Methylobacterium Genome Sequences: A Reference Blueprint to Investigate Microbial Metabolism of C1 Compounds from Natural and Industrial Sources

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation

Published Version
doi:10.1371/journal.pone.0005584

Accessed
June 16, 2017 4:20:27 AM EDT

Citable Link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:4454165

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

(Article begins on next page)
**Methylobacterium** Genome Sequences: A Reference Blueprint to Investigate Microbial Metabolism of C1 Compounds from Natural and Industrial Sources

Stéphane Vuilleumier1,9, Ludmila Chistoserdova2,9, Ming-Chun Lee3, Françoise Bringel1, Aurélie Lajus4, Yang Zhou5, Benjamin Girion6, Valérie Barbe7, Jean Chang5, Stéphane Cruveiller4, Carole Dossat7, Will Gillett5, Christelle Gruffaz1, Eric Haugen5, Edith Hourcade1, Ruth Levy5, Sophie Mangenot7, Emilie Muller1, Thierry Nadalig1, Marco Pagni8, Christian Penny1, Rémi Peyraud6, David G. Robinson3, David Roche4, Zoé Rouy4, Channakhone Saenampechek5, Grégory Salvignol4, David Vallenet4, Zaining Wu5, Christopher J. Marx3, Julia A. Vorholt9, Maynard V. Olson5,9,10, Rajinder Kaul5,10, Jean Weissenbach4, Claudine Médigue4, Mary E. Lidstrom2,11

1 Département Micro-organismes, Génomes, Environnement, Université de Strasbourg, UMR7156 CNRS, Strasbourg, France, 2 Department of Chemical Engineering, University of Washington, Seattle, Washington, United States of America, 3 Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States of America, 4 Institut de Génomique, Laboratoire de Génomique Comparative, CEA and UMR8030 CNRS, Génoscope, Evry, France, 5 Genome Center, University of Washington, Seattle, Washington, United States of America, 6 Institute of Microbiology, ETH Zurich, Zurich, Switzerland, 7 Institut de Génomique, CEA, Génoscope, Evry, France, 8 Swiss Institute of Bioinformatics, Vital-IT group, Lausanne, Switzerland, 9 Department of Genome Sciences, University of Washington, Seattle, Washington, United States of America, 10 Department of Medicine, University of Washington, Seattle, Washington, United States of America, 11 Department of Microbiology, University of Washington, Seattle, Washington, United States of America

---

**Abstract**

**Background:** Methylo trophy describes the ability of organisms to grow on reduced organic compounds without carbon-carbon bonds. The genomes of two pink-pigmented facultative methylotrophic bacteria of the Alpha-proteobacterial genus *Methylobacterium*, the reference species *Methylobacterium extorquens* strain AM1 and the dichloromethane-degrading strain DM4, were compared.

**Methodology/Principal Findings:** The 6.88 Mb genome of strain AM1 comprises a 5.51 Mb chromosome, a 1.26 Mb megaplasmid and three plasmids, while the 6.12 Mb genome of strain DM4 features a 5.94 Mb chromosome and two plasmids. The chromosomes are highly syntenic and share a large majority of genes, while plasmids are mostly strain-specific, with the exception of a 130 kb region of the strain AM1 megaplasmid which is syntenic to a chromosomal region of strain DM4. Both genomes contain large sets of insertion elements, many of them strain-specific, suggesting an important potential for genomic plasticity. Most of the genomic determinants associated with methylotrophy are nearly identical, with two exceptions that illustrate the metabolic and genomic versatility of *Methylobacterium*. A 126 kb dichloromethane utilization (dcm) gene cluster is essential for the ability of strain DM4 to use DCM as the sole carbon and energy source for growth and is unique to strain DM4. The methylamine utilization (mau) gene cluster is only found in strain AM1, indicating that strain DM4 employs an alternative system for growth with methylamine. The *dcm* and *mau* clusters represent two of the chromosomal genomic islands (AM1: 28; DM4: 17) that were defined. The *mau* cluster is flanked by mobile elements, but the *dcm* cluster disrupts a gene annotated as chelatase and for which we propose the name “island integration determinant” (iid).

**Conclusion/Significance:** These two genome sequences provide a platform for intra- and interspecies genomic comparisons in the genus *Methylobacterium*, and for investigations of the adaptive mechanisms which allow bacterial lineages to acquire methylotrophic lifestyles.

---


**Editor:** Niyaz Ahmed, University of Hyderabad, India

**Received** January 24, 2009; **Accepted** March 30, 2009; **Published** May 18, 2009

**Copyright:** © 2009 Vuilleumier et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** SV was supported by a CNRS ATIP career development award, a CNRS USA mobility grant and a Génoscope sequencing grant, and also by the Integrated Computational Genomics Resources of the Swiss Institute of Bioinformatics (RTA-CT-2006-026204) together with MP. Annotation was supported by a grant from MRT/ANR FFTV 2007, MicroScope project. CP, EH and EM were supported by PhD grants from the Government of Luxembourg, the French Ministry of Research and Région Alsace, respectively. RP was supported by ETH (ETH-25 08-2). ML, DGR, and CJM were supported by NSF (108-0612591), the Clarke-Cooke Fund, and the Harvard University Microbial Sciences Initiative. LC and MEL were supported by NIH (GM 58933). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: vuilleumier@unistra.fr (SV); milachis@uwashington.edu (LC)

9 These authors contributed equally to this work.
Introduction

Pink-pigmented facultative methylotrophs of the genus Methylobacterium are ubiquitous in soil, air and water environments [1]. The common trait of all Methylobacterium species is the ability to grow on one or several reduced one carbon (C1) compounds other than methane, most prominently methanol, which is a major volatile organic compound emitted by vegetation [2]. Accordingly, strains of Methylobacterium are often found in association with plants, either involved in bona fide synbioses as endophytes, or as epiphytes on leaf surfaces [3–7]. The potential of strains from this genus to provide biotechnological products of high added value has attracted sustained scientific attention [8,9].

Of all Methylobacterium strains, M. extorquens strain AM1 (formerly Pseudomonas AM1, Methylobacterium sp. AM1) is the best studied, and has served as a model organism for over four decades. It was first isolated in 1960 in Oxford, England, as an airborne contaminant growing on methanoline [10]. It was then used as a workhorse to characterize the serine cycle for assimilation of the C1-unit of methylene tetrahydrofolate, a central intermediate in methylotrophic metabolism, and more recently the ethylmalonyl-CoA pathway for glyoxylate regeneration [8,11–13] (Fig. 1). Enzymatic systems for oxidation of both methanol [14,15] and methylene [16], which involve the use of specific cofactors pyrroloquinoline quinone (PQQ) and tryptophan tryptophylquinone (TTQ), respectively [17], were characterized in strain AM1. Bacterial tetrahydromethanopterin (H4MPT)-dependent enzymes, now known to occur in most methylotrophs [18,19] but originally thought to be unique to archaeal methanogens [20] were also first demonstrated in this strain. In Methylobacterium, the H4MPT-dependent pathway has been shown to play a major role in both energy generation and protecting cells from formaldehyde poisoning [21]. As to the analogous tetrahydrofolate (H4F)-linked pathway that involves two enzymes encoded by *mtdh* and *fch*, also first discovered in this organism [22,23], its major role in assimilatory metabolism was recently identified in supplying C1 units into the serine cycle [24,25] (Fig. 1).

Draft genome data for M. extorquens AM1 have been available since 2003 [11] and have enabled transcriptomic and proteomic approaches (see e.g. [26,27]), as well as metabolomic studies (see e.g. [24,28,29]). Combined with the large complement of genetic tools developed for Methylobacterium (see e.g. [30,31]), this has established *M. extorquens* AM1 as a model for systems level investigations.

