Title: Whole-Genome Analysis of Methyl *tert*-Butyl Ether (MTBE)-Degrading Beta-Proteobacterium *Methylibium petroleiphilum* PM1

Running Title: Genome Sequence Analysis of *Methylibium petroleiphilum* PM1

Staci R. Kane[1*†], Anu Y. Chakicherla[1*], Patrick S. G. Chain[1,4], Radomir Schmidt[2], Maria W. Shin[1], Tina C. Legler[1], Kate M. Scow[2], Frank W. Larimer[3,4], Susan M. Lucas[4], Paul M. Richardson[4], and Krassimira R. Hristova[2]

[1]Lawrence Livermore National Laboratory, Livermore, CA; [2]Dept of Land Air and Water Resources, University of California, Davis, CA; [3]Genome Analysis Group, Oak Ridge National Laboratory, Oak Ridge, TN; [4]Joint Genome Institute Production Genomics Facility, Walnut Creek, CA.

[*]Designates equal contribution to first authorship, [†]Corresponding author

Email: kane11@llnl.gov

Address: Lawrence Livermore National Laboratory,
7000 East Avenue, L-542,
Livermore, CA, 94550

Phone: 925-422-7897 Fax: 925-422-3800

Abstract:

Methylibium petroleiphilum PM1 is a methylotroph distinguished by its ability to completely metabolize the fuel oxygenate methyl *tert*-butyl ether (MTBE). Strain PM1 also degrades aromatic (benzene, toluene, xylene) and straight chain (C_5-C_{12}) hydrocarbons present in petroleum products. Whole genome analysis of PM1 reveals a ~4-Mb circular chromosome and ~600-kb megaplasmid containing 3831 and 646 genes, respectively. Aromatic hydrocarbon and alkane degradation, metal resistance, and methylotrophy are encoded on the chromosome. The megaplasmid contains an unusual t-RNA island, numerous insertion sequences and large repeated elements including a 40-kb region also present on the chromosome and a 29-kb tandem repeat encoding phosphonate transport and cobalamin biosynthesis. The megaplasmid also codes for alkane degradation and was shown to play an essential role in MTBE degradation through plasmid curing experiments. Discrepancies between the IS element distribution pattern, the distribution of best BLASTP hits among major phylogenetic groups, and G+C content of the chromosome (69.2%) and plasmid (66%) together with comparative genome hybridization experiments suggest the plasmid was recently acquired and apparently carries the genetic information responsible for PM1's ability to degrade MTBE. Comparative genomic hybridization analysis with two PM1-like MTBE-degrading environmental isolates (~99% identical 16S rDNA sequences) showed that this plasmid was highly conserved (ca. 99% identical), whereas, the chromosomes were too diverse to conduct resequencing analysis. PM1's genome sequence provides a foundation to investigate MTBE biodegradation and explore genetic regulation of multiple biodegradation pathways in *M. petroleiphilum* and other MTBE-degrading beta-proteobacteria.

Methylibium petroleiphilum strain PM1, a newly described genus and species (57), is a motile bacterium belonging to the Comamonadaceae family of the beta-Proteobacteria and an important member of subsurface microbial communities in many gasoline-contaminated aquifers. Furthermore, PM1 is a methylotroph that can grow aerobically on the fuel oxygenate methyl *tert*-butyl ether (MTBE) and oxidize it completely to carbon dioxide (9, 35). MTBE is a suspected carcinogen that has contaminated drinking water wells throughout the US due to the preponderance of underground leaking storage tanks, the widespread usage of MTBE and its recalcitrance and mobility in groundwater. PM1 can also oxidize aromatic hydrocarbons (toluene, benzene, o-xylene, and phenol) (20) and *n*-alkanes (C_5-C_{12}) (57; K. Hristova, unpublished data), and has been used in two bioaugmentation field trials in gasoline-contaminated aquifers in California (71) and Montana (18, 73). In contaminated sites amended with oxygen, *in situ* MTBE degradation was observed and corresponded to increases in native populations of *Methylibium* sp. (~99% similarity to PM1 based on 16S rDNA) (40, 71, 81). PM1-like bacteria occur naturally in a number of MTBE-contaminated aquifers in the US (47, 48, 82), Mexico (21) and Europe (55, 61), and their presence has been correlated with MTBE degradation activity in numerous sites (48, 71, 82) using real-time PCR analysis (38). These results suggest that PM1-like organisms may play a major role in MTBE biodegradation under aerobic conditions in contaminated aquifers. The genetic basis for MTBE metabolism is not currently understood although there is general agreement that the initial enzymatic steps are similar to co-metabolic degradation pathways (27, 68, 74), and recent reports have described genes involved in degradation of MTBE downstream metabolites, 2-methyl-1,2-propanediol (29) and 2-hydroxyisobutyrate (61). The complex regulation of the metabolism of fuel hydrocarbons and MTBE, often occurring in mixtures, is relatively unknown, while limited studies showed that MTBE degradation could be inhibited in mixtures with BTEX compounds (20, 48).

In this paper, we present our analysis of the whole genome sequence of *M. petroleiphilum* PM1. We present comparative sequence analysis results between PM1 and bacteria with homologous individual genes and operons as well as comparative whole genomic hybridization analysis between PM1 and PM1-like MTBE-degrading isolates (~99% identical 16S rDNA sequences) from gasoline-contaminated sites.

General genome features are discussed including interesting repeated elements, as well as genes and operons involved in methylotrophy, degradation of aromatic hydrocarbons, degradation of cyclic and straight-chain alkanes, cofactor biosynthesis, motility, secretion, and heavy metal resistance and transport. A noteworthy finding was the presence of a large ~600 bp plasmid in PM1 that was highly conserved among PM1-like bacteria. Furthermore, plasmid-curing experiments showed that the plasmid was essential for MTBE and TBA biodegradation in PM1. The PM1 genome sequence has provided a foundation for understanding novel pathways and interactions in this important subsurface bacterium as well as those in phylogenetically similar MTBE-degrading bacteria.

