scaffolds, or segments of assembled genomes, using MAUVE v.2.4.0 with the progressiveMauve algorithm and default parameter settings [35, 36]. Figures were generated using a GenoPlotR v.0.8.7 [37] for R v.3.4.3 [38] in RStudio v.1.1.423 [39].

GC content in the *Methylobacterium sp*. AMS5 genome was calculated in R using the alphabetFrequency function from Biostrings v.2.46.0 [40]; for the plot in Fig. S1, a sliding window of 5 kb was used for the full genome and 500 bp for the catabolic island region. Genome signature difference was calculated using a custom script in R, using the formula for the δ* difference [41] with tetranucleotides rather than dinucleotides.

Growth assays on aromatic substrates

All organisms were grown on MPIPES medium [42] with the addition of 1x Wolfe's Vitamins [43], 25 μM $LaCl_3$ (as has been found to facilitate growth in some *Methylobacterium spp*. [44]), and carbon substrates as specified. All conditions contained 16 mM of carbon (4 mM succinate, 2.67 mM glucose, 2.29 mM benzoate, 2.29 mM PHBA, 2.29 mM PCA, 2 mM vanillate, 1.6 mM ferulate, 0.4 g/L Kraft lignin).

Methylobacterium sp. 4-46 was also tested on methanol without $LaCl_3$. For each assay, one colony was inoculated from a culture plate into 5 mL MPIPES with 3.5 mM disodium succinate and grown 24 hours to obtain a stationary-phase culture, then diluted 1:64, and grown again until stationary. Using that inoculum, the growth experiment was conducted as follows: each replicate was diluted 1:64 into MPIPES and aliquoted into several wells of a Costar 48-well tissue culture-treated plate (product #3548, Corning Inc., Corning, NY) for non-volatile substrates, or into Balch-type glass culture tubes for vanillin, along with carbon substrate provided at concentrations such that all conditions were equimolar in carbon (16 mM). Total volume was 640 µL per well in culture plates or 5 mL per tube in Balch tubes. Each strain was tested in biological triplicate (3 separate colonies). Balch tubes were sealed with butyl rubber serum stoppers to prevent escape of vanillin.

Plates were incubated at 30 °C in a LPX44 Plate Hotel (LiCONiC, Woburn, MA) with shaking at 250 RPM. Optical density was assessed using a Wallac 1420 Victor2 Microplate Reader (Perkin Elmer, Waltham, MA), reading OD_{600} for 0.4 seconds at intervals of between 2 and 5 hours. Culture tubes were incubated in a TC-7 tissue culture roller drum (New Brunswick Scientific, Edison, NJ) at a speed setting of 7, and OD_{600} was measured in the culture tubes using a Spectronic 200 spectrophotometer (Thermo Fisher Scientific, Waltham, MA). Outputs from the Wallac 1420 software were collated using Curve Fitter [42]; data cleaning, analysis, and plotting were then conducted using custom scripts in R (Supplemental Files S1, S2). Because many of the strains formed clumps, maximum OD_{600} was used as a semi-quantitative metric for growth. Non-growth was defined as no OD_{600} above 0.025.

Formaldehyde production during growth on vanillate

Growth of *Methylobacterium* strains on vanillate was initiated with stationary-phase MPIPES-succinate cultures as described above; these were diluted 1:64 into 5 mL MPIPES with 2 mM vanillate and grown 2 days until stationary phase. This inoculum was diluted 1:32 into culture flasks containing 20 mL of MPIPES with 2 mM vanillate and incubated, shaking, at 30 °C. Cultures were sampled every 4 hours: OD_{600} was

measured in a SmartSpec Plus Spectrophotometer (Bio-Rad Laboratories, Hercules, CA), and the supernatant of 500 µL of centrifuged culture was used for the measurement of formaldehyde using a colorimetric assay [45]. All strains were tested in biological triplicate.

Formaldehyde production of non-methylotrophs during vanillate growth was assayed similarly. Stationary-phase vanillate-grown cultures of *Pseudomonas putida* KT2440 and *Rhodococcus jostii* RHA1 were diluted 1:64 into 5 mL MPIPES medium in culture tubes, with sampling every 2 hours. Vanillate was provided at 4 mM; although this was greater than in the *Methylobacterium* experiment, we have found that vanillate concentration has little effect on the resulting concentration of formaldehyde in the medium (Fig. S3).

Analysis of *Methylobacterium vanA* genes in published metagenomes

To assess the abundance of *Methylobacterium*-encoded *vanA* genes in the environment, we searched all published metagenomes available through the IMG/M portal on December 20, 2017. The "Function Search" tool was used to identify all metagenomes containing genes annotated with KO:K03862 (*vanA*). To increase our chances of observing meaningful ecological patterns, we restricted our analysis metagenomes containing >100 *vanA* genes. Genes on scaffolds with a phylogenetic lineage assignment of *Methylobacterium* were designated "*Methylobacterium vanA.*" and the ecological context was taken from the Ecosystem Type field in the sample metadata associated with each gene. Because it was unclear whether all researchers used consistent definitions, we combined phylloplane with phyllosphere, and rhizoplane with rhizosphere.

To provide a tree-based assessment of *vanA* diversity in addition to the USEARCH-based IMG/M phylogenetic lineage assignment [46], we downloaded the *Methylobacterium vanA* gene sequences and placed them in the reference *vanA* phylogeny (described above). Pplacer v.1.1 [47] was used with HMMer v.3.1b.2 [48], RAxML v.8.2 [49], and taxtastic v.0.8.8 for Python 3. The resulting tree was paired with metadata from sequence scaffolds in Phyloseq v.1.22.3 for R [50], to generate the plot

in Fig. 8.

The relative abundances of all *vanA* or all *Methylobacterium* reads within each metagenome were found using the Statistical Analysis tool in IMG [25]. For each Ecosystem Type, the average and standard error among metagenomes was calculated manually, in order to omit datasets with "NA" values due to unassembled genes that impeded the Statistical Analysis tool. Because the *Methylobacterium vanA* genes analyzed here originated from only the subset of metagenomes having >100 *vanA* genes, the ratio of *Methylobacterium vanA* to all *vanA* was calculated using only the metagenomes analyzed in that part of the study.

Results

Genetic capacity for metabolism of methoxylated aromatic compounds is present primarily in two clades of *Methylobacterium*

We searched the published genomes of 26 *Methylobacterium* strains for the presence of *vanA*, *pcaG*, *pobA*, and *ech*. We found that all four genes are indeed present within the genus *Methylobacterium*. They are primarily limited to two closely related clades, the *aquaticum* clade and the *nodulans* clade; furthermore, they are present in nearly all organisms in these clades (Fig. 1). However, there are exceptions. One member each of the *extorquens* clade (*Methylobacterium sp.* AMS5) and the *radiotolerans* clade (*M. pseudosasiacola*) also carry some aromatic catabolism genes (Fig. 1). Two members of the *nodulans* clade (*sp.* 4-46 and *sp.* WSM2598) appear to lack *vanA,* though *Methylobacterium sp*. 4-46 carries *ech* (Fig. 1) and *vdh* (not shown), conferring the capacity to produce vanillate from ferulate, suggesting that the organism may have previously had the capacity to utilize the product of these reactions but has since lost *vanA*. Among the closest relatives to *Methylobacterium*, *Enterovirga rhinocerotis* has none of the four genes of interest, and *Microvirga flocculans* has only *pobA* and *pcaH*.

Taken together, these patterns suggest that the capacity to metabolize several lignin-derived aromatic compounds was acquired by the *aquaticum*/*nodulans* clade early in its divergence within the *Methylobacterium*, but members of other clades have acquired the genes more recently by horizontal gene transfer. To investigate this evolutionary story further, we analyzed the phylogenies of the genes themselves, and also examined the genomic contexts of *vanA* and *pcaG*.

Phylogeny of aromatic catabolism genes supports ancestral origins in *M. aquaticum* and *M. nodulans* clades and horizontal acquisition by two other species

The phylogenies of *pobA*, *pcaG*, and *vanA* were largely congruent with the phylogeny of conserved housekeeping genes within the *Methylobacterium* but not between *Methylobacterium* and their closest relatives, supporting the hypothesis that the four genes are ancestral to the *aquaticum*/*nodulans* clades but were not necessarily acquired together (Fig. 2). The *ech* phylogeny was the most difficult to interpret, as the *Methylobacterium* separated into three non-monophyletic clades (Fig. 2); it is possible that not all gene homologs in the phylogeny code for enzymes specific to ferulate. Notably, *Methylobacterium sp*. AMS5 and *M. pseudosasiacola* were more closely related to each other than to any other species in both their *pobA* and *pcaG* sequences, despite their phylogenetic distance from each other at the genome level. And while the *pobA* sequences of those two strains fell within those of other *Methylobacterium*, their *pcaG* sequences clustered with the *Sphingomonas*, further support not only for the hypothesis that these two disparate species both acquired the beta-ketoadipate pathway via horizontal gene transfer, but that they may have acquired it from the same donor or that one served as the donor to the other.