*Methylobacterium* strain DM4 has been isolated from industrial wastewater sludge in Switzerland, as part of efforts to characterize microorganisms able to degrade the organohalogenated pollutant dichloromethane (DCM) [32]. Unlike methanol and methylene, which are mainly produced naturally, DCM is better known as a synthetic compound [33,34]. Rated as potentially carcinogenic for humans and the most highly produced chlorinated organic compound [http://www.eurochlor.org/solvents], DCM is highly volatile (b.p. 38°C) and water-soluble, making it a widespread contaminant in the environment [35]. Aerobic methylotrophic bacteria capable of using DCM as the sole source of carbon and energy [36] express high levels of DCM dehalogenase, which transforms DCM into formaldehyde and two molecules of HCl [37]. The genotoxic effects of DCM in both mammals [38] and bacteria [39] are due to a short-lived intermediate in the enzymatic transformation of DCM to formaldehyde [40]. Growth with DCM, as the main trait that distinguishes strain DM4 from *M. extorquens* AM1, has led to its classification as a separate *Methylobacterium* species [41]. Basing on 16S rRNA gene sequence and DNA-DNA relatedness, it was recently proposed that strain DM4 should be reclassified as *M. extorquens* [42].

The primary objective of this work was to define a fully assembled and annotated reference genomic blueprint for *Methylobacterium*, to assist future experimental investigations of methylotrophic metabolism by global approaches. We report here complete genomic sequences of strains AM1 and DM4, and describe the genomic make-up and potential for genomic plasticity that underlies the extensive capacity of *Methylobacterium* for physiological adaption to methylotrophic lifestyles. Availability of complete genomic sequences of the two strains also provides the opportunity to define the conserved complements of genes associated with methylotrophy, and to investigate the differences between the two strains associated with strain-specific adaptations.

Results and Discussion

Genomic structure

The genome of *M. extorquens* AM1 totals 6.88 Mb and consists of five replicons: a chromosome of 5.51 Mbp (Acc. No. CP001510), a megaplasmid of 1.26 Mbp (CP001511), and three plasmids (25 kb (CP001512), 38 kb (CP001513), and 44 kb (CP001514)), with an average GC content of 68.5% (Table 1, Fig. 2). The genome of strain DM4 is somewhat smaller (6.12 Mb), and features only three replicons: a chromosome of 5.94 Mbp (Acc. No. FP103042) and two plasmids (141 kb (FP103043) and 38 kb (FP103044)), with an average GC content of 68.0% (Table 1, Fig. 2). Based on their sizes and the relative distribution of sequencing reads for each replicon, plasmids p1META and p2META are predicted to be present at 2–3 copies per strain AM1 genome, the replicon p3META at 1–2 copies per genome, and the megaplasmid at one copy per genome. Predicted copy numbers of 0.4–0.5 and 0.6–0.7 per genome were obtained for DM4 plasmids p1METDI and p2METDI, respectively.

By convention, the origin of both chromosomes was set upstream of the *dnaA* gene, as no GC skew was observed to help in predict initiation and termination of replication [43]. The chromosomes of the two strains are remarkably similar in both gene content and synteny (Fig. 2, Fig. 3). 85% of the *M. extorquens* AM1 chromosomal genes have full-length homologs (higher than 30% identity) on the chromosome of strain DM4. Of these, 89% have homologs at higher than 95% identity, underlining the orthologous nature of most genes in the two strains. Ribosomal genes (23S, 16S, and 5S) are identical in all five copies of the AM1 chromosomal genes and full-length homologs (higher than 30% identity) on the chromosome of strain DM4. Of these, 89% have homologs at higher than 95% identity, underlining the orthologous nature of most genes in the two strains. Ribosomal genes (23S, 16S, and 5S) are identical in all five copies of the ribosomal operon of strain AM1 and DM4. The intergenic spacer length between 16S and 23S genes is identical in all five copies of the ribosomal operon of the same strain, but its length differs markedly between the two strains (905 nt in *M. extorquens* AM1, 602 nt in strain DM4). These data confirm the already mentioned recent suggestion [42] that strain DM4 belongs to the species *M. extorquens*.

The distribution of functional categories according to the COG classification ([44], Table 2) was as expected for a free-living proteobacterium with a versatile lifestyle and the observed genome size. COG functional class assignments are more frequent and diverse for chromosomal genes than for plasmid genes (Table 2). No significant differences in functional classes were evident between AM1 and DM4 chromosomes, except for the larger proportion of genes associated with recombination, replication and repair in strain AM1, a reflection of the larger set of IS elements in that strain (Table 1, and see below).

The Mage annotation platform [45] and Alien Hunter [46] were used to detect genes and genome regions found in one strain but not in the other. Several unique chromosomal regions, termed genomic islands and ranging from a few genes to hundreds of
genes, which represent approximately 632 kb (11.5%) and 1,054 kb (17.7%) of the chromosome for strains AM1 and DM4, respectively, were defined (Table 3). With the exception of the dcm and mau gene clusters (see below), few of these islands appear to encode functions important for central metabolism or methylo-
troph. One remarkable genomic island in strain AM1 (Table 3) contains a hypothetical gene of unknown function of 47.5 kb (META1_2412) which encodes a 15,831 residue-long repeat-rich polypeptide (Pfam PF00353 (hemolysin-type calcium-binding region); PF05594, (haemagglutinin, bacterial); COG3210 (large exoproteins involved in heme utilization or adhesion), and COG2931 (RTX toxins and related Ca2+–binding proteins). This gene product, if expressed, would represent one of the largest proteins known in biology [47].
Table 1. Genome statistics for M. extorquens AM1 and DM4.

<table>
<thead>
<tr>
<th>Gene Category</th>
<th>Strain AM1</th>
<th>Strain AM1</th>
<th>Strain AM1</th>
<th>Strain AM1</th>
<th>Strain AM1</th>
<th>Strain DM4</th>
<th>Strain DM4</th>
<th>Strain DM4</th>
<th>Strain DM4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome Megaplasmid</td>
<td>1261460</td>
<td>44195</td>
<td>37858</td>
<td>24943</td>
<td>6879778</td>
<td>5943768</td>
<td>141504</td>
<td>38579</td>
<td>6123851</td>
</tr>
<tr>
<td>GC (%)</td>
<td>68.7</td>
<td>67.7</td>
<td>67.9</td>
<td>65.3</td>
<td>66.9</td>
<td>68.5</td>
<td>68.1</td>
<td>65.3</td>
<td>63.7</td>
</tr>
<tr>
<td>Repeat regions (%)</td>
<td>8.3</td>
<td>7.9</td>
<td>0.2</td>
<td>2.7</td>
<td>0.3</td>
<td>8.0</td>
<td>9.3</td>
<td>1.6</td>
<td>0</td>
</tr>
<tr>
<td>Genes</td>
<td>5315</td>
<td>1318</td>
<td>46</td>
<td>45</td>
<td>35</td>
<td>6759</td>
<td>5857</td>
<td>137</td>
<td>41</td>
</tr>
<tr>
<td>Protein-coding genes</td>
<td>5227</td>
<td>1312</td>
<td>46</td>
<td>45</td>
<td>35</td>
<td>6665</td>
<td>5769</td>
<td>137</td>
<td>41</td>
</tr>
</tbody>
</table>

Average length

- CDS (bp): 891.7
- Intergenic (bp): 176.4
- rRNA operons: 0
- rRNA: 5

Total length: 2.4%

Intact IS: 41 (25)

Partial IS: 22 (13)

MITEs: 1

<table>
<thead>
<tr>
<th>IS Type</th>
<th>AM1</th>
<th>AM1</th>
<th>AM1</th>
<th>AM1</th>
<th>DM4</th>
<th>DM4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total length (%)</td>
<td>2.4</td>
<td>7.6</td>
<td>22.0</td>
<td>27.5</td>
<td>15.5</td>
<td>3.7</td>
</tr>
<tr>
<td>Intact IS</td>
<td>93 (19)</td>
<td>41 (25)</td>
<td>3 (3)</td>
<td>4 (4)</td>
<td>1 (1)</td>
<td>142 (39)</td>
</tr>
<tr>
<td>Partial IS</td>
<td>8 (7)</td>
<td>22 (13)</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td>0</td>
<td>32 (19)</td>
</tr>
<tr>
<td>MITEs</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Extra-chromosomal replicons are highly strain-specific and show little similarity in size, gene content or synteny with each other. However, an approximately 130 kb region of the AM1 megaplasmid is globally syntenic to a region of similar length in the chromosome of strain DM4 (Fig. 3). Plasmids encode mostly proteins of currently unknown function (Table 2) or proteins associated with plasmid-related functions. Exceptions include a cluster in AM1 encoding two acyl homoserine lactone synthases [48]; and a cluster of copper resistance genes on plasmid p1META1 (p1META1_0021/p1META1_0022). A cluster of copper resistance genes on plasmid p2META1 (p2META1_0029/p2META1_0030); a truncated luxI gene (pMETA1_0049) recently shown to be essential for the operation of two bona fide, chromosomally-located luxI genes, and encoding two acyl homoserine lactone synthases [48]; and an UsDC system involved in SOS DNA repair. Unlike in strain DM4, which has two complete copies of usDC, all chromosomal copies of usDC are on the megaplasmid (META2_0643/META2_0644), while a truncated copy of usDC is found on the chromosome (META1_4790).