MATERIALS AND METHODS

Bacterial strains used in genome sequence and comparative hybridization analyses. *Methylibium petroleiphilum* strain PM1 was used for whole genome sequencing at the Joint Genome Institute (Walnut Creek, CA). Strain PM1 was isolated from a sewage treatment plant biofilter used for treating discharge from oil refineries (34,9). Two MTBE-degrading bacterial pure cultures (MG4 and 312) were obtained from two different gasoline-contaminated aquifers in Northern California (47; Travis Air Force Base, Travis, CA and San Mateo, CA, respectively). Enrichment culturing was conducted in 10 mg/L MTBE mineral salts media (MSM; 56) with shaking at 150 rpm at room temperature. Enrichment cultures were plated onto 0.1X trypticase soy agar (TSA), and individual colonies were picked and grown in MSM with 10 mg/L MTBE and analyzed for MTBE degradation activity using purge-and-trap gas chromatography/mass spectrometry with reference to d_{12}-MTBE as an internal standard (47). Culture purity was tested by plating (0.1X TSA) and 16S rDNA sequence analysis of colony genomic DNA.

Sequencing, gene prediction and annotation. Genomic DNA was isolated and purified from *M. petroleiphilum* PM1 and whole genome shotgun libraries (3-kb, 8-kb, and 40-kb

DNA inserts) were constructed and sequenced as previously described (14). After quality control of the 90,327 total initial reads of draft sequence, 83,180 sequences were assembled, producing an average of 10.7-fold coverage across the genome. The whole genome sequence was assembled using the Phred/Phrap/Consed package (P. Green, University of Washington) (25, 26, 33). The reads were assembled into 24 high-quality draft sequence contigs, which were linked into 3 larger scaffolds by paired-end sequence information. Gaps in the sequence were closed by either walking on gap-spanning clones or with PCR products generated from genomic DNA. Physical (un-captured) gaps were closed by combinatorial (multiplex) PCR. Sequence finishing and polishing added 308 reads, and the final assessment of the genome assembly was completed as described previously (14). The final genome assembly quality of PM1 adheres to conventional standards of less than one error per 10000 bp. Each base is covered by at least 2 quality sequences, with an average of 10.7 fold coverage. Proper assembly was verified by fosmid coverage coupled with PCR data. Gene modeling and genome annotation was performed as previously described (14) to identify open reading frames likely encoding proteins (coding sequences [CDS]).

Nucleotide sequence annotation and accession number. The annotation is available on the Joint Genome Institute web-site (http://genome.ornl.gov/microbial/rgel/) and has been deposited in the GenBank/EMBL database under accession number NZ_AAEM00000000.

Comparative genomics. Orthologs and CDSs unique to *M. petroleiphilum* PM1 were identified using the Integrated Microbial Genomes (IMG) system from the Joint Genome Institute. Results were based on BLASTP analysis with cutoff values of $E < 10^{-5}$ and 30% identity for orthologs and $E < 10^{-2}$ and 20% identity for unique CDSs.

Phylogenetic tree analysis. Homologs of *M. petroleiphilum* PM1 translated CDSs were identified using BLASTP searches against the non-redundant (NR) GenBank database from National Center for Biotechnology Information. Sequences were aligned and alignments were refined using ClustalX version 1.8 (42) along with manual adjustments.

The *protdist* program and the *neighbor* program of the Phylip package (28) were used to generate the phylogenetic tree for MpeA3393. The pairwise parameters included gap opening = 35 and gap extension = 0.75. The multiple alignment parameters included gap opening = 15, gap extension = 0.3, delay divergence = 30%, and Gonnet series for the protein weight matrix.

Comparative Genomic Hybridization and Comparative Genomic Sequencing analyses. Comparative Genomic Hybridization (CGH) was conducted in order to analyze conservation of genes from MTBE-degrading isolates MG4 and 312 with PM1 across the entire genome. High-density arrays (~400,000 oligomers) were designed and produced by NimbleGen Systems, Inc. (Madison, WI) using 29-mer probes every 26 bp for both strands of the entire genome and every 7 bases for both strands of the plasmid. Arrays were hybridized with labeled genomic preparations of MG4, 312 and PM1. Genomic DNA was isolated (5), and digested (5 μg per array) with 0.005 U DNase I (Amersham) in 1X One-Phor-All buffer (Amersham, Piscataway, NJ) at 37°C for 5 min with subsequent inactivation (95°C for 15 min). To the DNA digest were added 4 μL 5X Terminal Transferase Buffer, 1 nmol Biotin-N6-ddATP, and 25 U Terminal Transferase. The sample was incubated at 37°C for 90 min followed by inactivation at 95°C for 15 min. Hybridization of arrays was conducted in 1 X hybridization buffer for 16 hr at 45°C using a Hybriwheel device (NimbleGen). PM1 was used as a reference in the analysis and was hybridized to separate arrays. Duplicate arrays were processed per strain. Arrays were washed with non-stringent wash buffer (6X SSPE, 0.01% [v/v] Tween-20) followed by two 5 min washes with stringent buffer (100 mM MES, 0.1 M NaCl, 0.01% [v/v] Tween-20) at 47.5°C. Arrays were stained with Cy3-streptavidin conjugate (Amersham Piscataway, NJ) for 10 min followed by washing in nonstringent buffer. Signal amplification was achieved by secondary labeling with biotinylated goat anti-streptavidin (Vector Laboratories, Burlingame CA), washing in nonstringent buffer and restaining with Cy3-streptavidin. Finally, arrays were washed in non-stringent wash buffer, in 0.5 X SSC two times for 30 sec and in 70% ethanol for 15 sec. Arrays were spun dry by centrifugation. Scanning was conducted at 5-⌠m resolution with a Genepix 4000b scanner (Axon Instruments, Union City CA), and NimbleScan software (NimbleGen) was

used to obtain pixel intensities. For higher resolution resequencing of the MG4 and 312 plasmids, arrays were synthesized and hybridized with genomic DNA from each strain and scanned as above. Single nucleotide polymorphism (SNP) positions were tested for uniqueness in the genome using custom algorithms (NimbleGen). The PM1 annotation (http://genome.ornl.gov/microbial/rgel/) was used to generate the output file in SignalMap analysis software (NimbleGen). The predicted SNPs were confirmed by producing and sequencing amplicons using PCR primers located external to the SNP location.