Gene synteny in *pcaG* and *vanA* neighborhoods is consistent with phylogeny

We further examined the relationships of *vanA* and *pcaG* across species by comparing patterns of synteny within the genome neighborhood for each. In all strains with *pcaG*, the gene was located within a cluster of beta-ketoadipate pathway genes, and five of these genes were present in all strains in the same order. Almost all strains also

286 contained a gene annotated as an uncharacterized conserved protein, DUF849
287 (COG3246, beta-keto acid cleavage enzyme), between *pcaG* and *pcaB* (Fig. 3). We
288 found no genes annotated as *pcaI* and *pcaJ* (beta-ketoadipate:succinyl-CoA
289 transferase) or *pcaF* (encoding beta-ketoadipyl-CoA thiolase) within these operons,
290 which are necessary for the final steps of the beta-ketoadipate pathway [5, 7]. All
291 *Methylobacterium* strains with *vanA* also had the other two genes required to confer the
292 ability for the demethoxylation of vanillate: *vanB* (encoding the other subunit of vanillate
293 monooxygenase) and *vanR* (an AraC family transcriptional regulator). Some strains also
294 had genes encoding a Major Facilitator Superfamily transporter specific to
295 protocatechuate (*pcaK*) or vanillate (*vanK*).
296
297 The genomic contexts of *pcaG* and *vanA* among strains exhibited patterns that agreed
298 overall with the phylogenies described above (Fig. 3). Closely related strains shared
299 common genes in the regions near *pcaG* and *vanA*, with differences among the strains
300 increasing with increasing distance from the genes of interest. Also consistent with
301 phylogeny, *Methylobacterium sp.* AMS5 and *M. pseudosasiacola* shared no
302 commonalities with the *M. aquaticum* clades or with each other in the neighborhoods
303 surrounding the *pca* genes or the *van* genes. Notably, in most previously described
304 organisms, the *pca* gene is not co-located with the other genes of interest in this study
305 (Tables S1, S2). However, the reaction product of the enzyme encoded by *pobA* (PCA)
306 is the substrate for the gene encoded by *pcaGH*; the proximity of the genes therefore
307 suggests that if these two organisms gained the *pca* genes via horizontal transfer, the
308 *pob* genes may have been transferred simultaneously as part of a catabolic island. We
309 found further support for the catabolic island hypothesis in the proximity of *vanA* and
310 *ech* to the *pca* genes in *Methylobacterium sp.* AMS5.
311
312 ***Methylobacterium sp*. AMS5 carries a catabolic island conferring the ability to**
313 **degrade lignin-derived aromatic compounds, in an *M. extorquens*-like genome**
314 The co-localization of all four aromatic catabolism genes of interest in the genome of
315 *Methylobacterium sp*. AMS5 prompted us to examine the region more closely. AMS5
316 was originally isolated in 2011 from the stem of a hypernodulating strain of soybean [51]

but is very closely related model organism *M. extorquens* PA1 [52, 53] (Fig. 1). We aligned the two genomes, as well as that of another clade member, *M. zatmanii* PSBB041, at the region where the aromatic catabolism genes are located (Fig. 4). The comparison revealed that the region appears to be prone to the insertion and excision of mobile genetic elements.

At the locus where *M. zatmanii* carries a gene encoding a putative IS5-type transposase (possibly nonfunctional due to a frameshift with an internal stop codon), AMS5 contained an additional 22-kb region: a catabolic island containing genes that appear to confer the full ability to degrade ferulate, vanillate, vanillin, *p*-hydroxybenzoate, and protocatechuate, flanked by two different transposase genes from the MULE superfamily. On the 3' side of this region, the two strains share 11 genes (14 kb), many related to nucleotide sugar metabolism, that are also not present in PA1 (Fig. 4), or in any of the other *Methylobacterium* species surveyed in this study. The corresponding locus in the PA1 genome carries only a putative site-specific integrase/recombinase, which is present only in the other *M. extorquens* strains (AM1, PA1, DM4; and *M. chloromethanicum, M. populi,* which have previously been identified as *M. extorquens* strains [54]). The gene adjacent to the integrase encodes UDP-*N*-acetylglucosamine 2-epimerase (*wecB*); the sequence appears to be complete in *M. zatmanii* (1,137bp) but truncated in PA1 (missing 912 bp from the 3' end). In all three genomes, the region of variability lies immediately at the 3' end of the genes for tRNA-Pro and tRNA-Arg.

The presence of the truncated gene in PA1 suggests that the region of sugar metabolism genes may originally have been present but were excised, whereas the arrangement of genes in the catabolic island suggest it was acquired into a *M. zatmanii*-like genomic background. The origins of the catabolic island itself are uncertain; as shown above (Fig. 2), there appear to be diverse phylogenetic origins represented among the different genes within the pathway, and BLAST search for the entire region in the NCBI Nucleotide database found no other organisms containing all genes in the same order. We could detect no genes in the region relating to replication or conjugation as might be expected in an integrative and conjugative element (ICE) [55].

Furthermore, no significant difference in tetranucleotide composition or in GC content was found between the inserted region (GC% = 71.4) and the full genome (GC% = 71.1) (Fig. S1), indicating that the catabolic island has likely been present in the genome long enough for amelioration [56].

Genome content predicts ability to grow on aromatic compounds in most *Methylobacterium* strains

We next wanted to assess whether *Methylobacterium* strains carrying genes for the catabolism of methoxylated aromatic compounds could indeed use those compounds as growth substrates. We chose 8 species from the *M. aquaticum*/*nodulans* clades, shown in Figure 5 (and Table S1). All of these strains carry all four genes of interest except *Methylobacterium sp.* 4-46, which lacks *vanA* and so was predicted not to grow on vanillate or ferulate (Fig. 1). We conducted a series of growth experiments on defined mineral medium with substrates of each of the four gene products of interest, as well as succinate (a known growth substrate for all the strains tested), glucose (used by only some *Methylobacterium* species), benzoic acid (another aromatic acid with a degradation pathway separate from the beta-ketoadipate pathway [57]), vanillin, and Kraft lignin (a soluble form of polymeric lignin). All strains grew on succinate and none on lignin; growth on PHBA, PCA, vanillate, and ferulate was as expected for all strains given their genome content, with the exceptions that *M. variabile* did not grow on vanillate and *M. nodulans* did not grow on ferulate (Fig. 5, Fig. S2, Supplemental File S2). No growth was observed on vanillin at the concentration tested; however, this may be due in part to the toxicity of vanillin [58], as we have observed growth on vanillin by *M. nodulans* when using lower concentrations in previous experiments.

Non-methylotrophs excrete formaldehyde during vanillate growth, whereas *Methylobacterium* do not

We were especially interested to ask whether methylotrophs are different from non-methylotrophs in their ability to cope with the formaldehyde produced during vanillate metabolism. Although, as mentioned above, it is recognized that the formaldehyde released by vanillate monooxygenase is a burden to organisms growing on vanillate, to

our knowledge no attempts have been made to measure the dynamics of accumulation of the toxin during growth. We therefore measured formaldehyde in the medium during the growth of two well-studied non-methylotrophic lignin degraders, *Pseudomonas putida* KT2440 and *Rhodoccus jostii* RHA1: we assayed each strain on vanillate as well as on PCA as a control compound, as PCA is the direct product of vanillate demethylation and therefore involves the same metabolism except for the effect of formaldehyde. In both strains, when growth occurred on vanillate, formaldehyde accumulated in the medium concomitant with the increase in optical density, peaking at concentrations of 0.94 mM (Fig. 6A, Table S4). When cultures entered stationary phase (approximately 10 hours for *P. putida*, 15 hours for *R. jostii*), formaldehyde concentrations began to decrease, ultimately returning to below the detection limit. No formaldehyde was detected at any time during growth on PCA. Moreover, growth on PCA was faster: stationary phase was reached at 7 hours for *P. putida* and 10 hours for *R. jostii*. These results suggest that while both *P. putida* and *R. jostii* can oxidize formaldehyde, removal is slower than production.