Comparative genomics of aerobic methylotrophy

Methylotrophy can be envisioned in terms of the assembly of discrete metabolic modules, each responsible for a specific metabolic task, which in combination define pathways for methylotrophic metabolism, several variants of which have been well characterized [11,49].

The Methyllobacterium blueprint. In Methylobacterium, the currently recognized methylotrophy genes and modules are found exclusively on the chromosomes of strains AM1 and DM4 (Fig. 2, Suppl. Table S1). Common genes associated with methylotrophy in AM1 are displayed at least 95% identity at the protein level (99.1% average), with complete synteny between the two strains [11]. Several methylotrophy genes are found as singletons, including several cases of genes that encode different subunits of the same enzyme (e.g. pemAB, pccAB, see Suppl. Table S1). Nevertheless, a majority of methylotrophy genes are found in large clusters. Only two known methylotrophy gene clusters are not shared between the two strains (Suppl. Table S1, and see below): the dcm (dichloromethane degradation) gene region present only in strain DM4, and the man gene cluster encoding methylamine dehydrogenase and accessory functions in strain AM1. One large multi-operon cluster (49.3 kb) encodes most of the serine cycle enzymes, most of the PQQ biosynthesis functions [17], genes for H4MPT-linked reactions and H4MPT biosynthesis, and H4F biosynthesis genes. It also contains genes encoding a homolog of methanol dehydrogenase (XoxFG) of still unknown function, often found nearby genes involved in G1 metabolism [50,51] and recently suggested to be involved in formaldehyde metabolism in the photosynthetic bacterium Rhodobacter sphaeroides [52].

Comparison of gene sets for methylotrophy in fully sequenced genomes. A steadily increasing number of genomes of methylotrophic microorganisms have been sequenced, assembled and annotated. We limit our comparative analysis of known genetic determinants and modules of methylotrophy (Suppl. Table S2) to completed, manually annotated and officially published methylotroph genomes (listed in Suppl. Table S3), six of which belong to the phylum Proteobacteria and one to the phylum Verrucomicrobia. Methyllobacterium capsulatus represents Gamma-proteobacterial methanotrophs [53], while Methyllobacterium petroleiphilum [54], Methyllobacterium flaxellaceum [55] and Methylophilales strain HTCC2181 [56] feature two different orders within Beta-proteobacteria (Burkholderiales and Methylophilales). Sipellobacter pinnatus, although not reported to grow methylotrophically, is an Alpha-proteobacterium of the family

<table>
<thead>
<tr>
<th>Strain AM1</th>
<th>Strain AM1</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein-coding genes</td>
<td>5227</td>
<td>1312</td>
</tr>
<tr>
<td>rRNA operons</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>rRNA</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

*Defined by the algorithm Nosferatu as implemented in Mage [45].

*Number of IS elements (number of IS types in brackets).
Rhodobacteriaceae capable of degrading methylated sulfur compounds [57]. Granulibacter bethesdensis is an emerging human pathogen of the family of Acetobacteriaceae within Alpha-proteobacteria [58] reported to grow on methanol [59]. Finally, strain V4 (candidatus “Methyloacidiphilum infernorum”) represents the recently discovered group of thermophilic and acidophilic methanotrophs of the phylum Verrucomicrobia [60].

**Methanol utilization.** The mxa gene cluster encoding the classic methanol dehydrogenase is nearly identical (over 99% identity at the protein level) between strains AM1 and DM4, and very similar in the genomes of M. capsulatus, M. flagellatus, G. bethesdensis and several other proteobacterial methylotrophs [61]. This conservation of both gene sequence and gene synteny suggests that the mxa gene cluster has most likely disseminated via lateral transfer among methylotrophs of different subclasses of Proteobacteria. This notwithstanding, no similar gene clusters are recognizable in the genomes of the other four organisms discussed here.

**Methylamine utilization.** The mau gene cluster encoding the canonical system for methylamine utilization was characterized for a large part in strain AM1, and the genome of M. flagellatus [55] contains a mau gene cluster very similar to it. The main difference is that the gene for the electron acceptor from methylamine dehydrogenase in strain AM1, amicyanin, is replaced by a gene for azurin, an analogous copper-containing electron acceptor protein in M. flagellatus. The mau cluster was not found in genomes of the other methylotrophs including strain DM4 discussed here, which were shown or assumed to grow with methylamine (Suppl. Table S2). Thus, as yet uncharacterized genetic determinants are responsible for methylamine utilization in most methylotrophs, including strain DM4 in particular.

**H₄MPT-dependent formaldehyde oxidation.** Tetrahydomethanopterin (H₄MPT)-dependent formaldehyde oxidation is the main pathway for both energy generation and formaldehyde detoxification in M. extorquens, and therefore absolutely essential for methylotrophy in this organism [21,62,63]. First defined in M. extorquens AM1, this pathway has also been described in a variety of other bacteria, including from
phyla whose methylotrophic ability has not yet been demonstrated such as Planctomycetes [18,64,65]. Phylogenetic analysis suggests that this pathway must be one of the most ancient in the context of methylotrophic metabolism. However, it is unessential in M. flagellatus [66], and is absent in some other methylotrophs. S. pomeroyi possesses an alternative glutathione-dependent (FlhA/FghA) system for oxidation of formaldehyde similar to that of P. denitrificans [67] and R. sphaeroides [52]. No formaldehyde oxidation systems were identified in the genomes of Methylophilales HTCC2181 or Verrucomicrobia strain V4.

**Conversion of formate to CO₂.** M. extorquens strains possess four different functional formate dehydrogenases for the final step of energy generation from carbon oxidation [68]. The other methylotrophs included in our analysis also encode one or several FDH homologs (Suppl. Table S2), but only one, FHD2, is consistently detected. These observations suggest that formate oxidation, as a transformation ubiquitous to life, does not strictly qualify as a methylotrophy-specific reaction, and may thus involve analogous [69] enzymatic systems.

**C1 assimilation via methylene tetrahydrofolate and the serine cycle.** The serine cycle is essential for carbon assimilation in Methylobacterium and comprises reactions specific to methylotrophy as well as reactions involved in multicarbon metabolism (Fig. 1, see also [8,11]). Genes involved in the serine cycle can be ascribed to two categories on the basis of mutational analysis [11]: methylotrophy-specific genes (glyA, sga, hpr, gck, ppc, mtkAB and mcl), and genes which are essential under non methylotrophic growth conditions (eno and mth). Recent evidence [24,25] and mutant analyses [23,70–73] suggest that genes for the C1 transfer pathway linked to H₄F (mtdA, fch and fflL) are specifically involved in assimilatory metabolism in Methylobacterium. Six methylotrophy-specific serine cycle genes, along with mtdA and fch, belong to gene clusters associated with methylothrophy on the chromosomes of strains AM1 and DM4 (Fig. 4), while the three remaining genes (glyA, gck and fflL) are not parts of methylotrophy gene clusters and are located elsewhere on the chromosome.