Random mutagenesis, mutant characterization and plasmid curing of PM1. In order to label the megaplasmid with a selectable marker, random transposon mutagenesis was employed using the mini transposon derivative, pTnMod-SmO containing the streptomycin/spectinomycin adenylyltransferase gene (*aadA*) and an oriR origin of replication between the inverted repeats (22). Electrocompetent PM1 cells were prepared by culturing in 0.5X Tryptic Soy Broth (TSB) at 27°C with shaking to log phase. Cells were collected by centrifugation, washed in 10% glycerol four times, and suspended in 10% glycerol to a final volume of 100 ⌈l. A mixture of 50 ⌈l cells and 2 ⌈l pTnMod-SmO DNA (1 µg/µl) was electroporated in 0.1 mm gap cuvettes using 1.8 kV, 200 Ohms, and 25 ⌈F capacitance settings (22) in a BioRad Gene Pulser Electroporator (BioRad, Hercules, CA). Following a 4 h recovery in 0.5X TSB at 27°C with shaking, transposon mutants were selected on 0.5X TSA plates with 50 ⌈g/ml streptomycin (Sm). Sm-resistant colonies were present after incubation for 5-6 days at 27°C and stable transposon integration was confirmed by PCR analysis of genomic DNA using pTnMod-SmO specific primers.

Using the rapid cloning strategy outlined by Dennis et al. (22), the <SmO> insert location was mapped in several PM1 subclones containing the oriR within the transposon. Briefly genomic DNA was extracted, digested with *Ava* II restriction endonuclease, self-ligated, and transformed into *E. coli* TOP10 cells (Invitrogen, Carlsbad, CA). The resulting transformants were selected on LB agar containing 50 ⌈g/ml Sm. Sequencing with primers against the ends of the <SmO> insert was used to determine the exact insert location. One transposon-mutant MP0005 was shown to have

the <SmO> insert on the megaplasmid (in MpeB636). MP0005 was subjected to plasmid curing by heat stress as described by Trevors (78). Specifically, strain MP0005 was incubated at 37°C for 6-8 h before plating on 0.5X TSA. Following replica plating on 0.5X TSA with and without 50 ⌈g/ml Sm, Sm-sensitive colonies were selected and megaplasmid loss was confirmed by PCR analysis. MTBE and TBA degradation activity by strain MP0005 and a megaplasmid-free strain MP0007 were determined by gas chromatography analysis as previously described (35).

Results and Discussion

General genome features of chromosome and megaplasmid. The *Methylibium petroleiphilum* strain PM1 genome consists of a circular chromosome of 4,044,225 bp (Figure 1a.), and a megaplasmid (pPM1) of 599,444 bp (Figure 1b.) (Table 1). The genome encodes 4,477 putative CDSs, of which 964 are unique to PM1 based on BLASTP searches against NR. The pPM1-encoded proteins account for a disproportionately large number (382) of these unique genes. Of the remaining proteins, 2801 could be assigned a putative function based on the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. Analysis of the top BLAST hits (against completed genomes in KEGG) revealed the closest homolog was most often found in other beta-proteobacterial genomes (2332), with the most hits (790) to *Ralstonia solanacearum* followed by *Burkholderia pseudomallei* (497) and *Azoarcus* sp. EbN1 (413). This distribution appears to reflect that of the chromosome: 2210, 589 and 364 to beta-, gamma-, and alpha-proteobacteria, respectively (Table 1). Interestingly, in contrast to the chromosome where beta- and gamma-proteobacteria account for 57.7% and 15.4% of its top BLAST hits respectively, the distribution of top hits between beta- (18.9%) and gamma- (15.6%) proteobacteria is nearly equivalent on the megaplasmid. The lower fraction of beta-proteobacteria-like CDSs in the megaplasmid is balanced by the large proportion of CDSs with no hits to KEGG genomes (47.7%) compared with the CDSs on the chromosome (9.9%). This surprising difference in the phylogenetic distribution of best hits together with the discrepancy in G+C content between the plasmid (66%) and the chromosome (69.2%) points to the likelihood that the plasmid was horizontally

acquired; further evidence for this statement is provided by conservation of the megaplasmid in other phylogenetically similar MTBE-degrading bacteria (discussed in detail later). Analysis of Clusters of Orthologous Genes (COG) distribution (77) showed that the most abundant groups (excluding no COG or general function) were amino acid transport and metabolism (7.0%), energy production (6.4%), and transcription (6.3%) on the chromosome, and replication, recombination and repair (8.0%), coenzyme transport and metabolism (7.0%), and inorganic ion transport and metabolism (5.3%) on the plasmid.

The chromosome contains a single ribosomal *rrn* operon (16S-tRNA^{ala}-tRNA^{ile}-23S-5S) and all genes coding for ribosomal proteins. Structural RNA genes for SRP RNA, rnpB, and tmRNA were present. Forty-two tRNA genes, evenly distributed on the chromosome (with the exception of a few clusters of 2 or 3 tRNAs), correspond to 40 tRNA acceptors and can recognize all possible codons. A very unusual feature of pPM1 is that it contains a single large cluster of 27 additional tRNA genes (25 are redundant with those on the chromosome, the two others do not have clear anticodons). This is the first report of such a large tRNA gene island, and the first report of such a cluster on a plasmid. The role of this island in translation, in genome evolution, or in positive selection of the plasmid in this or other bacterial strains is unclear.