We conducted a similar experiment on the five strains of *Methylobacterium* showing the greatest growth on vanillate in our laboratory conditions: *M. aquaticum*, *M. nodulans*, *M. tarhaniae*, *Methylobacterium sp*. AMS5, and *M. aquaticum* MA-22A. We incubated all cultures with vanillate for 33 hours, long enough for AMS5 to show marked growth and all other species to consume all the substrate and reach stationary phase. No detectable formaldehyde was produced at any time (Fig. 6B, Table S4). These results suggest that *Methylobacterium* species are able to consume the formaldehyde generated from vanillate demethylation as rapidly as it is produced, likely via the same pathways by which the formaldehyde generated from methanol is used for energy generation and biosynthesis.

Methylobacterium-derived _vanA_ reads in published metagenomes are predominantly from the _M. nodulans_ cluster, with highest relative abundance in the phyllosphere

Aside from the environments in which they were isolated (Table S1), there exists scant information on the ecological niches of the aromatic-degrading *Methylobacterium* clades—for instance, whether they are more likely than the other *Methylobacterium* to inhabit ecosystems rich in lignin, such as soil or the rhizosphere. We therefore sought to learn more about the prevalence and abundance of *vanA*-carrying *Methylobacterium* species in the environment by querying the publicly available metagenome datasets on the JGI IMG/M database. The distinctness of the *Methylobacterium* genes within the phylogeny of known *vanA* sequences (Fig. 2) makes it possible to deduce phylogeny from DNA sequence. *vanA* is present as a single copy in most genomes in which it is found.

Our study set comprised 1,651 metagenomes, with a total assembled gene count of 5.60×10^9; we retrieved 317,816 scaffolds carrying *vanA*, of which 348 had *Methylobacterium* as their IMG phylogenetic lineage assignment (a frequency of 0.11%) (Tables S5, S6, S7). However, our phylogenetic placement using pplacer found that only 182 of these genes actually fell within the *Methylobacterium* clades, whereas 31 clustered more closely with *Azospirillum halopraeferens*, and the remaining 135 were distributed among more distantly related organisms (Fig. S4). We therefore focused only on the genes identified by both IMG and pplacer as *Methylobacterium* for the remainder of our analyses.

In absolute numbers, *vanA* genes belonging to *Methylobacterium* were highest in samples from the soil (43% of *Methylobacterium vanA*) and rhizosphere/rhizoplane (40%), the ecosystems in which *vanA* is most abundant generally (Fig. 7). However, we were especially interested in environments where *Methylobacterium* are most prominent within the community of vanillate degraders (high relative abundance of *Methylobacterium vanA* genes among total *vanA*). We found that there were many ecosystem types in which *vanA* was found but *Methylobacterium vanA* was not, including engineered ecosystems and those associated with animal hosts (Table S5). Where *Methylobacterium* were found, they made up the largest proportion of *vanA* in phyllosphere/phylloplane samples (0.28±0.13% of *vanA* genes) (Fig. 7). Relative

abundances in the other ecosystem types were one order of magnitude lower (freshwater: 0.024±0.016%; rhizoplane/rhizosphere: 0.022±0.010%; peat moss: 0.017%; soil: 0.014±0.005%; sediment: 0.011±0.005%, where ± denotes standard error; only 1 peat moss genome contained *Methylobacterium vanA*). This distribution across ecosystem types was qualitatively similar to that of total *Methylobacterium* reads in all metagenomes (Fig. 7).

The metagenome sequences identified as *Methylobacterium vanA* fell into three phylogenetic clusters (Fig. 8). The majority (114 genes, 63%) were most closely related to *M. nodulans*; 17% (31 genes) clustered with *Methylobacterium sp*. AMS5, and 20% (37 genes) clustered with the remaining *vanA*-carrying species, in the *M. aquaticum* clade (Fig. 8). These three clades were distributed approximately evenly across all ecosystem types.

Discussion

We have provided the first comprehensive description of growth of *Methylobacterium* on lignin-derived methoxylated aromatic acids. *Methylobacterium* species with this capability can use vanillate as a sole carbon substrate without the transient formaldehyde accumulation in the environment that is observed with non-methylotrophic vanillate degraders. The genes encoding aromatic catabolism are found almost exclusively in all members of the *M. aquaticum* and *M. nodulans* clades, but acquisition by horizontal gene transfer has been observed in other *Methylobacterium* species. These findings shed new light on the ecology of the genus.

Although there are important implications for our understanding of the bacterial lignin-degrading community, we are far from proposing that *Methylobacterium* be classified as lignin degraders: the species we studied were not found to grow on Kraft lignin, nor is there evidence that *Methylobacterium* can produce the peroxidases or laccases necessary for lignin depolymerization. In the environment, these organisms may depend

upon aromatic acids released by the action of other lignin-degrading organisms, which have been found to be a prominent component of leaf litter leachate [59]. Alternatively, *Methylobacterium* may encounter these compounds primarily in plant root exudates [60]. In addition to acting as a growth substrate, aromatic acids play an important role in plant-microbe signaling: vanillate, ferulate, *p*-hydroxybenzoate, and protocatechuate all influence on the process and productivity of root nodulation by other members of the *Rhizobiales* [61, 62]. This is significant in light of the fact that the members of the *M. nodulans* clade were, as the species name suggests, isolated from root and stem nodules of their plant hosts [21, 63]; and that we found this clade to be the most abundant among the *Methylobacterium vanA* genes found in the environment. Furthermore, *Methylobacterium sp*. AMS5 was isolated in a study on soybean epiphytes that are particularly responsive to host nodulation phenotype [51]. It is likely that at least part of the importance of aromatic catabolism in the *Methylobacterium* is to facilitate the relationships of these plant-associated organisms with their hosts. A link between aromatic catabolism and plant-microbe symbioses could help to explain our finding that organisms from the *M. nodulans* and *Methylobacterium sp*. AMS5 clades are dominant among the *Methylobacterium vanA* sequences found in the environment, despite the fact that there are few genome-sequenced representatives.

Perhaps the most compelling results from this study are the new insights into the evolution of the genus *Methylobacterium*. One element is the absence of the *vanA* gene in *Methylobacterium sp*. 4-46 and *sp.* WSM298, the only two species in the *M. nodulans*/ *M. aquaticum* clades to lack the capacity for methoxydotrophic growth. These species do carry genes for transforming ferulate to vanillin (*ech*) and vanillin to vanillate (*vdh*), which may be remnants from predecessors that were able to metabolize vanillate but lost the capability. Notably, *Methylobacterium* sp. 4-46 is also one of very few *Methylobacterium* species reported to be unable to grow on methanol [64]; our lab has found that it can indeed use methanol as a growth substrate but only in the presence of $LaCl_3$ (Fig. S5, Table S8), suggesting the involvement of a XoxF-type methanol dehydrogenase [44], and possibly a different role for methanol oxidation in this organism's ecology. Given our hypothesis that methanol oxidation and vanillate

demethylation require the same pathways for metabolizing the formaldehyde produced, it is possible that the loss of *vanA* in *Methylobacterium* sp. 4-46 might be related to its different style of methylotrophy. We have also observed that this and several non-*Methylobacterium* lignin-degrading species possess the genes to use PCA (the aromatic product of vanillate demethylation) but not the methoxy group (Fig. 1). Yet we have found no species in which the reverse is true, although it would theoretically be possible for a methylotroph with *van* genes but no *pca* genes to carry out methoxydotrophic growth by utilizing only the methoxy group of vanillate.

The other unexpected evolutionary finding relates to the acquisition of aromatic catabolism genes by horizontal transfer in two *Methylobacterium* species from outside of the *M. nodulans*/*aquaticum* clades. The discovery of a catabolic island in *Methylobacterium sp.* AMS5 is itself not unusual; *Methylobacterium* species have long been recognized to carry an abundance of IS (insertion sequence) elements, and it has been postulated that the associated genome rearrangements and horizontal gene transfer associated are important mechanisms of evolution in the genus [65–67]. Relevant to the present study is the prior observation that diverse features of the genomic background—and not necessarily those predicted by phylogeny—influence whether a newly introduced set of genes are immediately useful to the recipient organism [68]. Are there particular features of the *M. pseudosasiacola* and *Methylobacterium sp.* AMS5 genomes that allowed them to acquire the capacity for the degradation of methoxylated aromatics when no other known members of their clades did, or is the maintenance of this genomic capability the result of selective pressure specific to their ecological niche? Further work on the ecology of AMS5 and the biology of aromatic catabolism in *Methylobacterium* is necessary to address these questions.