As exemplified here for M. petroleiphilum [54], Beta-proteobacterial methylotrophs may also employ the serine cycle for C1 carbon assimilation [54,74]. As that of the Alpha-proteobacterium S. pomeroyi, the genome of M. petroleiphilum contains a single gene cluster encoding all required functions of the serine cycle (Fig. 4). In S. pomeroyi however, the organisation of this gene cluster is quite different from that of Methylobacterium, Granulibacter bethesdensis and M. petroleiphilum (Fig. 4), and contains tandem genes for two distantly related bifunctional methylene-H₄F dehydrogenase/methenyl-H₄F cyclohydrolase (FoID) enzymes instead of the isofunctional mtdA/fch genes found in the other genomes discussed here. Moreover, the hpr, gck and sga genes inferred from the genomic context display only modest sequence identity with Methylobacterium prototypes (Fig. 4), further suggesting that the serine cycle in S. pomeroyi belongs to an independent evolutionary lineage.

Extending the analysis to methylotrophic organisms able to grow with methane, the Gamma-proteobacterial methanotroph M. capsulatus also harbors serine cycle gene homologs in its genome, including the mtdA/fch pair, but few of them are clustered (Fig. 4). However, the gene for one key enzyme of the serine cycle, the methylotrophy-specific phosphoenolpyruvate carboxylase...
Table 2. Functional classes in *M. extorquens* AM1 and DM4 replicons.

<table>
<thead>
<tr>
<th>Class - Description</th>
<th>Strain AM1</th>
<th>Strain AM1</th>
<th>Strain AM1</th>
<th>Strain AM1</th>
<th>Strain AM1</th>
<th>Strain DM4</th>
<th>Strain DM4</th>
<th>Strain DM4</th>
<th>Strain DM4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chrom. megaplasmid p1</td>
<td>p2</td>
<td>p3</td>
<td>Class (%)</td>
<td>Proc. (%)</td>
<td>chrom.</td>
<td>p1</td>
<td>p2</td>
<td>Class (%)</td>
</tr>
<tr>
<td>O - Cell cycle control, cell division, chromosome partitioning</td>
<td>34</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0.66</td>
<td>16.28</td>
<td>38</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>M - Cell wall/membrane/envelope biogenesis</td>
<td>240</td>
<td>30</td>
<td>1</td>
<td>1</td>
<td>4.08</td>
<td>245</td>
<td>4</td>
<td>4.18</td>
<td></td>
</tr>
<tr>
<td>N - Cell motility</td>
<td>126</td>
<td>16</td>
<td>3</td>
<td>2.18</td>
<td>121</td>
<td>1</td>
<td>2.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O - Posttranslational modification, protein turnover, chaperones</td>
<td>164</td>
<td>32</td>
<td>2</td>
<td>2.97</td>
<td>201</td>
<td>2</td>
<td>3.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T - Signal transduction mechanisms</td>
<td>34</td>
<td>1</td>
<td>1</td>
<td>4.53</td>
<td>304</td>
<td>1</td>
<td>1</td>
<td>5.14</td>
<td></td>
</tr>
<tr>
<td>U - Intracellular trafficking, secretion, and vesicular transport</td>
<td>35</td>
<td>8</td>
<td>1</td>
<td>0.66</td>
<td>32</td>
<td>3</td>
<td>0.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V - Defense mechanisms</td>
<td>66</td>
<td>13</td>
<td>1</td>
<td>1.20</td>
<td>80</td>
<td>2</td>
<td>1.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B - Chromatin structure and dynamics</td>
<td>0.05</td>
<td>14.12</td>
<td>4</td>
<td>0.07</td>
<td>13.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J - Translation, ribosomal structure and biogenesis</td>
<td>188</td>
<td>14</td>
<td>1</td>
<td>3.05</td>
<td>199</td>
<td>1</td>
<td>3.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K - Transcription</td>
<td>205</td>
<td>46</td>
<td>2</td>
<td>2</td>
<td>3.86</td>
<td>243</td>
<td>5</td>
<td>2</td>
<td>4.20</td>
</tr>
<tr>
<td>L - Replication, recombination and repair</td>
<td>292</td>
<td>161</td>
<td>11</td>
<td>11</td>
<td>3</td>
<td>7.17</td>
<td>290</td>
<td>39</td>
<td>7</td>
</tr>
<tr>
<td>C - Energy production and conversion</td>
<td>262</td>
<td>26</td>
<td>1</td>
<td>4.34</td>
<td>245</td>
<td>10</td>
<td>5.18</td>
<td>27.39</td>
<td></td>
</tr>
<tr>
<td>E - Amino acid transport and metabolism</td>
<td>423</td>
<td>31</td>
<td>1</td>
<td>6.84</td>
<td>468</td>
<td>12</td>
<td>8.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F - Nucleotide transport and metabolism</td>
<td>79</td>
<td>11</td>
<td>1.35</td>
<td>80</td>
<td>1.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G - Carbohydrate transport and metabolism</td>
<td>141</td>
<td>9</td>
<td>2.25</td>
<td>144</td>
<td>2.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H - Coenzyme transport and metabolism</td>
<td>124</td>
<td>7</td>
<td>1</td>
<td>1.98</td>
<td>123</td>
<td>2.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I - Lipid transport and metabolism</td>
<td>164</td>
<td>11</td>
<td>2.63</td>
<td>174</td>
<td>2</td>
<td>2.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P - Inorganic ion transport and metabolism</td>
<td>200</td>
<td>35</td>
<td>3</td>
<td>3.59</td>
<td>220</td>
<td>2</td>
<td>3.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q - Secondary metabolites biosynthesis, transport and catabolism</td>
<td>89</td>
<td>12</td>
<td>1</td>
<td>1.53</td>
<td>92</td>
<td>1</td>
<td>1.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R - General function prediction only</td>
<td>355</td>
<td>40</td>
<td>2</td>
<td>5.96</td>
<td>11.24</td>
<td>388</td>
<td>1</td>
<td>2</td>
<td>6.57</td>
</tr>
<tr>
<td>S - Function unknown</td>
<td>312</td>
<td>35</td>
<td>2</td>
<td>5.28</td>
<td>327</td>
<td>2</td>
<td>1</td>
<td>5.55</td>
<td></td>
</tr>
<tr>
<td>CDS with at least one COG hit</td>
<td>3768</td>
<td>578</td>
<td>30</td>
<td>22</td>
<td>10</td>
<td>4408</td>
<td>66.15</td>
<td>4076</td>
<td>84</td>
</tr>
<tr>
<td>Total CDS</td>
<td>5227</td>
<td>1311</td>
<td>46</td>
<td>45</td>
<td>35</td>
<td>6664</td>
<td>5772</td>
<td>137</td>
<td>41</td>
</tr>
</tbody>
</table>

*Only the first COG hit of each CDS is considered (CDS may be associated with several COGs).

Processes: Cellular processes and signaling (D,M,N,O,T,U,V); information storage and processing (B,J,K,L); metabolism (C,E,F,G,H,I,P,Q); and poorly characterized (R,S).