Cell motility, secretion and transport systems. *M. petroleiphilum* PM1 possesses the genes necessary for flagellar biosynthesis (for one polar flagellum), chemotactic response, type IV pili synthesis, the type II secretion pathway as well as several genes related to the *Agrobacterium tumefaciens* type IV secretion pathway (Figure 2; Supplemental Table 1). Type IV secretion mechanisms are often involved in pathogenesis. However, homologs to only three of the five core type IV secretion proteins (VirB9, 10, 11, not VirB4 or 7) (6) were identified, so it is unclear at present if PM1 possesses a functional type IV secretion pathway. PM1 likely moves both by flagellar-facilitated swimming and pili-facilitated twitching motility. Three copies of *tra* genes on pPM1 suggest that PM1 may be capable of conjugative transfer, a possibility currently under investigation. Thirteen chromosomal and one plasmid born methyl-accepting chemotaxis proteins (MCP's) allow PM1 to respond to a range of

environmental stimuli (Supplemental Table 1). As in other organisms (83), MCP's in PM1 are found scattered throughout the genome. Only three MCP's are found within taxis operons – the *pilG-L* operon, the *cheYA*(MCP)*W* operon and the *flg/flh* gene cluster. The apparent mobility of MCP's, together with the fact that six PM1 MCP's appear to be paralogs, complicate function assignation. Nevertheless, there are five other MCP's in addition to those already mentioned whose gene environment may offer insight into their possible functions: two paralogous MCP's are located immediately downstream of and may be part of the same operon as the two toluene/benzene monooxygenase pathways; one of the *aer*-like MCP's is immediately downstream of, and in the same orientation as the LysR-type activator of the ribulose 1,5-bisphophate carboxylase/oxygenase (RuBisCO) operon, which is upstream of the two regulatory genes; one MCP may be co-transcribed with a gene showing similarity to the direct oxygen sensor *dos* of *E. coli*; and one MCP may be co-transcribed with a gene showing low percent similarity to bacteriophytochromes and motility sensors. Neighbor-joining analysis of the putative PM1 MCP's against their homologs showed that eight MCP's cluster close to MCP1-4 of *E. coli*/ MCPA-D of *S. typhimurium*, two appear related to the aerotaxis and energy sensor AER of *E. coli*, and one is very similar to the twitching motility protein PilJ of *P. aeruginosa.*

Strain PM1 has two sets of genes coding for form I RuBisCO, *cbbL* (*mpeA1478* and *mpeA2782*) and *cbbS* (*mpeA1479* and *mpeA2783*) and associated enzymes required for CO_2 fixation via the Calvin cycle (Supplemental Table 1); however, this activity has not been demonstrated for PM1. A thorough search of the PM1 genome revealed the absence of key enzymes from each of the three other known CO_2 fixation pathways: 2-oxoglutarate:ferredoxin oxidoreductase and ATP citrate lyase (reductive TCA cycle); the acetyl-CoA synthase/CO dehydrogenase (reductive acetyl-CoA pathway); malonyl-CoA reductase and propionyl-CoA synthase (3-hydroxypropionate cycle) (4, 41). This strain possesses several ABC transporters for transport of inorganic ions such as nitrate, sulfate, magnesium, potassium, phosphate, phosphonate, as well as amino acids, branched chain amino acids, carbohydrates, long chain fatty acids, dipeptides/oligopeptides, polyamines, and antibiotics (Figure 2; Supplemental Table 2). In addition, putative

regulatory/signaling proteins, and cytochromes (based on CXXCH motifs) have been identified (Supplemental Table 3).

Repeated elements. The genome has a number of complex repetitive elements, including eight families of insertion sequences (ISmp1-8) (up to 12 copies) and two large genomic segments (29 and 40 kb) flanked by IS elements that have undergone what appears to be recent duplications (Figure 1). The two replicons do not equally share the repeated insertion elements; five of the eight families are located only on the chromosome and one family is strictly found on the plasmid. The distribution patterns of the IS elements, lends support to the dissimilar phylogenetic distribution of best KEGG hits among sequenced genomes and strengthens the notion of the megaplasmid's recent acquisition.

Parallel copies of ISmp8 flank two tandem copies of a 29-kb repeat, each consisting of two operons involved in phosphonate and cobalamin metabolism. The phosphonate operons (*PhnFDC-HtxFGHIJKLMN*) include putative C-P lyase subunits 54-83% similar to those of *Pseudomonas stutzeri* WM88 (81) (Supplemental Table 2). The Htx and Phn C-P lyases support growth on methylphosphonate or additional alkylphosphonates, respectively; growth on these substrates is not yet known for PM1. Also contained in the repeat are cobalamin (vitamin B_{12}) synthesis genes encoding the conversion of uroporphyrinogen III to cobinamide and the synthesis of dimethylbenzimidazole (DMB) in the aerobic pathway for cobalamin biosynthesis (*mpeB437-453*, *B472-488*) (Supplemental Table 1). Downstream of the tandem repeat are the remaining genes (*mpeB509-522*) for the covalent linkage of DMB, cobinamide and a phosphoryl group to complete the cobalamin synthesis pathway. Genes coding for the anaerobic pathway of cobalamin biosynthesis (*cbi* genes) are also present in the *cob* clusters, however, a complete pathway is lacking. PM1 also lacks a *cobG* encoding the monooxygenase that converts precorrin-3 to precorrin-4 in the aerobic pathway, however one or more of the multiple copies of *cbiG* (*mpeB479, 480, 444, 445*) may code for a functional enzyme that performs this reaction without oxygen (60). Cobalt, and cobalamin (vitamin B_{12}) have been shown to enhance PM1's ability to grow on and degrade MTBE and its primary metabolite, *tert*-butyl alcohol (TBA) (K. Hristova, unpublished data) so it is not surprising that multiple copies of genes involved in

cobalamin synthesis are present in PM1 including tandem repeats of *cob* and *cbi* genes. Recently, Rohwerder et al. (61) showed that cobalamin synthesis affected the growth rate on the MTBE metabolites, TBA and 2-hydroxyisobutyrate (2-HIBA) for a beta-proteobacterial MTBE-degrading strain that was phylogenetically similar to PM1 (95.6% identical based on 16S rDNA sequence). In this strain, cobalt or cobalamin was necessary for activity of an enzyme, isobutyryl-CoA mutase involved in metabolism of 2-HIBA (61), and 99% identical homologs to this two-component mutase are present in PM1 (*mpeB538/541*). As mentioned, a relatively large percentage of predicted proteins on the plasmid (7.0%) belong to the COG category for coenzyme transport and metabolism. A cluster of ethanolamine utilization (*eut*) genes (*mpeB0499-502*) found between the tandem repeats and third cobalamin cluster on the plasmid encode putative proteins 48-85% similar to EutJEMN from the cobalamin-dependent ethanolamine utilization pathway of *S. typhimurium* (50). The latter two proteins are homologs of carboxysome shell proteins (CcmKL). While *S. typhimurium* also contains the ethanolamine lyase subunits and regulator (EutBC and EutR) in its *eut* operon, in PM1 these genes (*mpeA2417-8*, *mpeA2415*) are on the chromosome.