This study has benefited from the wealth of knowledge that already exists on microbial degradation pathways for the degradation of lignin-derived aromatics in other organisms, and on methylotrophic metabolism in *Methylobacterium*, to deduce the likely fate of the methoxy group from vanillate. We found that in most cases, the gene annotations in IMG enabled us to correctly predict the substrates each strain could grow

on (the two exceptions were both cases of no growth, and we cannot rule out the possibility that growth could occur under different conditions). However, we did find some novel features of the *Methylobacterium* pathways: almost all the strains we studied appear have no homologs of *pcaI*, *pcaJ, or pcaF* encoded within the *pca* gene cluster. It is possible that the functions of these genes are carried out by homologs located elsewhere in the genome, as has been found in some other organisms [6]. A second possibility is raised by a previous study that carried out enzymatic screening and active site modeling on the DUF849 family of genes [69]: several of the DUF849 genes found in these *Methylobacterium* gene clusters were classified as beta-keto acid cleavage enzymes (G4 BKACE) predicted to act on betaketoadipate, raising the possibility that they might carry out the *pcaI/pcaJ* function and thus constitute a novel variant of the already diverse family of beta-ketoadipate pathway configurations [6, 7].

Likewise, our work benefited from the abundance of metagenome data that are publicly available, as a way of achieving a broad overview of the ecology our organisms of interest. However, *Methylobacterium* likely make up ~0.01-0.2% of organisms carrying *vanA* in environments where vanillate degraders are found; this low abundance means that even searching the entire set of available data on IMG/M yielded only 182 gene sequences of interest, and the majority of metagenomes with *Methylobacterium vanA* genes contained only one such gene, making statistical analysis difficult. Furthermore, the very environment where *Methylobacterium* make up a greater part of the *vanA* population is that where total *vanA* reads are low and therefore poorly represented in our dataset: the phyllosphere. Taken together, these caveats mean that the broad-scale ecological patterns observed in this initial survey should be interpreted with caution. However, as the first such survey carried out, our work provides a dataset on which to base more targeted assays to obtain a deeper understanding of the diversity and distribution of aromatic-degrading *Methylobacterium* in the environment.

561 **Acknowledgments**

562

563 We are grateful to Nicholas Shevalier, Alyssa Baugh, and Tomislav Ticak for their
564 assistance with experiments and support during the scientific process. We thank
565 Armando MacDonald for technical assistance with measurements, Akio Tani for
566 providing the *M. aquaticum* MA-22A culture, Andrea Lubbe and Trent Northen for the *P.*
567 *putida* and *R. jostii* cultures, and José de la Torre for assistance with calculating
568 genome signatures. We are thankful to Dipti Nayak, Mete Yuksel, Elizabeth Winters,
569 Brittany Baker, and Tomislav Ticak for their valuable feedback on the manuscript. This
570 work was funded by grants from the US Department of Energy Genomic Science
571 Program in Systems Biology for Energy and Environment, award DE-SC0012627, and
572 from the US National Science Foundation Dimensions of Biodiversity Program, award
573 DEB-1831838.

574

References

1. Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, et al. Lignin valorization: Improving lignin processing in the biorefinery. *Science* 2014; **344**: 1246843.
2. Welte CU. A microbial route from coal to gas. *Science* 2016; **354**: 184–184.
3. Ruiz-Dueñas FJ, Martínez ÁT. Microbial degradation of lignin: how a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microb Biotechnol* 2009; **2**: 164–177.
4. Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. Lignin biosynthesis and structure. *Plant Physiology* 2010; **153**: 895–905.
5. Harwood CS, Parales RE. The beta-ketoadipate pathway and the biology of self-identity. *Annu Rev Microbiol* 1996; **50**: 553–590.
6. Parke D. Acquisition, reorganization, and merger of genes: novel management of the β-ketoadipate pathway in *Agrobacterium tumefaciens*. *FEMS Microbiol Lett* 1997; **146**: 3–12.
7. Buchan A, Neidle EL, Moran MA. Diverse organization of genes of the β-ketoadipate pathway in members of the marine *Roseobacter* lineage. *Appl Environ Microbiol* 2004; **70**: 1658–1668.
8. Brunel F, Davison J. Cloning and sequencing of *Pseudomonas* genes encoding vanillate demethylase. *J Bacteriol* 1988; **170**: 4924–4930.
9. Mitsui R, Kusano Y, Yurimoto H, Sakai Y, Kato N, Tanaka M. Formaldehyde fixation contributes to detoxification for growth of a nonmethylotroph, *Burkholderia cepacia* TM1, on vanillic acid. *Appl Environ Microbiol* 2003; **69**: 6128–6132.

10. Merkens H, Beckers G, Wirtz A, Burkovski A. Vanillate metabolism in *Corynebacterium glutamicum*. *Curr Microbiol* 2005; **51**: 59–65.
11. Sudtachat N, Ito N, Itakura M, Masuda S, Eda S, Mitsui H, et al. Aerobic vanillate degradation and C1 compound metabolism in *Bradyrhizobium japonicum*. *Appl Environ Microbiol* 2009; **75**: 5012–5017.
12. Chen H-P, Chow M, Liu C-C, Lau A, Liu J, Eltis LD. Vanillin catabolism in *Rhodococcus jostii* RHA1. *Appl Environ Microbiol* 2012; **78**: 586–588.
13. Chen NH, Djoko KY, Veyrier FJ, McEwan AG. Formaldehyde stress responses in bacterial pathogens. *Front Microbiol* 2016; **7**.
14. Hibi M, Sonoki T, Mori H. Functional coupling between vanillate-O-demethylase and formaldehyde detoxification pathway. *FEMS Microbiol Lett* 2005; **253**: 237–242.
15. Lessmeier L, Hoefener M, Wendisch VF. Formaldehyde degradation in *Corynebacterium glutamicum* involves acetaldehyde dehydrogenase and mycothiol-dependent formaldehyde dehydrogenase. *Microbiology* 2013; **159**: 2651–2662.
16. Roca A, Rodríguez-Herva JJ, Ramos JL. Redundancy of enzymes for formaldehyde detoxification in *Pseudomonas putida*. *J Bacteriol* 2009; **191**: 3367–3374.
17. Kato N, Yurimoto H, Thauer RK. The physiological role of the ribulose monophosphate pathway in bacteria and archaea. *Biosci Biotechnol Biochem* 2006; **70**: 10–21.

621 18. Mayumi D, Mochimaru H, Tamaki H, Yamamoto K, Yoshioka H, Suzuki Y, et al.
622 Methane production from coal by a single methanogen. *Science* 2016; **354**: 222–
623 225.
624 19. Chistoserdova L. Modularity of methylotrophy, revisited. *Environmental*
625 *Microbiology* 2011; **13**: 2603–2622.
626 20. Ito H, Iizuka H. Taxonomic studies on a radio-resistant *Pseudomonas*. *Agric Biol*
627 *Chem* 1971; **35**: 1566–1571.
628 21. Jourand P. *Methylobacterium nodulans* sp. nov., for a group of aerobic,
629 facultatively methylotrophic, legume root-nodule-forming and nitrogen-fixing
630 bacteria. *Int J Syst Evol Biol* 2004; **54**: 2269–2273.
631 22. Ventorino V, Sannino F, Piccolo A, Cafaro V, Carotenuto R, Pepe O.
632 *Methylobacterium populi* VP2: plant growth-promoting bacterium isolated from a
633 highly polluted environment for polycyclic aromatic hydrocarbon (PAH)
634 biodegradation. *Sci World J* 2014; **2014**.
635 23. Masai E, Katayama Y, Fukuda M. Genetic and biochemical investigations on
636 bacterial catabolic pathways for lignin-derived aromatic compounds. *Biosci*
637 *Biotechnol Biochem* 2007; **71**: 1–15.
638 24. Kelly DP, McDonald IR, Wood AP. The family *Methylobacteriaceae*. In: Rosenberg
639 E, DeLong EF, Lory S, Stackebrandt E, Thompson F (eds). *The Prokaryotes*, 4th
640 ed. 2014. Springer Berlin Heidelberg, pp 313–340.
641 25. Chen I-MA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, et al. IMG/M v.5.0:
642 an integrated data management and comparative analysis system for microbial
643 genomes and microbiomes. *Nucleic Acids Res* 2019; **47**: D666–D677.