doi:10.1371/journal.pone.0005584.t002
<table>
<thead>
<tr>
<th>Start CDS (META1_1)</th>
<th>Start (nt)</th>
<th>End CDS (META1_1)</th>
<th>End (nt)</th>
<th>Length (bp)</th>
<th>[Left border]</th>
<th>[Right border]</th>
<th>IS</th>
<th>Features (proposed role)</th>
<th>Total CDS (%)</th>
<th>Unique CDS (%)</th>
<th>AH CDS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Methylobacterium extorquens AM1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MITE-1</td>
<td>0035</td>
<td>37982</td>
<td>0046</td>
<td>51448</td>
<td>13467</td>
<td>[none][GC][IS]</td>
<td>1</td>
<td>14</td>
<td>11 (78.6)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>0149</td>
<td>15686</td>
<td>0165</td>
<td>17421</td>
<td>17859</td>
<td>[none][int-AH][tRNA]</td>
<td>0</td>
<td>16</td>
<td>13 (81.3)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>1078</td>
<td>1126013</td>
<td>1115</td>
<td>116485</td>
<td>38802</td>
<td>[IS][GC-mob-tRNA-tRNA][IS]</td>
<td>8</td>
<td>45</td>
<td>19 (42.2)</td>
<td>4 (8.9)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>1226</td>
<td>1283284</td>
<td>1257</td>
<td>130815</td>
<td>24832</td>
<td>[tRNA][none][tRNA]</td>
<td>0</td>
<td>sulfur metabolism</td>
<td>33</td>
<td>29 (87.9)</td>
<td>0</td>
</tr>
<tr>
<td>tRNA</td>
<td>149</td>
<td>15686</td>
<td>165</td>
<td>174721</td>
<td>18778</td>
<td>[tRNA][GC][int]</td>
<td>1</td>
<td>17</td>
<td>16 (94.1)</td>
<td>5 (29.4)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>2402</td>
<td>2475649</td>
<td>2415</td>
<td>2534027</td>
<td>276379</td>
<td>[tRNA-AH][none][GC]</td>
<td>4 (1)</td>
<td>34</td>
<td>25 (73.5)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>2657</td>
<td>2787350</td>
<td>2652</td>
<td>2781501</td>
<td>27708</td>
<td>[tRNA][none][tRNA]</td>
<td>0</td>
<td>metal transport</td>
<td>31</td>
<td>31 (100)</td>
<td>0</td>
</tr>
<tr>
<td>tRNA</td>
<td>2755</td>
<td>2814961</td>
<td>2697</td>
<td>2823774</td>
<td>8814</td>
<td>[tRNA][GC][int]</td>
<td>2</td>
<td>16</td>
<td>14 (87.5)</td>
<td>2 (13.3)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>3934</td>
<td>4056642</td>
<td>4086</td>
<td>4163355</td>
<td>106714</td>
<td>[tRNA][GC][int]</td>
<td>12</td>
<td>zinc metabolism</td>
<td>148</td>
<td>145 (98.0)</td>
<td>25 (16.9)</td>
</tr>
<tr>
<td>tRNA</td>
<td>4747</td>
<td>4874344</td>
<td>4792</td>
<td>4901372</td>
<td>27028</td>
<td>[tRNA][AH][int-mob(3)-int-AH]</td>
<td>8</td>
<td>28 (57.1)</td>
<td>25 (42.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Methylobacterium extorquens DM4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MITE-1</td>
<td>0036</td>
<td>37946</td>
<td>0052</td>
<td>51383</td>
<td>13438</td>
<td>[Mite][int - GC][IS]</td>
<td>2</td>
<td>16</td>
<td>13 (81.3)</td>
<td>11 (68.8)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>0137</td>
<td>137916</td>
<td>0147</td>
<td>146122</td>
<td>8207</td>
<td>[tRNA][none][AH]</td>
<td>2 (1)</td>
<td>SOS repair (umuCD)</td>
<td>8</td>
<td>0</td>
<td>4 (50.0)</td>
</tr>
<tr>
<td>tRNA</td>
<td>0225</td>
<td>232397</td>
<td>0336</td>
<td>321727</td>
<td>89331</td>
<td>[tRNA][AH][none]</td>
<td>4</td>
<td>sensors/regulators, transport; carbon metabolism</td>
<td>109</td>
<td>62 (56.9)</td>
<td>27 (24.8)</td>
</tr>
<tr>
<td>tRNA</td>
<td>0345</td>
<td>328235</td>
<td>0426</td>
<td>390700</td>
<td>262379</td>
<td>[tRNA][AH][int]</td>
<td>12</td>
<td>copper resistance</td>
<td>148</td>
<td>145 (98.0)</td>
<td>25 (16.9)</td>
</tr>
<tr>
<td>tRNA</td>
<td>0707</td>
<td>673123</td>
<td>0725</td>
<td>683063</td>
<td>9941</td>
<td>[int][AH][none]</td>
<td>1</td>
<td>17</td>
<td>16 (94.1)</td>
<td>5 (29.4)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>0769</td>
<td>717385</td>
<td>0782</td>
<td>730568</td>
<td>13184</td>
<td>[tRNA][none][AH]</td>
<td>2</td>
<td>sulfur metabolism</td>
<td>33</td>
<td>29 (87.9)</td>
<td>0</td>
</tr>
<tr>
<td>tRNA</td>
<td>0786</td>
<td>733379</td>
<td>0426</td>
<td>321727</td>
<td>89331</td>
<td>[tRNA][AH][int]</td>
<td>4</td>
<td>sensors/regulators, transport; carbon metabolism</td>
<td>109</td>
<td>62 (56.9)</td>
<td>27 (24.8)</td>
</tr>
<tr>
<td>tRNA</td>
<td>1157</td>
<td>1065633</td>
<td>1204</td>
<td>1108451</td>
<td>42819</td>
<td>[tRNA][none][AH]</td>
<td>2</td>
<td>16</td>
<td>13 (81.3)</td>
<td>11 (68.8)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>1209</td>
<td>1114522</td>
<td>1243</td>
<td>1136772</td>
<td>22251</td>
<td>[tRNA][none][AH]</td>
<td>2</td>
<td>16</td>
<td>13 (81.3)</td>
<td>11 (68.8)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>1275</td>
<td>1161594</td>
<td>1336</td>
<td>1215994</td>
<td>53491</td>
<td>[tRNA][none][AH]</td>
<td>5</td>
<td>30</td>
<td>21 (33.9)</td>
<td>19 (30.6)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>1341</td>
<td>1219900</td>
<td>1375</td>
<td>125218</td>
<td>32619</td>
<td>[tRNA][AH][none]</td>
<td>1</td>
<td>16</td>
<td>13 (81.3)</td>
<td>11 (68.8)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>1382</td>
<td>1257095</td>
<td>1424</td>
<td>1290282</td>
<td>3188</td>
<td>[tRNA][AH][none]</td>
<td>5</td>
<td>30</td>
<td>21 (33.9)</td>
<td>19 (30.6)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>1586</td>
<td>1452662</td>
<td>1650</td>
<td>1512578</td>
<td>59197</td>
<td>[tRNA][none][AH]</td>
<td>7</td>
<td>30</td>
<td>21 (33.9)</td>
<td>19 (30.6)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>1656</td>
<td>1502641</td>
<td>1633</td>
<td>1548181</td>
<td>27941</td>
<td>[tRNA][none][AH]</td>
<td>0</td>
<td>putative sulfur compound transport/amidase region</td>
<td>28</td>
<td>26 (92.9)</td>
<td>9 (32.1)</td>
</tr>
<tr>
<td>tRNA</td>
<td>1860</td>
<td>1745124</td>
<td>1886</td>
<td>1767524</td>
<td>26001</td>
<td>[tRNA][none][AH]</td>
<td>0</td>
<td>carbon utilisation, transport,25 molybdothionin-related</td>
<td>28</td>
<td>26 (92.9)</td>
<td>9 (32.1)</td>
</tr>
<tr>
<td>tRNA</td>
<td>1917</td>
<td>1800884</td>
<td>1956</td>
<td>1830642</td>
<td>29759</td>
<td>[tRNA][none][AH]</td>
<td>3</td>
<td>16</td>
<td>13 (81.3)</td>
<td>11 (68.8)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>2330</td>
<td>2227106</td>
<td>2333</td>
<td>2244317</td>
<td>17206</td>
<td>[tRNA][none][IS]</td>
<td>1</td>
<td>16</td>
<td>13 (81.3)</td>
<td>11 (68.8)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>2348</td>
<td>2263918</td>
<td>2380</td>
<td>2300422</td>
<td>36505</td>
<td>[tRNA][AH][none]</td>
<td>0</td>
<td>16</td>
<td>13 (81.3)</td>
<td>11 (68.8)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>2551</td>
<td>2460335</td>
<td>2682</td>
<td>2587253</td>
<td>126919</td>
<td>[tRNA][mob-tRNA-AH-</td>
<td>none</td>
<td>]</td>
<td>6</td>
<td>129</td>
<td>127 (98.4)</td>
</tr>
<tr>
<td>tRNA</td>
<td>3361</td>
<td>3293492</td>
<td>3383</td>
<td>3308142</td>
<td>14651</td>
<td>[tRNA][tRNA-AH][int]</td>
<td>0</td>
<td>16</td>
<td>13 (81.3)</td>
<td>11 (68.8)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>4329</td>
<td>4266669</td>
<td>4356</td>
<td>4283718</td>
<td>17050</td>
<td>[tRNA][AH][int]</td>
<td>3</td>
<td>16</td>
<td>13 (81.3)</td>
<td>11 (68.8)</td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>4367</td>
<td>4291118</td>
<td>4487</td>
<td>4393993</td>
<td>102876</td>
<td>[tRNA][none][AH]</td>
<td>5</td>
<td>carbon utilisation, transport,25 molybdothionin-related</td>
<td>28</td>
<td>26 (92.9)</td>
<td>9 (32.1)</td>
</tr>
</tbody>
</table>
gene, is missing [53], consistent with the extensive biochemical studies demonstrating that the main pathway for C1 assimilation in *Methylococcus capsulatus* is the RuMP pathway.