The largest (40 kb) repeated element is present on both the plasmid and chromosome, and encodes a putative PinR-like site-specific recombinase, a replicative DNA helicase, 2 putative spoJ-like transcriptional regulators or plasmid partitioning proteins, a spoIIIE-like DNA translocase, a tellurite resistance protein, and many hypothetical products. The presence of the repeat on both replicons suggests a recent duplicative transfer event. Though the types of genes found in this region suggest a plasmid origin, both copies interrupt similar but non-identical copies of *dcd* genes (dCTP deaminase), and the direction of duplicative transfer remains to be proven. Since this duplication, the 40-kb repeat on pPM1 has been interrupted by a transposase between genes *mpeB0184* and *mpeB0187*.

Heavy metal tolerance and metal homeostasis. One interesting outcome of the genome analysis is evidence of PM1's potential resistance to heavy metals suggesting promise in using the organism to treat sites containing mixed wastes. Arsenic extrusion in PM1 is probably mediated by *arsRBC* present in two copies on the chromosome that are 58-81%

similar to each other (*mpeA1581-1584, arsHBCR*; *mpeA2343-4, A2347*, *arsCB* and *arsR*). While some other bacteria have five-gene operons (*arsRDABC*) and use the ArsAB pump, PM1 probably extrudes arsenite by a carrier protein, ArsB alone energized by membrane potential (62), although resistance to arsenic oxyanions needs to be evaluated. ArsC encodes an arsenic reductase responsible for the transformation of As(V) into As(III) and ArsR is a transcriptional repressor that responds to As(III) and Sb(III) (62). The function of the fourth gene product, ArsH, found in several bacteria (*Yersinia enterocolitica*, *Acidothiobacillus ferrooxidans*, *Pseudomonas aeruginosa*, *P. putida* KT2440) still remains unclear (10, 63). Three chromosomal copies of *chrA* (*mpeA2204*, *mpeA2205* and *mpeA2526*), belonging to the CHR family of transporters, may mediate chromate resistance in PM1. One copy of ChrA (*mpeA2204*) is 63 and 61% similar to its homolog in *Dechloromonas aromatica* RCB and *P. putida* KT2440, respectively, although a homologous chromate reductase (ChrR; 43) was not evident in PM1.

Genome analysis revealed 15 copper resistance genes in a large cluster (~14.4 kb at positions 1760297 – 1775675) on the PM1 chromosome with structural analogy to the plasmid-mediated (pMOL30) copper resistance cluster *copOAIPRSFG* in *Ralstonia metallidurans* (54) (Supplemental Table 2). The *CopOAIP-CopRS* cluster in PM1 is likely involved in the efflux of periplasmic copper (analogous to *cop* system of *P. syringae*, *R. metallidurans*, and *R. solanacearum*, and the *cos* system of *E. coli* [13]), whereas, the efflux of cytoplasmic copper is mediated by a P1-ATPase, CopF1F2 (analagous to *R. metallidurans* CopF) (54). The genome of PM1 also has a putative chemiosmotic antiporter efflux system similar to CzcCBA of *R. metallidurans,* conferring resistance to Cd, Zn and Co (54). In addition to *copF*, there are two other genes encoding putative metal-transporting P1-type ATPases, *mpeA2479* and *mpeA3535*. Additional proteins potentially involved in metal transport include NikBCDE for nickel (*mpeA3117-3120*), CbiOQMK for cobalt and nickel (*mpe*A2799-2802), and ModABC (*mpeA3707*, *mpeA3714*, *mpeA3715*) for molybdenum uptake (Supplemental Table 2).

Ferric iron has also been shown to enhance PM1's ability to grow on and degrade MTBE and TBA (K. Hristova, unpublished data) so it is likely that active transport of iron is of particular importance. As with other Gram-negative bacteria, PM1 acquires its iron supply via Fe^{3+}-siderophores. *Fep* genes, which function in the synthesis of

polypeptides required for uptake of ferric enterobactin were identified inside each of the four *cob* operons in PM1 (Supplemental Table 1). *Polaromonas* sp., *R. ferrireducens*, and *M. flagellatus* all have iron transport genes either within or in close proximity to their *cob* operons (data not shown). Minimally, FepABC are required for ferric enterobactin uptake. *MpeA2292* and *mpeA2605* have been annotated as *fepA*, coding for an outer membrane receptor for an iron siderophore, however it is possible that *btuB* genes located near the *febBDC* genes are also involved in iron assimilation. The TonB-dependent energy transduction complex (*tonB*, *exbB* or *tolQ*, and *exbB*; Supplemental Table 1.) coded on the chromosome likely provides the mechanism for active transport of iron siderophores and cobalamin across the outer membrane (8, 38). The PM1 genome encodes about 39 putative proteins involved in iron transport and homeostasis, which implies the importance of iron in its physiology.

Methylotrophy. Methylotrophic metabolism of PM1 is of great interest because formaldehyde and formate are common intermediates of both methanol and MTBE or TBA oxidation by PM1 and other degraders (10, 59). *M. petroleiphilum* PM1 is capable of aerobic growth on methanol, formate, and succinate. Unlike other methylotrophic beta-proteobacteria, PM1 grows on MTBE, toluene, benzene, ethylbenzene, and dihydroxybenzoates as sole carbon sources (57, 59). PM1 possesses genes for the serine cycle and methylotrophy scattered in several different clusters on its chromosome (Figure 1; Table 2). The strain does not grow on methylamine (K. Hristova, unpublished data), lacks a gene encoding the methylamine dehydrogenase (MADH) large subunit, and likely lacks MADH activity. Despite the ability of PM1 to grow on methanol, its genome lacks true homologs of *mxaF* and *mxaI*, known genes coding for the methanol dehydrogenase (MeDH) large and small subunits, present in several methylotrophs known to date. PM1 contains a MeDH-like cluster *XoxF-J* (*mpeA3393-5*) that is present in *Methylobacterium extorquens* AM1 (17), which also contains the true *mxaF* cluster. Comparative sequence analysis of the product of gene *mpeA3393* revealed high similarity to MxaF (74% to *M. extorquens* AM1) and the XoxF homolog present in several non-methylotrophs (77% to *Burkholderia fungorum*). Based on phylogenetic analysis, *MpeA3393* clusters with the MxaF homologs of unknown function from other methylotrophic and non-methylotrophic

Rhizobia and *Burkholderia* spp., while true MeDH large subunits (MxaF) cluster together and are distinct from the MxaF homologs (Figure 3).