644 26. Green PN, Ardley JK. Review of the genus *Methylobacterium* and closely related
645 organisms: a proposal that some *Methylobacterium* species be reclassified into a
646 new genus, *Methylorubrum* gen. nov. *Int J Syst Evol Microbiol* 2018; **68**: 2727–
647 2748.
648 27. Parte AC. LPSN – List of prokaryotic names with standing in nomenclature
649 (bacterio.net), 20 years on. *Int J Syst Evol Microbiol* 2018; **68**: 1825–1829.
650 28. Menna P, Barcellos FG, Hungria M. Phylogeny and taxonomy of a diverse
651 collection of *Bradyrhizobium* strains based on multilocus sequence analysis of the
652 16S rRNA gene, ITS region and *glnII, recA, atpD* and *dnaK* genes. *International*
653 *Journal of Systematic and Evolutionary Microbiology* 2009; **59**: 2934–2950.
654 29. Zhang YM, Tian CF, Sui XH, Chen WF, Chen WX. Robust markers reflecting
655 phylogeny and taxonomy of rhizobia. *PLOS ONE* 2012; **7**: e44936.
656 30. Gaunt MW, Turner SL, Rigottier-Gois L, Lloyd-Macgilp SA, Young JP. Phylogenies
657 of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia.
658 *International Journal of Systematic and Evolutionary Microbiology* 2001; **51**: 2037–
659 2048.
660 31. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable
661 generation of high-quality protein multiple sequence alignments using Clustal
662 Omega. *Mol Syst Biol* 2011; **7**: 539.
663 32. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective
664 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*
665 2015; **32**: 268–274.

666 33. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for
667 inference of large phylogenetic trees. *Proceedings of the Gateway Computing*
668 *Environments Workshop (GCE)*. 2010. New Orleans, LA, pp 1–8.
669 34. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast,
670 accurate, and powerful alternative. *Syst Biol* 2006; **55**: 539–552.
671 35. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple alignment of
672 conserved genomic sequence with rearrangements. *Genome Res* 2004; **14**: 1394–
673 1403.
674 36. Darling AE, Mau B, Perna NT. progressiveMauve: Multiple genome alignment with
675 gene gain, loss and rearrangement. *PLOS ONE* 2010; **5**: e11147.
676 37. Guy L, Roat Kultima J, Andersson SGE. genoPlotR: comparative gene and
677 genome visualization in R. *Bioinformatics* 2010; **26**: 2334–2335.
678 38. R Core Team. R: A language and environment for statistical computing. 2018. R
679 Foundation for Statistical Computing, Vienna, Austria.
680 39. RStudio. 2018. RStudio, Boston, MA.
681 40. Pagès H, Aboyoun P, Gentleman R, DebbRoy S. Biostrings: Efficient manipulation
682 of biological strings. 2019.
683 41. Karlin S. Detecting anomalous gene clusters and pathogenicity islands in diverse
684 bacterial genomes. *Trends Microbiol* 2001; **9**: 335–343.
685 42. Delaney NF, Kaczmarek ME, Ward LM, Swanson PK, Lee M-C, Marx CJ.
686 Development of an optimized medium, strain and high-throughput culturing
687 methods for *Methylobacterium extorquens*. *PLOS ONE* 2013; **8**: e62957.

688 43. Atlas RM. Handbook of Microbiological Media, Fourth Edition, 4 edition. 2010. CRC
689 Press, Washington, D.C. : Boca Raton, FL.
690 44. Skovran E, Martinez-Gomez NC. Just add lanthanides. *Science* 2015; **348**: 862–
691 863.
692 45. Nash T. The colorimetric estimation of formaldehyde by means of the Hantzsch
693 reaction. *Biochem J* 1953; **55**: 416–421.
694 46. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Tennessen
695 K, et al. The standard operating procedure of the DOE-JGI Metagenome
696 Annotation Pipeline (MAP v.4). *Standards in Genomic Sciences* 2016; **11**: 17.
697 47. Matsen FA, Kodner RB, Armbrust EV. pplacer: linear time maximum-likelihood and
698 Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC*
699 *Bioinformatics* 2010; **11**: 538.
700 48. Eddy SR. HMMER - biosequence analysis using profile hidden Markov models.
701 2007.
702 49. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis
703 of large phylogenies. *Bioinformatics* 2014; **30**: 1312–1313.
704 50. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive
705 analysis and graphics of microbiome census data. *PLOS ONE* 2013; **8**: e61217.
706 51. Anda M, Ikeda S, Eda S, Okubo T, Sato S, Tabata S, et al. Isolation and genetic
707 characterization of *Aurantimonas* and *Methylobacterium* strains from stems of
708 hypernodulated soybeans. *Microbes Environ* 2011; **26**: 172–180.

52. Nayak DD, Marx CJ. Genetic and phenotypic comparison of facultative methylotrophy between *Methylobacterium extorquens* strains PA1 and AM1. *PLoS ONE* 2014; **9**: e107887.

53. Knief C, Frances L, Vorholt JA. Competitiveness of diverse *Methylobacterium* strains in the phyllosphere of *Arabidopsis thaliana* and identification of representative models, including *M. extorquens* PA1. *Microb Ecol* 2010; **60**: 440–452.

54. Marx CJ, Bringel F, Chistoserdova L, Moulin L, Farhan Ul Haque M, Fleischman DE, et al. Complete genome sequences of six strains of the genus *Methylobacterium*. *J Bacteriol* 2012; **194**: 4746–4748.

55. Wozniak RAF, Waldor MK. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* 2010; **8**: 552–563.

56. Lawrence JG, Ochman H. Amelioration of bacterial genomes: Rates of change and exchange. *J Mol Evol* 1997; **44**: 383–397.

57. Moreno R, Rojo F. The target for the *Pseudomonas putida* Crc global regulator in the benzoate degradation pathway is the *BenR* transcriptional regulator. *J Bacteriol* 2008; **190**: 1539–1545.

58. Fitzgerald DJ, Stratford M, Gasson MJ, Ueckert J, Bos A, Narbad A. Mode of antimicrobial action of vanillin against *Escherichia coli*, *Lactobacillus plantarum* and *Listeria innocua*. *J Appl Microbiol* 2004; **97**: 104–113.

59. Kuiters AT. Role of phenolic substances from decomposing forest litter in plant–soil interactions. *Acta Botanica Neerlandica* 1990; **39**: 329–348.