**C1 assimilation and the ethylmalonyl-CoA pathway for glyoxylate regeneration.** The assimilation of C1 units by the serine cycle requires the regeneration of glyoxylate from acetyl-CoA. It has been a long standing puzzle how strain AM1 achieves this given that it lacks isocitrate lyase activity, the key enzyme of the classical glyoxylate regeneration pathway [8]. Indeed, and the corresponding gene was not detected in the *Methylobacterium* genome. Glyoxylate regeneration via the recently elucidated ethylmalonyl-CoA pathway [12] has now been demonstrated in strain AM1 [13], and the corresponding genes have been identified [11,12]. The genomes of the other bacteria compared here present a contrasting picture in this respect. The genome of *S. pomeroyi* also contains a complete set of the genes for the ethylmalonyl-CoA pathway (not shown), and as in *M. extorquens*, these genes are not clustered on the chromosome. In *M.

### Table 3. Cont.

<table>
<thead>
<tr>
<th>Start CDS (META1)</th>
<th>Start (nt)</th>
<th>End CDS (META1)</th>
<th>End (nt)</th>
<th>Length (bp)</th>
<th>[Left border][Right border]</th>
<th>IS</th>
<th>Features (proposed role)</th>
<th>Total CDS (%)</th>
<th>Unique CDS (%)</th>
<th>AH CDS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4495</td>
<td>4400484</td>
<td>4514</td>
<td>4455275</td>
<td>14792</td>
<td>[IS][AH][IS]</td>
<td>2</td>
<td>efflux determinant</td>
<td>20</td>
<td>16 (80.0)</td>
<td>16 (80.0)</td>
</tr>
<tr>
<td>trRNA/4746</td>
<td>4641080</td>
<td>4776</td>
<td>4668490</td>
<td>27411</td>
<td>[trRNA][int-mob-trRNA-AH][none]</td>
<td>0</td>
<td>virulence determinant</td>
<td>30</td>
<td>26 (86.7)</td>
<td>14 (46.7)</td>
</tr>
<tr>
<td>trRNA/5356</td>
<td>5288636</td>
<td>5390</td>
<td>5336993</td>
<td>48358</td>
<td>[trRNA][int][none]</td>
<td>0</td>
<td></td>
<td>35</td>
<td>32 (91.4)</td>
<td>22 (62.9)</td>
</tr>
<tr>
<td>trRNA/5552</td>
<td>5522860</td>
<td>5566</td>
<td>5532024</td>
<td>9165</td>
<td>[trRNA][int-AH][none]</td>
<td>0</td>
<td></td>
<td>16</td>
<td>15 (93.8)</td>
<td>15 (93.8)</td>
</tr>
</tbody>
</table>

*Int: integrase; mob: mobility determinant; GC: region with atypical GC content; AH: region rich in genes detected by Alien Hunter; IS: Insertion Sequence; MITE: Miniature Inverted Repeat Transposable Element.

*No homolog with >80% identity/0.8 minLrap value in the chromosome of the compared strain.

*As detected by Alien Hunter ([46], see Materials and Methods).

*Including one putative MITE.

Figure 4. Clustering and conservation of serine cycle and other genes important for methylotrophic metabolism in sequenced methylotrophic bacteria. Sequences were retrieved from Genbank and visualized using CLC Sequence Viewer 5 (www.clcbio.com). Chromosome sequence positions are indicated, as well as the percent identity at the protein level with *Methylobacterium* prototypes (nd: not detectable). Formate tetrahydrofolate ligase/formyl-tetrahydrofolate synthetase (*ftfL*, black); serine hydroxymethyltransferase (*glyA*, pink); serine glyxoylate aminotransferase (*sga*, yellow); hydroxypyruvate reductase (*hprA*, red); glycerate kinase (*gck*, purple); phosphoenolpyruvate carboxylase (*ppc*, orange); malyl-CoA lyase/b-methylmalyl-CoA lyase (*mcl*, dark green); malate thikokinese (*mtkA/mtkB*, light green); NAD(P)-dependent methylene-tetrahydromethanopterin/methylene-tetrahydrofolate dehydrogenase (*mtdA*, dark blue); methenyl tetrahydrofolate cyclohydrolase (*fch*, light blue); bifunctional methylene-tetrahydrofolate dehydrogenase/methenyl-tetrahydrofolate cyclohydrolase (*folD*, grey); transcriptional regulator (pale green); other (white); trRNA (black rectangles).

doi:10.1371/journal.pone.0005584.t003
doi:10.1371/journal.pone.0005584.g004
Methylobacterium methylotrophic bacteria including unknown function are found nearby serine cycle genes in all genome. However, the genes of several probable regulators of not located in the proximity of known serine cycle genes in the cycle regulator in doi:10.1371/journal.pone.0005584.g005 within each family in AM1 and DM4, respectively.

Figure 5. IS family distribution of intact ISs in M. extorquens AM1 and DM4. The bar length shows the total intact IS copy number of each IS family in DM4 (right) and AM1 (left) (see Suppl. Table S4). Differently colored regions represent different replicons: blue – DM4 chromosome, cyan – DM4 plasmid p1METDI, green – DM4 plasmid p2METDI, pink – AM1 chromosome, orange – AM1 megaplasmid, dark yellow – AM1 plasmid p1META1, light yellow – AM1 plasmid p2META1, light green – AM1 plasmid p3META1. Open circles and squares represent the numbers of different types of ISs present instead [54]. In M. capsulatus, neither ethylmalonyl-CoA pathway nor the isocitrate lyase shunt appear to be encoded within the genome [49], consistent with the operation of the RuMP pathway as the predominant pathway for C1 assimilation in M. capsulatus.

Transcriptional regulation of carbon assimilation in methylotrophic metabolism. The gene of the global serine cycle regulator in Methylobacterium (QscR, a LysR-type regulator homologous to CbR), is essential for methylotrophic growth. It activates transcription of the clustered serine cycle genes as well as of glyA, and negatively regulates its own transcription [75] but it is not located in the proximity of known serine cycle genes in the genome. However, the genes of several probable regulators of unknown function are found nearby serine cycle genes in all methylotrophic bacteria including Methylobacterium discussed here (Fig. 4).

Analysis of IS elements uncovers a significant potential for genome plasticity in Methylobacterium

Methylobacterium genomes display an IS content comparable to that other microbial genomes [76], but with a clear differential distribution of highly diverse IS elements in AM1 and DM4 (Fig. 5, Suppl. Table S4). In AM1, 39 different IS types (defined by a 95% amino acid identity threshold), belonging to 14 IS families (defined as broad groupings of related elements in ISfinder [77]), were detected, compared to 42 IS types belonging to 14 IS families in DM4. Overall diversity of IS types is higher in DM4, but the total number of IS elements in AM1 is twice as high as in DM4 (Table 1). A total of 71 intact and 23 partial IS elements were detected in strain DM4, representing about 2% of the genome (Table 1). With 9 and 7 copies, respectively, ISMex15 and ISMex17 were the two most abundant IS elements in this strain. In comparison, strain AM1 featured 142 intact and 32 partial IS elements, representing 3.7% of the genome (Table 1). At 37, 16 and 23 intact copies, respectively, ISMex1, ISMex2 and ISMex3 of AM1 (with average pairwise nucleotide differences between different gene copies of only 0.01%, 0.11% and 0.03% respectively), the most abundant IS elements identified, may have undergone recent expansion. In addition, one miniature inverted-repeat transposable element (MITE), MiniMd3, was detected in both strains (Suppl. Table S4). This element (~400 bp) is related to ISMd3 but lacks the transposase gene. Few studies so far have identified the presence of both non-autonomous and autonomous transposable elements in the same bacterial genome (see e.g. Out of a total of 70 IS types identified in this work, only 11 IS types are shared between the two strains [Suppl. Table S4, intact IS]. IS5 and IS110 are the most abundant shared IS families, each family featuring 5 to 7 different types of IS (Fig. 5). This suggests that substantial IS loss and/or acquisition has occurred during the relatively short period of time since both strains have emerged from a common ancestor.