A cytochrome c-555 (*mpeA3394*) 56% similar to the c_H cytochrome of *M. capsulatus* Bath (electron donor to the oxidase in methylotrophic bacteria [1]) is found adjacent to the *mpeA3394*. A putative MxaJ/XoxJ (*mpeA3395*) shows 54% similarity with XoxJ from *Paracoccus denitrificans* and 42% similarity with MxaJ from *M. capsulatus* Bath. Five genes (*mpeA3829*, *2585-8*) are involved in the biosynthesis of pyrroloquinoline quinone (PQQ), a cofactor of MeDH as well as quinoprotein ethanol dehydrogenase. A cluster of genes required for MeDH synthesis, *mxaLKCASR* (*mpeA3273-3278*) is also present. To date, none of the gene clusters containing the XoxF homolog have been shown to be involved in methanol oxidation. Therefore, it is possible that a new enzyme is responsible for this function in the beta-proteobacterium *M. petroleiphilum* PM1.

Three different formate dehydrogenases are present in the PM1 genome, with homologs to *M. extorquens* and *M. capsulatus* Bath. The function of the tungsten-dependent formate dehydrogenase *fdh1* (*mpeA0337-339*), NAD-linked formate dehydrogenase *fdh2* (*mpe*A3708-12), and cytochrome-linked formate dehydrogenase *fdh3* (*mpeA1170-71, 1173*) for energy generation during growth on C_1 substrates or for MTBE oxidation needs to be further explored. *Mpe*A3377 coding for a putative ABC-type tungstate transport system permease links gene clusters *fdh1* and *fdh2* (Table 2). The *fdh2* genes in PM1 have the same gene arrangement and significant sequence identity (52-81%) to the NAD-dependent formate dehydrogenase cluster *fdsGBACD* of *Ralstonia eutropha* (58). Pathways involved in metabolism and detoxification of formaldehyde, a central intermediate of both methanol and MTBE degradation by PM1 and other strains (36,58), may also function in MTBE metabolism. PM1 has two pathways for formaldehyde oxidation to CO_2, an H_4MPT-linked metabolic module that includes an archaeal-like gene cluster and an H_4F-linked metabolic module. Recently, phylogenetic analysis of a subset of bacterial and archaeal H_4MPT-linked C_1 transfer genes placed PM1 sequences with other described beta-proteobacteria (45).

Fuel hydrocarbon degradation pathways. PM1 contains an operon (*mpeA0814-0821*) likely encoding for conversion of benzene to phenol (and catechol), and toluene to methylphenol (and methylcatechol) that is 62-74% similar to the benzene monooxygenase pathway in *P. aeruginosa* JI104 (BmoA-D1) (49) and 50-71% similar to the toluene para-monooxygenase (TpMO) pathway in *Ralstonia pickettii* PKO1 (TbuA1UBVA2CX) (76) (Table 3). A second operon (*mpeA2539-2547*) is 55-74% identical to the first operon, however, it likely does not yield a functional monooxygenase since the TbuA1 homolog is interrupted by a transposon insertion and the TbuC homolog is a pseudogene. Both operons have two-component response regulator-sensor histidine kinases upstream and divergently transcribed (*mpeA0811-812*; *mpeA2536-2537*) although *mpeA2537* may be truncated due to the transposon insertion. *MpeA821* encodes a putative TbuX (65% similar to that in PKO1), an outer membrane protein regulated by TbuT and involved in toluene uptake (44). The BMO pathway has been implicated in benzene and toluene degradation (49), as has the TpMO pathway (76). In addition to benzene and toluene, PM1 has been shown to degrade *o*-xylene (19), although the biochemical pathway has not been elucidated to date. It is likely that *m*- and *p*-xylene can also be metabolized via the toluene monooxygenase (TMO) pathway of PM1 as described for PKO1 and other bacteria (30).

M. petroleiphilum PM1 grows on phenol, and two distinct clusters of dimethylphenol (*dmp*)-like genes are present (*mpeA2265-67, 2272-86*; *mpeA3305-13*, *3321-25*) although the latter lacks the key structural gene *dmpP* so may not yield a functional phenol hydroxylase (PH). Gene products from the first cluster *dmpRKLMNOPQBCDEHFGI* are 60-83% similar to those on pVI150 in *Pseudomonas sp*. strain CF600 (66), including a multi-component PH, catechol 2,3-dioxygenase and the meta-cleavage pathway for catechol (Table 3). The second operon has transposon insertions inside *dmpC* and adjacent to *dmpO*. The PH subunits for the two operons are 44-69% similar. The DmpR homologs (*mpeA3310, A2286*) are similar to TbuT (69 and 65%) and may regulate TMO, PH and the meta-cleavage genes, since TbuT was shown to regulate these genes in PKO1 via separate promoters (11). However these operons are located together in PKO1, whereas, they are quite distant in PM1. PM1 can grow on phenol, and based on the presence of a complete *dmp* operon, it can likely degrade

alkylphenols as well, although it is not clear whether PH is essential for methylphenol degradation (as described for *P. stutzeri* OX1 [12]) or whether the TMO alone is capable of converting toluene to methylcatechol (as described for strain PKO1 [30]).