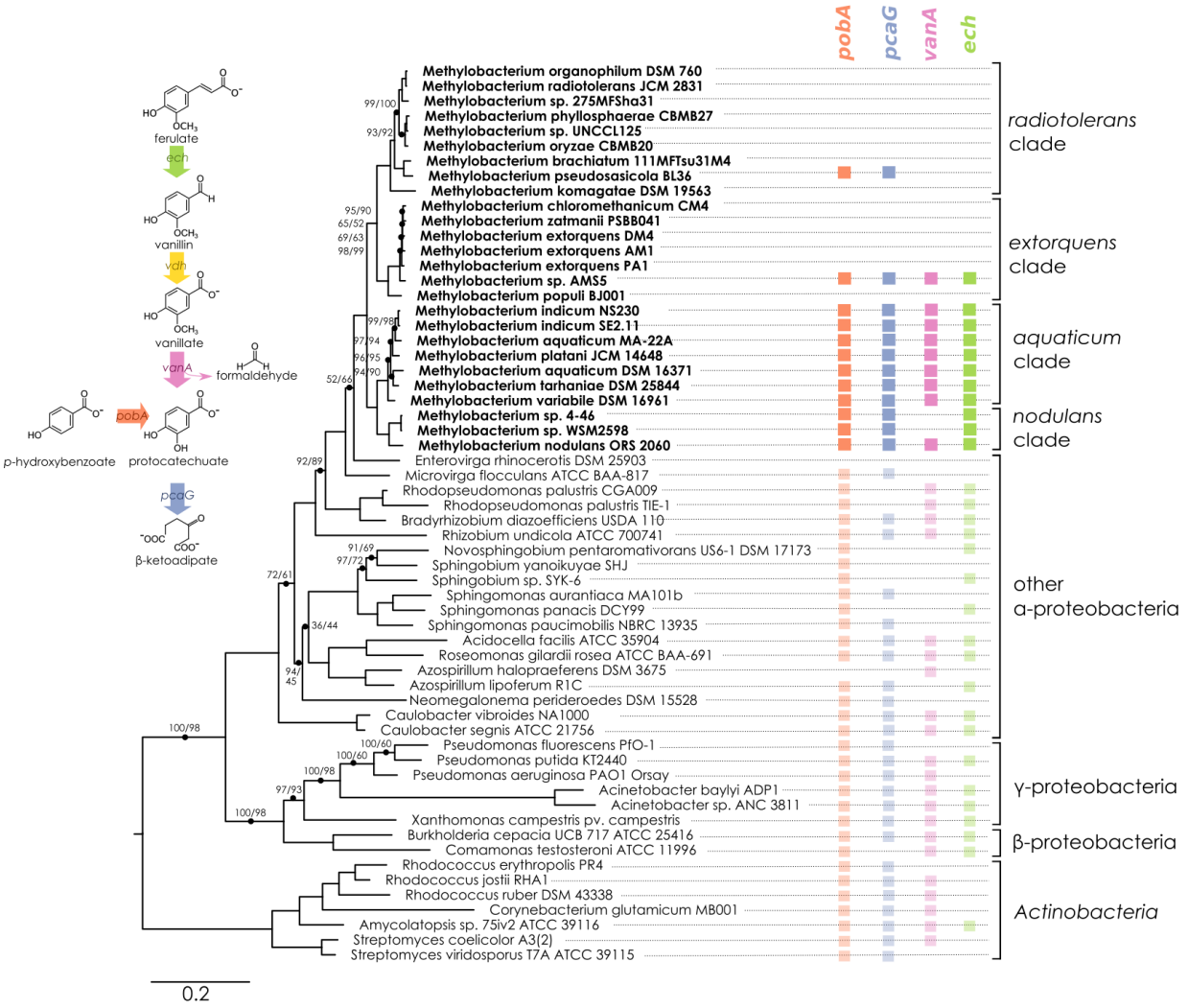
60. Zhalnina K, Louie KB, Hao Z, Mansoori N, Rocha UN da, Shi S, et al. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat Microbiol* 2018; **3**: 470–480.
61. Mandal SM, Chakraborty D, Dey S. Phenolic acids act as signaling molecules in plant-microbe symbioses. *Plant Signal Behav* 2010; **5**: 359–368.
62. Seneviratne G, Jayasinghearachchi HS. Phenolic acids: Possible agents of modifying N_2-fixing symbiosis through rhizobial alteration? *Plant and Soil* 2003; **252**: 385–395.
63. Fleischman D, Kramer D. Photosynthetic rhizobia. *Biochim Biophys Acta Bioeng* 1998; **1364**: 17–36.
64. Ardley JK, O'Hara GW, Reeve WG, Yates RJ, Dilworth MJ, Tiwari RP, et al. Root nodule bacteria isolated from South African *Lotononis bainesii*, *L. listii* and *L. solitudinis* are species of *Methylobacterium* that are unable to utilize methanol. *Arch Microbiol* 2009; **191**: 311–318.
65. Vuilleumier S, Chistoserdova L, Lee M-C, Bringel F, Lajus A, Zhou Y, et al. *Methylobacterium* genome sequences: A reference blueprint to investigate microbial metabolism of C1 compounds from natural and industrial sources. *PLOS ONE* 2009; **4**: e5584.
66. Schmid-Appert M, Zoller K, Traber H, Vuilleumier S, Leisinger T. Association of newly discovered IS elements with the dichloromethane utilization genes of methylotrophic bacteria. *Microbiology* 1997; **143**: 2557–2567.
67. Nayak DD, Marx CJ. Experimental horizontal gene transfer of methylamine dehydrogenase mimics prevalent exchange in nature and overcomes the

methylamine growth constraints posed by the sub-optimal N-methylglutamate pathway. *Microorganisms* 2015; **3**: 60–79.

68. Michener JK, Vuilleumier S, Bringel F, Marx CJ. Phylogeny poorly predicts the utility of a challenging horizontally transferred gene in *Methylobacterium* strains. *J Bacteriol* 2014; **196**: 2101–2107.

69. Bastard K, Smith AAT, Vergne-Vaxelaire C, Perret A, Zaparucha A, De Melo-Minardi R, et al. Revealing the hidden functional diversity of an enzyme family. *Nature Chemical Biology* 2014; **10**: 42–49.

764 # Figures and legends
765



766
767
768 **Figure 1. Genes associated with aromatic catabolism are present in two clusters of**
769 **closely-related *Methylobacterium* species, with some exceptions.**
770 Genomes of *Methylobacterium* species and known aromatic-degrading bacteria were
771 searched for four genes involved in different steps of the degradation of lignin-derived
772 methoxylated aromatic compounds (upper left). Colored squares indicate the presence
773 of each gene. Among the reference organisms (not *Methylobacterium*), all species that
774 do not have *pcaG* do have the genes for protocatechuate 4,5-dioxygenase, indicating an
775 alternative ring cleavage mechanism and different pathway for the catabolism of PCA.
776 The phylogeny was composed using a concatenated alignment of four housekeeping
777 genes (*atpD, recA, rpob*, and 16S rRNA), with the cyanobacterium species
778 Prochlorococcus marinus as an outgroup (not shown). Branch labels indicate SH-aLRT
779 branch support/ UltraFast bootstrap values; all unlabeled branches have values of
780 100/100. Accession numbers and published references are given in Tables S1 and S2.
781

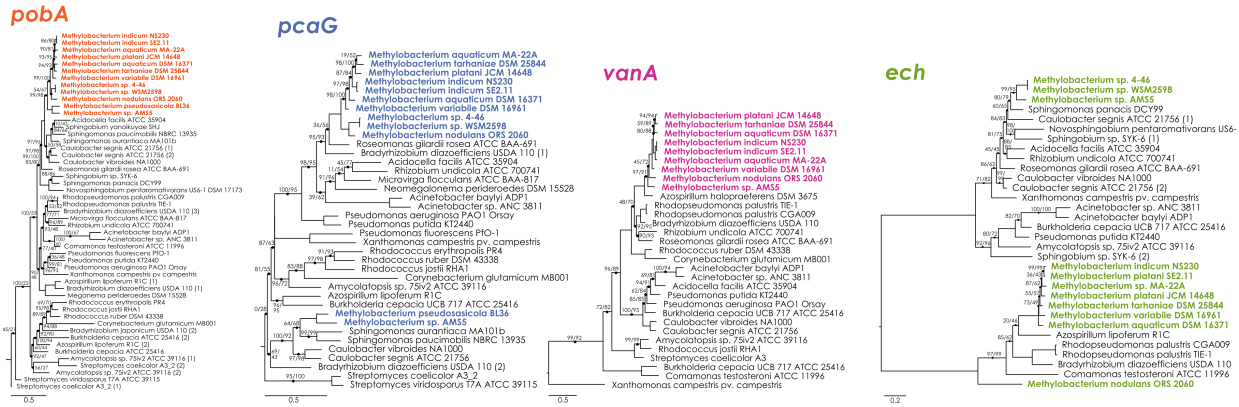


Figure 2. Phylogenies of aromatic catabolism genes suggest all four are ancestral to the *M. aquaticum/nodulans* clades, and have been horizontally acquired by two other *Methylobacterium* species.

For all organisms in Fig. 1, the sequences of *pobA* (*p*-hydroxybenzoate 3-monooxygenase), *pcaH* (protocatechuate 3,4-dioxygenase beta subunit), *vanA (*vanillate monooxygenase subunit A), and *ech* (*trans*-feruloyl-CoA hydratase/vanillin synthase) were aligned and phylogenies constructed using maximum likelihood, with homologs from archaeal organisms as outgroups (not shown). Accession numbers for all genes are given in Tables S1 and S2. Branch labels indicate SH-aLRT branch support/UltraFast bootstrap values; all unlabeled branches have values of 100/100. Comparison of these phylogenies and that in Fig. 1 suggest that all four genes are restricted to and prevalent in the *M. aquaticum* and *M. nodulans* clades, with a few losses (*vanA* by *Methylobacterium sp*. 4-46 and WSM2598) and some gains by horizontal gene transfer (in *M. pseudosasiacola* and *Methylobacterium sp*. AMS-5).