The distribution of IS element localization within each genome displays clear-cut, non-random features. Plasmids harbor a higher density of IS elements than the chromosomes. Over 20% of the length of the DM4 plasmid p2METDI and of the AM1 plasmids p1META and p2META encode IS elements. Similarly, IS elements comprise about 8% of the length of the AM1 megaplasmid (Table 1), a significantly higher proportion than in the chromosome (x^2 test, p<0.0001). Moreover, several IS families are significantly over-represented on particular replicons. For example, all 16 copies of ISMex2, an IS element belonging to the IS481 family that is specific to strain AM1, are found on its chromosome while all 5 copies of the IS elements belonging to the IS710 family are on the megaplasmid. In contrast, 13 out of the 14 copies of IS elements of this group in DM4 are located on the chromosome. For some IS elements, however, a more homogeneous distribution was noted. For example, the Tri3 family element ISMex22 unique to strain AM1 is found in one copy per replicon. Transposition immunity was described for this type of IS...
element [78], suggesting the occurrence of transposition saturation in this case.

The observed non-random IS density across replicons may be due to one or more of three potential causes: (1) biased transposition rates by different IS types across replicons, such as local hopping or plasmid specificity; (2) biased selective effects of transposition events, such as over-representation in regions with high density of genes with little or no selective value, such as plasmids or IS elements themselves; or (3) insufficient time for reaching equilibrium, e.g. for IS elements acquired via recent plasmid-mediated transmission. A second pattern in the distribution of IS locations was noted within each replicon. There is an over-representation of IS elements by 7-fold and 39-fold in chromosomal regions unique to AM1 and DM4, respectively, relative to the regions shared between the two strains (x^2 test, p<0.0001; also see Fig. 2). These could represent regions with fewer essential genes and therefore relaxed selection against DNA insertions. Alternatively, they could have been IS-rich regions dating back to the common ancestor of these two strains. This would have then led to increased rates of deletion between two co-directional copies of the same IS element, causing such IS-rich regions to be lost more frequently.

**IS elements linked to methylotrophy.** The two strain-specific methylotrophy regions containing mau (in AM1) and dcm (in DM4) gene clusters (Table 3) are closely associated with IS elements. In strain DM4, genes demR and demA are embedded within several overlapping IS elements ([79], Table 3 and see below, Fig. 6). In strain AM1, the mau cluster (12 kb) lies between 2 copies of ISMex15 (~30 kb), as part of a larger (approx. 66 kb) gene cluster unique to this strain (Table 3). This suggests that such methylotrophy-associated gene clusters may be prone to lateral gene transfer and/or deletion. Indeed, it has been shown recently that the presence of the mau gene cluster is variable in closely related environmental strains of *Methylotenera*, a betaproteobacterial methylotroph [80]. This phenomenon may be involved in the emergence of new ecotypes of methylotrophs.

The genomic island for DCM utilization: a new type of mobility determinant?

Unlike methylvamine and methanol which are produced naturally in large amounts [2,81], DCM is produced naturally at low levels only [33], and presumably occurs at significant concentrations in the environment due to industrial production. The dcm genomic island unique to strain DM4 with the dcmA gene encoding DCM dehalogenase required for growth of *Methylobacterium* with DCM is located on the chromosome (Table 3, Fig. 6), just 20 genes downstream of the large conserved 49 kb methylotrophy gene cluster (Fig. 2, Suppl. Table S1). This 126 kb DNA region, of markedly different GC content (60.5%) from the genome average,
was most likely acquired by horizontal transfer. The sequences upstream and downstream of the unique \( dcm \) region are in complete synteny between the genomes of strains DM4 and AM1. The integration point of the \( dcm \) region features the 5’-end and 3’-end of a “chelatase-like” (COG0606, predicted ATPase with chaperone activity). Although most currently known genomic islands are located at the 3’ end of a tRNA locus, other genes serving as integration sites have been described, such as the gln (glutamate racemase) gene of the \( H. pylori \) pathogenicity island [82]. Clues on the mode of integration of the \( dcm \) region within the \( M. extorquens \) chromosomal framework were obtained by a more detailed analysis. The first CDS within the \( dcm \) region encodes a putative recombinase. Arrangements of non-overlapping 5’ and 3’ fragments of such a “chelatase” gene bordering an internal DNA fragment beginning with a recombinase gene are also evident in three other published complete genomes (Table 4). Additional DNA motifs associated with such structures include 5–20 bp direct repeats and palindromic sequences located immediately up- and downstream of the 5’- and 3’-fragments of the disrupted gene, respectively (Table 4). DNA sequences encoding “chelatase” homologs are often apparent pseudogenes, partial sequences, or sequences containing one or several internal stop codons, suggesting that such sequences may have experienced insertion and subsequent excision of DNA fragments. It is tempting to speculate that such sequences represent novel determinants of genome plasticity, and we propose the term “island integration determinant” (iid) to describe them.

The \( dcm \) region features only few genes that can be associated with confidence with methylotrophic metabolism (Fig. 6). The majority of the genes within this region (74/128, 58%) are hypothetical or conserved hypothetical proteins (compared to the chromosomal average of 41.1% and plasmid average of 43.3% for such proteins, Table 1). Several genes of the \( dcm \) region are interrupted by IS elements (e.g. a glutathione S-transferase METDI2660/2663), or are present in truncated form (e.g. a DNA helicase METDI2648). Many CDS seem associated with DNA modification, stability and mobility. Moreover, 7 IS elements were identified in this region, with 4 in close proximity to \( dcmA \) [79]. The structural elements of a \( tona file repABC \) plasmid [83], i.e. a canonical \( repABC \) operon encoding plasmid replication and maintenance function with its counter-transcribed small RNA in divergent orientation upstream of \( repC \), and a palindromic 16 nt sequence GTTCTCAGCTGAGAAC fitting the \( par \) binding site consensus sequence [83] upstream of \( repA \), were also found within the \( dcm \) region. The 8 kb region centered around \( repABC \) displays extensive synteny with several rhizobial plasmids and with several regions on the chromosome of \( N. hamburgensis \) X14 [84]. This suggests that part or all of the \( dcm \) region may have once existed as an extrachromosomal element and contributed to the spread of the metabolic capacity to degrade DCM in the environment. Nevertheless, introduction of the \( dcmA \) gene into strain AM1, with expression of active DCM dehalogenase at high levels, failed to enable growth on DCM [85]. Thus, specific adaptations are required beyond the presence of DCM dehalo-