PM1 has nine CDSs encoding putative proteins with varying similarity to cyclohexanone monooxygenases (CHMOs) sometimes referred to as Baeyer-Villiger-type MOs (*mpeB579, B607, B610, A393, A898, A1038, A1351, A2885, and A2915*). Their protein products may play a role in hydroxylation of either alicyclic, aliphatic or aryl ketones to form a corresponding ester, which can easily be hydrolyzed. Alicyclic hydrocarbons represent up to 12% (wt/wt) of total hydrocarbons in petroleum mixtures (American Petroleum Institute). Aryl ketones such as acetophenone can be produced directly from atmospheric breakdown of ethylbenzene (a major petroleum component) or following abiotic conversion of ethylbenzene to ethylphenol (3) and subsequent biological conversion to the ketone. The putative CHMO genes are scattered across the genome and are not present in operons with other genes coding for subsequent metabolism after the MO reaction (i.e., esterases, alcohol and aldehyde dehydrogenases). The CHMOs have a narrow substrate range, possibly explaining the number of different flavoprotein MOs in PM1 with varying levels of similarity with representatives from this class (Table 2); the putative CHMOs in PM1 were more similar to phenylacetone MO (46-67%; ref. 53) than 4-hydroxyacetophenone MO (43-53%; ref. 45). In PM1, the nine Baeyer-Villiger MOs have a putative $NADP^+$-binding site that is 72-88% similar to the proposed site in a CHMO from *Acinetobacter* sp. strain NCIMB 9871 (16).

An alkane monooxygenase pathway on pPM1 may facilitate PM1's growth on *n*-alkanes. In addition, alkane monooxygenase (hydroxylase) has been proposed to play a role in cometabolic MTBE hydroxylation since acetylene, an inactivator of short-chain alkane monooxygenase, was shown to inhibit MTBE degradation (68). In PM1 the hydroxylase subunit, AlkB (*mpeB0606*) is 69% and 66% similar to that of *Alcanivorax borkumensis* AP1 (72) and *P. putida* PGo1 (contained on the OCT plasmid) (79) respectively, and contains all 8 of the conserved His residues observed in other integral membrane binuclear-iron hydrocarbon monooxygenases (34). Also present are two rubredoxin genes (*mpeB0603* and *mpeB0602*), whose products are 76 and 78% similar, respectively, to rubredoxin 3 and 4 in *Gordonia sp*. strain TF6 (32). The rubredoxin (Rd)

coded by *mpeB603* is an AlkG1-type Rd, whereas, that coded by *mpeB602* is an AlkG2-type Rd, based on the CXXCG motif criteria described by van Beilen et al. (80); Only AlkG2-type Rds were shown to be functional in electron transfer from the rubredoxin reductase to alkane hydroxylase. In addition, *mpeB0601* codes for an ATP-dependent transcriptional regulator 38% similar to AlkS from *A. borkumensis* SK2 (36). Separated from the putative *alkS* by three hypothetical genes is a rubredoxin reductase, *alkT* (*mpeB0597*) whose protein product is 49% similar to that in *Gordonia* sp. TF6. PM1 does not appear to possess any long-chain alkane ($>C_{13}$) oxidation pathways such as an alkane dioxygenase (64), P-450 monooxygenase (2) or two alkane hydroxylase complexes (AlkMa and AlkMb) similar to *Acinetobacter* sp. strain M-1 (75), although PM1's single AlkB is 54% similar to both AlkMa and AlkMb. The gene organization of the *alk* operon in PM1 is somewhat similar to that of *Gordonia* sp. TF6 (*alkB2G1G2T*), except a transposase (*mpeB605*) and putative esterase (*mpeB604*) are between *alkB* and *alkG1G2* and as mentioned, a putative *alkS* and three hypothetical genes (*mpeB600-598*) are located between *alkG1G2* and *alkT* in PM1. Homologs to AlkHJKL from *P. putida* GPo1 coding for aldehyde dehydrogenase, alcohol dehydrogenase, acyl CoA synthetase, and outer membrane protein (79), respectively, were not present on the plasmid, although the PM1 chromosome contains homologs to AlkH (*mpeA2324*, 47% similar), AlkJ (*mpeA3803*, 58% similar), AlkK (*mpeA1769*, 71% similar) and AlkL (*mpeA3010*, 49% similar).

In addition, the PM1 chromosome contains a putative propane monooxygenase pathway (*mpeA950-953*) whose predicted proteins are 41-64% identical to PrmABCD in *Gordonia* sp. TY-5, coding for the large hydroxylase subunit, the NADH-dependent acceptor oxidoreductase, the small hydroxylase subunit and the coupling protein, respectively. The Prm complex in strain TY-5 was shown to catalyze the subterminal oxidation of propane yielding 2-propanol (51), while propane oxidation in PM1 is currently under investigation. As for PrmA in *Gordonia* TY-5, a pair of conserved Glu-X-X-His sequences are present in the putative PrmA of PM1 at residues 138-141 and 237-240. The presence of these sequences is consistent with other monooxygenases in the binuclear-iron oxygenase family including soluble methane monooxygenases (23, 67) suggesting PrmA in PM1 may catalyze hydroxylation of propane. Like the operon in

strain TY-5, a chaperone similar to GroEL was adjacent to the *prm* cluster in PM1 (*mpeA954*). Finally, PM1 has homologs to strain TY-5 alcohol dehydrogenases, *adh1* (*mpeA936*) and *adh3* (*mpeA599*) that are 72 and 83% similar, respectively, that may facilitate 2-propanol degradation. The putative monooxygenases in PM1 are summarized in Supplemental Table 3 including methanesulfonate monooxygenase, *msmA* and alkanesulfonate monooxygenase *ssuD*, which are part of *msmABDCEFGHG* and *ssuAADCB* operons, respectively. PM1 may not utilize methanesulfonate since its *msm*A lacks the sequence CXH-X_{26}-CXXH unique to methanesulfonate utilizers (7). In general, PM1 possesses several homologous genes with other soil bacteria including *Gordonia*, *Alcinovorax*, and *Pseudomonas* spp. capable of biodegradation of petroleum hydrocarbons as well as xenobiotic and recalcitrant compounds such as phthalates.