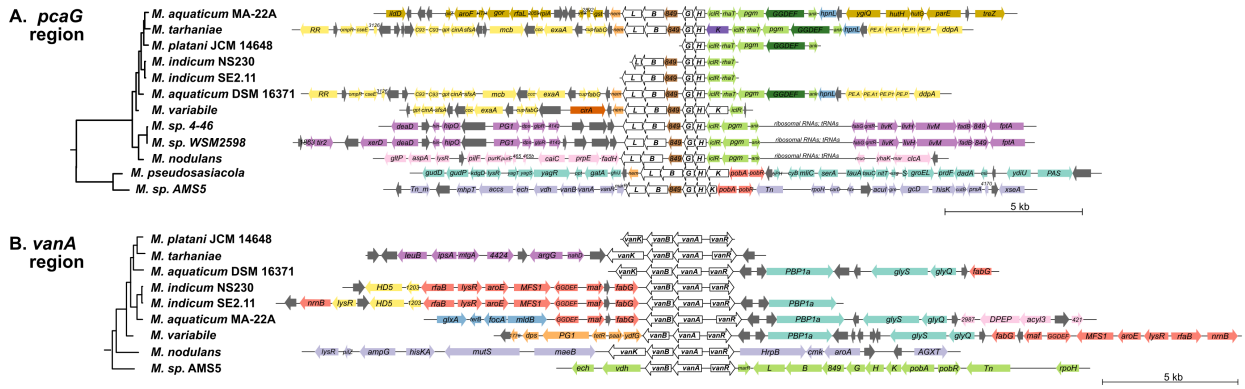


Figure 3. Genomic context of *vanA* and *pcaG* genes in *Methylobacterium* species supports a common evolutionary origin within the *M. aquaticum/nodulans* clades and separate origin in *M. pseudosasiacola* and *Methylobacterium sp.* AMS-5. Genomic regions (from fully assembled genomes) or scaffolds (from partially assembled genomes) are shown with A) all *pcaG* genes aligned or B) all *vanA* genes aligned. Gene phylogenies are excerpted from Fig. 2. Genes related to the beta-ketoadipate pathway (in A) and to vanillate demethylation (in B) are shown in white. Homologous regions shared across multiple genomes are shown in the same color to facilitate comparison among genomes. Gene labels are abbreviations based on functional annotations; full information on all genes is given in Table S3. Members of the *M. aquaticum* and *M. nodulans* clades share some synteny in the regions surrounding the *van* and *pca* gene clusters, with the degree of synteny roughly paralleling their phylogenetic distance; as an exception, *Methylobacterium sp.* AMS-5 and *M. pseudosasiacola* differ markedly from the other species.

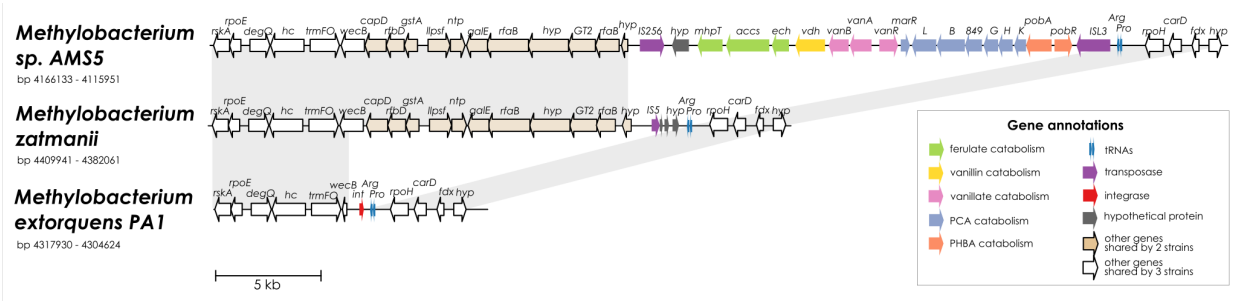


Figure 4. Alignment of three genomes from the *M. extorquens* clade reveals that *Methylobacterium sp*. AMS-5 harbors a catabolic island conferring the ability to degrade several aromatic substrates, in a genomic region prone to insertions and deletions.

The genome of *Methylobacterium sp*. AMS5 is shown aligned with that of two close relatives, *M. zatmanii* PSBB041 and *M. extorquens* PA1. Genes are color-coded to indicate function or commonality between genomes, as indicated by the key. Light gray shading between pairs of species connects regions shared by both. See Table S3 for full annotations corresponding to the short gene codes.

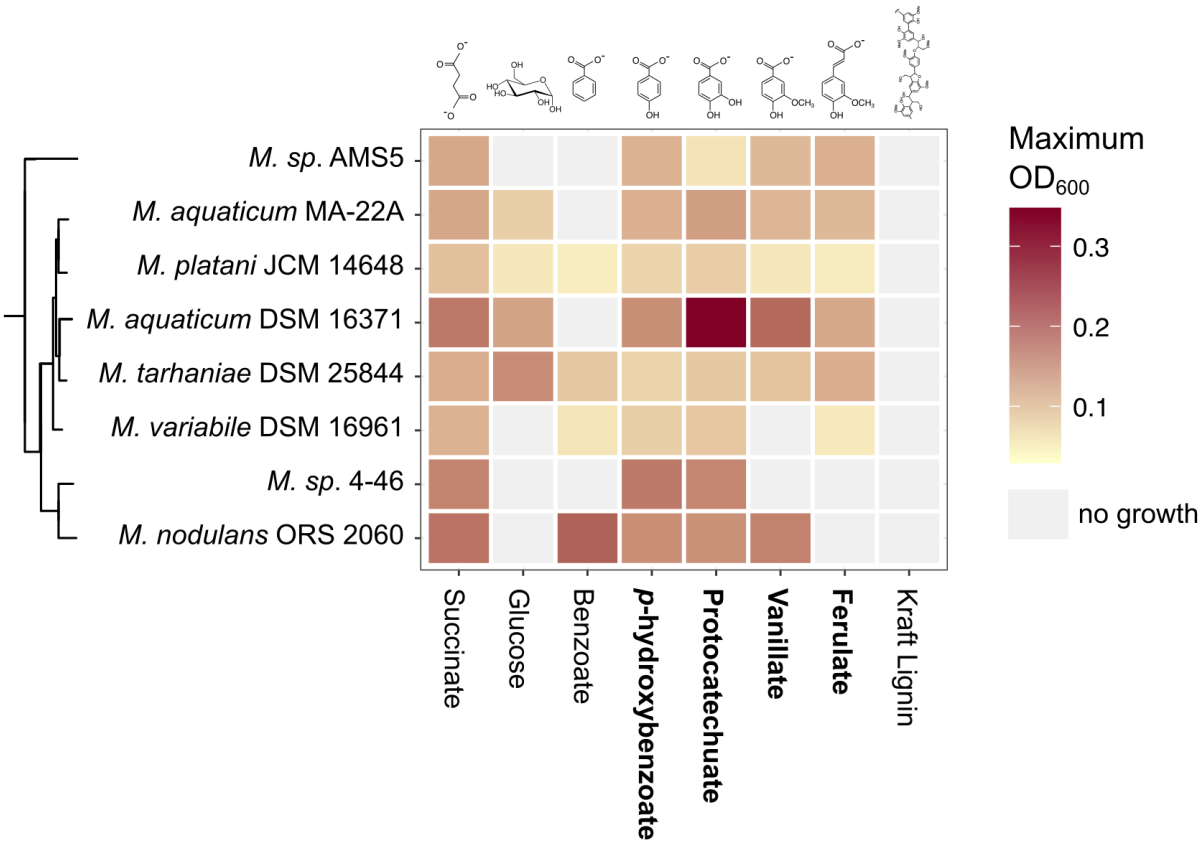


Figure 5. Growth of *Methylobacterium* species on aromatic compounds is predicted by genetic potential in most cases.
Methylobacterium species were grown in defined mineral medium with a single carbon substrate; all conditions contained 16 mM of carbon. Substrates corresponding to the four degradation pathways featured in this study are in bold; all species tested have genetic capacity for all four aromatic substrates except in the case of *Methylobacterium sp*. 4-46 and vanillate. Shading indicates maximum optical density (OD) at 600 nm measured during 80 hours of incubation, the average of three biological replicates. Original growth curves are shown in Fig. S2.