---

Table 4. Chelatase-like Island integration determinant (iid) associated with genomic islands in completed microbial genomes.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>M. strain DM4</th>
<th>M. extorquens PA1</th>
<th>Mesorhizobium loti</th>
<th>Nitrobacter hamburgensis X14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome accession number</td>
<td>FP103042</td>
<td>NC_010172</td>
<td>NC_002678</td>
<td>NC_007964</td>
</tr>
<tr>
<td>Disrupted CDS *</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5’-end fragment</td>
<td>METDI2550 (100)</td>
<td>Mext1904 (99)</td>
<td>Mii4733 (70)</td>
<td>Nhamic144 (75)</td>
</tr>
<tr>
<td>Integration position</td>
<td>329</td>
<td>215</td>
<td>214</td>
<td>214</td>
</tr>
<tr>
<td>3’-end fragment</td>
<td>METDI2684 (99)</td>
<td>Mext1923 (98)</td>
<td>Mii4667 (70)</td>
<td>Nhamic130 (73)</td>
</tr>
<tr>
<td>Associated recombinase (Rec)</td>
<td>METDI2551 (+, 100)</td>
<td>Mext1905 (+, 28)</td>
<td>Mii4668 (+, 28)</td>
<td>Nhamic143 (+, 28)</td>
</tr>
<tr>
<td>Genomic island organization ^</td>
<td>DR/DR2/PAL/Rec-[insert]-PAL/DR</td>
<td>DR/PAL-Rec-[insert]-PAL/DR</td>
<td>DR/PAL1-[insert]-Rec-PAL2/DR</td>
<td>DR-PAL-Rec-[insert]-PAL/DR</td>
</tr>
<tr>
<td>Direct repeat (DR) sequence(s)</td>
<td>GAACC (DR)</td>
<td>TGCTGATGA</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>Palindromic (PAL) sequence(s)</td>
<td>ATTCCCCACCTT&gt; X&lt;AAAGTTGGGGAAT</td>
<td>ATGACGTGGGCTATT&gt; X&lt;AAATGGCCACGTCA</td>
<td>GCCACAATCT&gt;GCTA&lt;ACATTTGTGGC (PAL1)</td>
<td>GTGATGCTACATAA&gt; X&lt;TTAATGTAATCAC</td>
</tr>
<tr>
<td>Associated genomic island</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size (kb)</td>
<td>126.4</td>
<td>18.4</td>
<td>50.5</td>
<td>17</td>
</tr>
<tr>
<td>%GC d</td>
<td>60.5 (68.1)</td>
<td>69.7 (68.2)</td>
<td>56.2 (62.8)</td>
<td>58.2 (61.7)</td>
</tr>
<tr>
<td>Number of CDS</td>
<td>127</td>
<td>18</td>
<td>65</td>
<td>13</td>
</tr>
<tr>
<td>Proposed role</td>
<td>DCM degradation</td>
<td>Arsenite resistance</td>
<td>unknown</td>
<td>unknown</td>
</tr>
</tbody>
</table>

*CDS number, % identity at protein level and genomic island integration position relative to the intact island integration determinant CDS META1_1797 of M. extorquens AM1.

^CDS number, orientation relative to the upstream island integration determinant and % identity at protein level with the recombinase of strain DM4.

D: direct repeat; PAL: palindrome; Rec: recombinase. Slash indicates sequence overlap, > and < indicate end and begin of minor palindromic sequences. In M. loti, the two palindromic sequence pairs are only separated by three and four bases, respectively. Overlapping segments of direct repeat and palindromic sequences are indicated in bold or underlined (5’-end fragment METDI2684 (99) Mext1923 (98) Mii4667 (70) Nhamic130 (73)) associated with genomic islands in completed microbial genomes.

*Genome %GC content given in brackets.

doi:10.1371/journal.pone.0005584.t004
genase to enable *Methylobacterium* to grow with this compound [36]. Additional genetic determinants needed for growth with DCM remain to be discovered, and the availability of genomic sequences will facilitate experimental efforts towards identifying them.

**Conclusions**

The assembled and complete genome sequences of two strains representing the pink-pigmented facultative methylotrophs of the genus *Methylobacterium* reveal extensive genome-wide homology and gene syntenies. Genomic determinants of methylotrophy are almost identical between the two strains, with the exception of the methylamine utilization cluster unique to strain AM1 and of the DCM utilization cluster unique to strain DM4. Still, the two strains differ in genome size and number of replicons, and feature a set of strain-specific genes, mostly of unknown function. The large number and extensive diversity of IS elements in *Methylobacterium* genomes, along with the often clustered organization of genes for utilization of C1 compounds, suggests that genome rearrangements and horizontal gene transfer most often associated with IS elements, represent key mechanisms of *Methylobacterium* evolution relating to growth-supporting nutrients and environmental conditions. The co-linearity of the two genomes and the absence of substantial large-scale sequence rearrangements are all the more striking in this context, and may indicate that purifying selection sets strong constraints against major alterations of the genome structure in *Methylobacterium*, despite the long laboratory history of the two strains, usually grown with different carbon sources (methanol for strain AM1 and DCM for strain DM4). These two genome sequences thus afford a refined picture of the potential of *Methylobacterium* for physiological flexibility and adaptation to specific environmental constraints within a conserved genomic framework, and provide the basis for renewed, systems level experimental investigations.

**Materials and Methods**

**Sequencing, assembly, and validation of the genome of**

*M. extorquens* AM1

Sequence data were obtained by whole genome shotgun sequencing as previously described [86]. BigDye terminator chemistry and capillary DNA sequencers (model 3700, Applied Biosystems) were used. Randomly picked blunt end-cloned small insert pUC19 vector-based plasmids (average ~3 kb insert size) were sequenced at both ends using universal forward and reverse sequencing primers, according to standard protocols established at the University of Washington Genome Center. In addition, a large insert fosmid library was constructed from SmalA partial-restricted genomic DNA cloned in *BamH*1 digested pFOS1 vector. About 1,920 randomly picked fosmid clones were end-sequenced and the data pooled with the small insert shotgun sequence data. Sequence data were assembled and visualized using Phred/Phrap/Consed software (www.phrap.com). The sequence quality and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87]. Manual finishing and assembly was improved by carrying out several rounds of experiments designed by the Autofinish tool in Consed [87].

The complete sequence of the genome of strain DM4 was obtained using three different libraries. Genomic DNA was fragmented by mechanical shearing, and 3 kb (A) and 10 kb (B) inserts were cloned, respectively, into plasmid vectors pNAV (a pcDNA2.1 Invitrogen derivative) and pCNS (a pSU18 derivative). In addition, a large insert BAC library (25 kb inserts, C) was constructed from SmalA partially digested total DNA by cloning into pBeloBAC11. Plasmid DNAs were purified and end-sequenced (79200 (A), 27648 (B), 13056 (C) paired match end-reads, respectively) using dye-terminator chemistry on ABI3730 sequencers. Assembly was realized as described [88] with Phred/Phrap/Consed software package (www.phrap.com). An additional 2170 sequences from selected clones were used in the finishing phase of assembly.

**Genome annotation and bioinformatic analysis**

Coding sequences were predicted using the AMIGene (Annotation of Microbial Genomes) software [89] and then submitted to automatic functional annotation using the set of tools listed in [45]. Putative orthology relationships between the two genomes were defined by gene pairs satisfying either the Bidirectional Best Hit criterion [90] or an alignment threshold (at least 40% sequence identity over at least 80% of the length of the smallest protein). These relationships were subsequently used to search for conserved gene clusters (synteny groups) among several bacterial genomes using an algorithm based on an exact graph-theoretical approach [91]. This method allowed for multiple correspondences between genes, detection of paralogy relationships, gene fusions, and chromosomal rearrangements (inversion, insertion/deletion). The ‘gap’ parameter, representing the maximum number of consecutive genes that are not involved in a synteny group, was set to five.

Manual validation of automatic annotations was performed in a relational database (MethylobacScope, https://www.genoscope.cns.fr/age/mage/wwwpklldb/Login/log.php?pid=26) using the MaGe web interface [45], which allows graphic visualization of
the annotations enhanced by a synchronized representation of synteny groups in other genomes chosen for comparison. Genomes were checked for the presence of genes without homologs in the parent genome using thresholds of 80% sequence identity threshold at the protein level and 80% of the length of the shorter homolog (minLgap 0.5). Chromosomal genes of potentially foreign origin were detected using Alien Hunter [46]. Potential genomic islands were searched for with the RGP [Region of Genomic Plasticity] tool of the Mage web-based interface [45] based on synteny breaks between compared genomes, and then checked the predicted regions manually. Only regions larger than 8 kb are reported here.

IS annotations were done by in-house computational tools (Robinson, Lee, Marx, unpublished) that incorporated IScan [92], followed by manual validation based on ISfinder [77]. IS elements were given names of type “ISMEx”, with “Mex” (for M. extorquens) and “Mdi” (for Methylobacterium degrading dichloromethane) indicating strains AM1 or DM4, respectively. The same type name was used for both strains for IS elements with >95% identity in protein sequence. An intact copy was defined as a sequence whose length was at least 99% of the length of the longest copy detected, and a partial IS was defined as a >500 bp fragment with >80% DNA identity to an intact copy.

**Supporting Information**

<table>
<thead>
<tr>
<th>Table S1</th>
<th>Methylotrophy genes in M. extorquens AM1 and DM4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Found at: doi:10.1371/journal.pone.0005584.s001</td>
<td>(0.07 MB DOC)</td>
</tr>
</tbody>
</table>

**References**