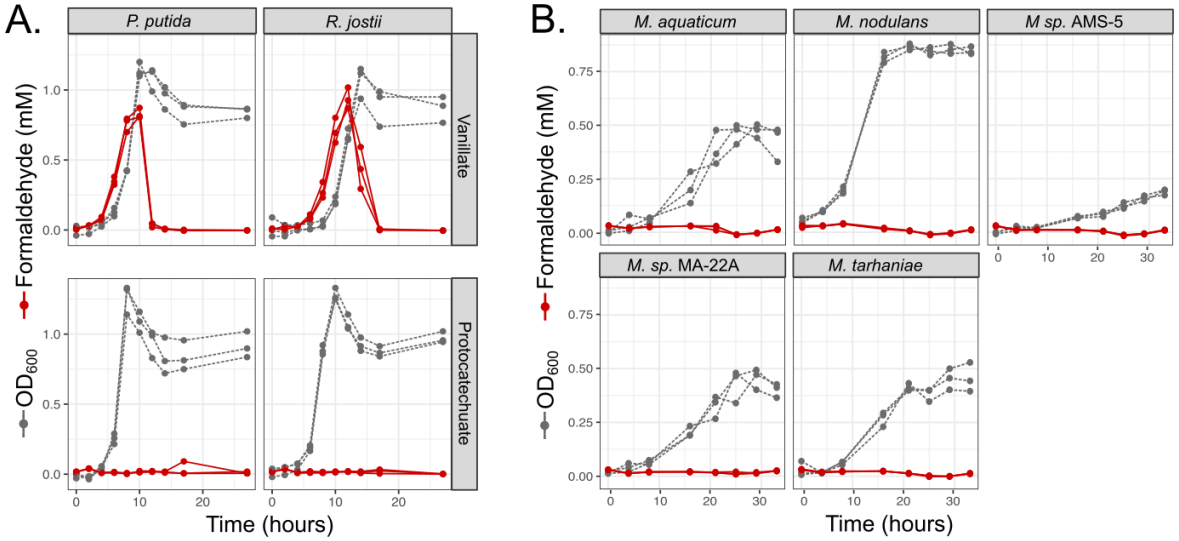


Figure 6. Non-methylotrophic lignin degraders *Pseudomonas putida* and *Rhodococcus jostii* produce formaldehyde when growing on methoxy-substituted aromatic compounds such as vanillate; *Methylobacterium* do not.
All organisms were grown in defined mineral medium with vanillate or protocatechuate (PCA) as a sole carbon source; growth was assayed by optical density at 600 nm (gray symbols and dashed lines) and formaldehyde in the medium was measured by a colorimetric assay (red symbols and solid lines). Each line represents one biological replicate. A) *P. putida* and *R. jostii* accumulate formaldehyde transiently in the medium when growing on vanillate, but not on the non-methoxylated compound PCA, and growth is slower on vanillate. B) All *Methylobacterium* species tested grew on vanillate without producing measurable formaldehyde.

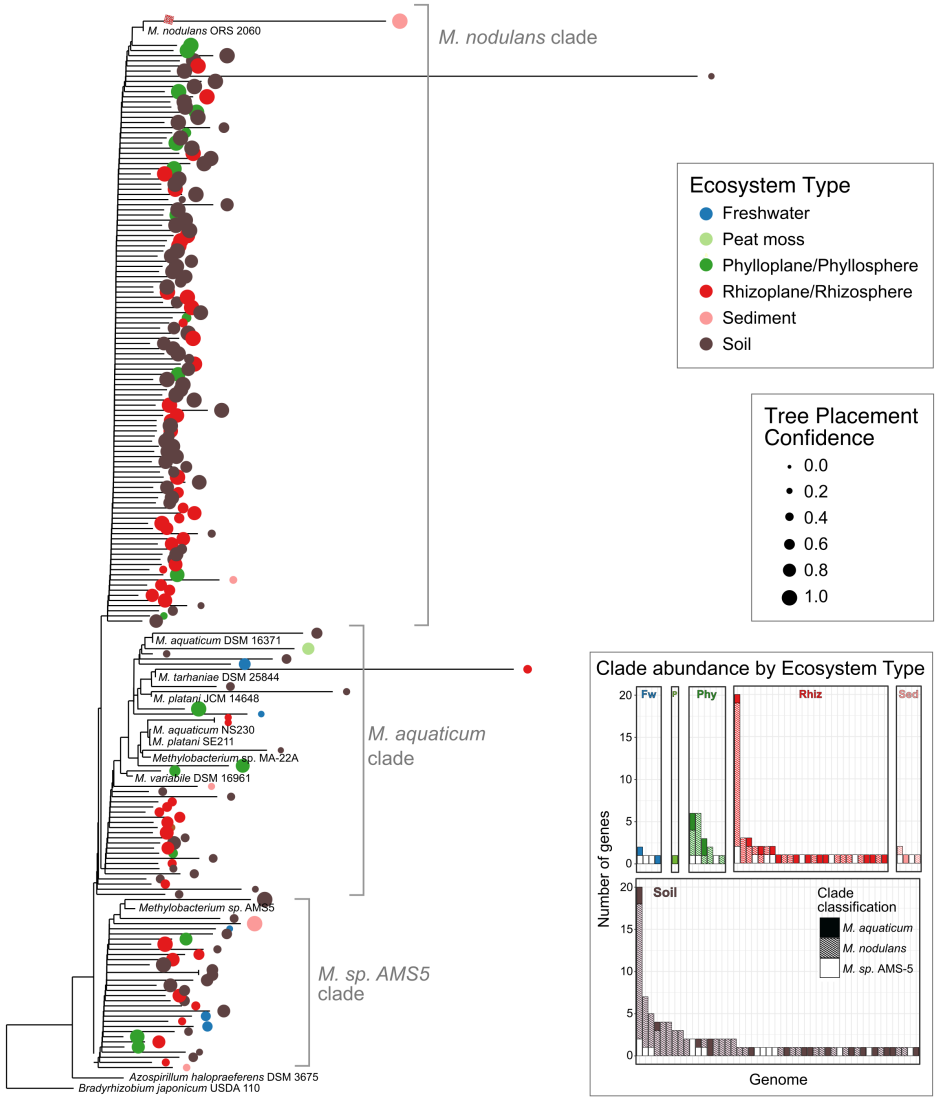


Figure 7. Organisms carrying *vanA* are most abundant in peat moss, rhizosphere, and soil; *Methylobacterium* are most abundant in the phyllosphere both overall and among *vanA*-carrying organisms.

Relative abundances of A) *vanA* genes and B*) Methylobacterium* genes as a percentage of all genes in each metagenome in the IMG/M database at the time of study. The numbers above each bar give the number of metagenomes searched in each ecosystem type (all metagenomes contained both *vanA* genes and *Methylobacterium* genes). C) Relative abundance of *Methylobacterium vanA* genes as a percentage of all *vanA* genes. The fraction above each bar denotes the number of metagenomes in which *Methylobacterium vanA* genes were found over the number of metagenomes searched for that ecosystem type (metagenomes were only searched if they had >100 *vanA* genes). Error bars denote the standard error among all metagenomes in the set.

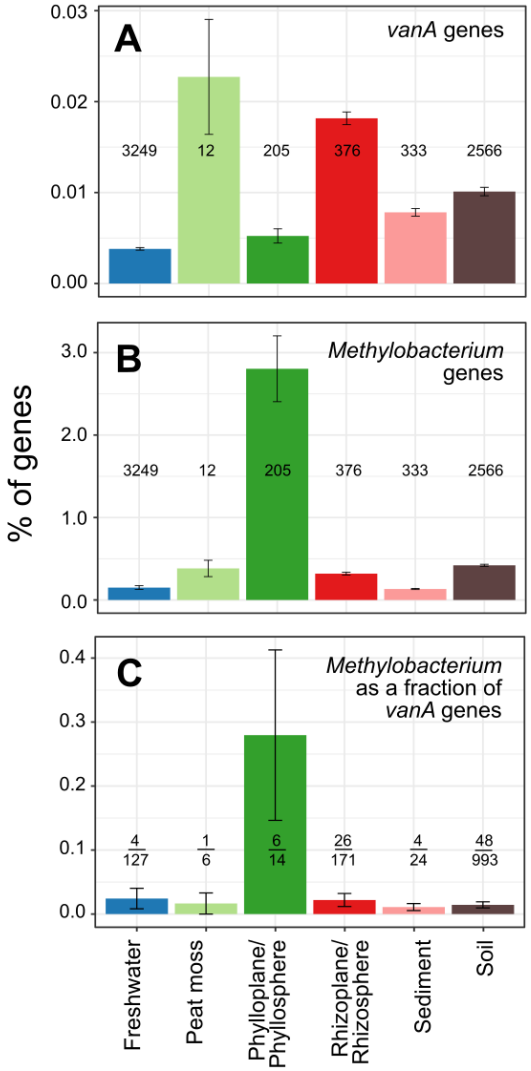


Figure 8. Among *Methylobacterium vanA* sequences in published metagenomes, the *M. nodulans* clade is most abundant.

Sequences of 182 *Methylobacterium*-associated *vanA* gene fragments found in published metagenomes were added using pplacer [47] to a phylogeny of full-length *vanA* genes sequenced from reference genomes. 166 genes that were classified by IMG as *Methylobacterium* but clustered outside of the genus are not shown. Colored dots indicate the ecosystem type from which the metagenome sample originated, with size scaled to the likelihood weight ratio of the pplacer classification as a measure of confidence in the placement. 114 genes (63%) are classified as *M. nodulans*; 31 genes (17%) as *Methylobacterium sp.* AMS5, and 37 genes (20%) as *M. aquaticum*. Inset bar plot: abundance of *Methylobacterium vanA* genes, shaded by clade, in each genome in which they were found. Each bar represents one genome (genome IDs not shown).