MTBE biodegradation. Though MTBE is a recent anthropogenic contaminant (released within the last 15 years), various microorganisms can utilize the compound for carbon and energy under aerobic conditions (31, 61, 65, 74). *M. petroleiphilum* PM1 is the best characterized of the few bacterial pure cultures reported to grow on and completely degrade MTBE and its daughter product TBA (20, 31, 37, 65). The genetic basis for MTBE and TBA conversion is not known, although different classes of monooxygenases have been proposed to play a role in metabolism or co-metabolism of these compounds (31, 37, 52, 68, 74), including P450-monooxygenase and alkane monooxygenase (hydroxylase) systems, the latter shown to play a role in cometabolic degradation of MTBE by *P. putida* GPo1 (69) and possibly also by *P. mendocina* KR-1 (70). A known inducer of alkane hydroxylase, dicyclopropylketone, was also shown to induce MTBE conversion to TBA in GPo1 (69). As reported above, an alkane MO (AlkB) system was detected in the PM1 genome on the megaplasmid. The AlkB in PM1 is likely involved in MTBE hydroxylation based on similarity to other AlkB proteins in organisms shown to be involved in MTBE degradation. Whereas, the K_s values for MTBE in *n*-alkane grown GPo1 was reported to be quite high (20-40 mM), the apparent half saturation constant for MTBE by PM1 was 88 μM, which is in the range of K_s values for MTBE by butane-degrading bacteria (52). Unlike GPo1 and KR-1, PM1 further degrades TBA, ultimately producing CO_2 and biomass. The putative AlkB in PM1 is proposed to only oxidize

MTBE and not TBA based on kinetics experiments with MTBE- and TBA-grown cells (19). Two separate enzyme systems were also suggested for MTBE and TBA degradation by *Hydrogenophaga flava* ENV735 (37). Because of its potential role in MTBE metabolism, the coding region for AlkB is the focus of current gene knockout studies. Biodegradation of a similar molecule, ethyl-*tert* butyl ether (ETBE), occurs via a cytochrome P450 pathway in *Rhodococcus ruber* IFP2001 (15). However, homologs of protein complexes involved in ETBE degradation from *R. ruber* were not found in PM1; like GPo1 (69), PM1 has not been shown to degrade ETBE.

Many pollutant degradation genes are located on bacterial catabolic plasmids. Significantly, the two strains, MG4 and 312 that are capable of MTBE degradation had a nearly identical plasmid to that of PM1 (ca. 99% identical) as determined by comparative genomic sequencing analysis. The MG4 and 312 plasmids showed only 5 or 4 SNPs respectively relative to PM1 (Table 3). MG4 and 312 plasmids also lacked transposase genes (three copies of Tra5 and transposase-8 and one copy of a DDE-domain transposase) that were present on the PM1 plasmid and a 1.2 kb deletion putatively containing an esterase/lipase gene (*mpeB604*) and a DDE-domain transposase (*mpeB605*) (Table 3). The promoter and coding region for *alkB* (*mpeB606*) did not appear to be affected by this deletion since it is significantly upstream although there was a SNP mapped within *alkB* of MG4 and 312 resulting in a putative amino acid change. As mentioned, two other PM1 plasmid-encoded genes *mpeB541* and *mpeB538* code for putative large and small subunits of isobutyryl-coenzyme A (CoA) mutase, respectively. The plasmids of strains MG4 and 312 also contained identical copies of *mpeB541* and *mpeB538*. These gene products were shown to have 99% identical homologs in *Ideonella* sp. strain L108 predicted to allow conversion of 2-HIBA to 3-hydroxybutyrate in the presence of CoA and ATP (61). It is not known whether these mutase genes are contained on a megaplasmid in L108 although horizontal gene transfer is often evoked when such high similarities in gene sequences between bacteria are observed. In addition to *alkB*, the PM1 plasmid (as well as MG4 and 312 plasmids) contains a gene coding for 3-hydroxybutyryl-CoA dehydrogenase (*mpeB0547*), putatively involved in conversion of 3-hydroxybutyryl-CoA, a proposed MTBE-metabolite (61), to acetoacetyl-CoA.

The role of the megaplasmid in MTBE and TBA degradation was clearly demonstrated by curing experiments. Chemical analysis of MTBE and TBA degradation by the MP0005 parent strain and the MP0007 strain lacking the megaplasmid (as evidenced by PCR analysis and loss of Sm-resistance) showed that only the former was able to degrade MTBE and TBA (Figure 4). This result is consistent with our proposal that key genes involved in both MTBE and TBA degradation are located on the PM1 megaplasmid. Since two different monooxygenases are proposed to be involved in MTBE and TBA degradation (19; K. Hristova, unpublished data), at least some of the required protein subunits for conversion of MTBE to TBA and conversion of TBA to the putative 2-methyl-1,2-propanediol are coded on the megaplasmid. In addition, the possible role of selected pPM1 proteins in MTBE/TBA oxidation, based on cDNA microarray results, is currently under investigation by gene knockout methods. It is noteworthy that the PM1 plasmid did not contain predicted proteins with significant homology to those proposed in degradation of 2-methyl-1,2-propanediol by *Mycobacterium austroafricanum* IFP 2012 (29), although the putative aldehyde dehydrogenase coded by *mpeA1909* on the chromosome was 54% similar to MpdC (hydroxyisobutyraldehyde dehydrogenase) and the alcohol dehydrogenase coded by *mpeA945* was 45% similar to MpdB (2-methyl-1,2-propanediol dehydrogenase). While the relevant alcohol and aldehyde dehydrogenases may be encoded on the chromosome, additional plasmid-encoded dehydrogenases may be more plausible and are currently being investigated for their putative role in the MTBE degradation pathway.

Concluding remarks. Prior to sequencing its genome, it was not known that PM1 possessed a 600-kb megaplasmid, much less that the plasmid contained candidate CDSs coding for the MTBE and TBA monooxygenases and enzymes involved in downstream reactions. It is noteworthy that MTBE-degrading strains from diverse locations including a biofilter treating wastewater in Southern California (PM1) and two distinct aquifers in Northern California (MG4 and 312) possess a nearly identical plasmid. The presence of this highly conserved megaplasmid among PM1-like MTBE-degraders, along with its different G+C content, its unique IS complement and the unique phylogenetic distribution of its gene products, together raise interesting questions concerning

horizontal gene/plasmid transfer and evolution of pathways via plasmid-mediated mechanisms. With the whole genome sequence, putative aromatic hydrocarbon and alkane degradation pathways were also identified providing a basis to study the complex regulation of fuel hydrocarbon degradation in this novel subsurface bacterium; this is important since substrate interactions are expected to influence the success of bioremediation strategies for gasoline-contaminated sites. In addition to comparative genomics approaches, whole genome microarray and 2-D gel electrophoresis experiments are being conducted to identify genes and proteins unique to MTBE degradation. PM1 can serve as a model for other MTBE-degrading methylotrophs such that the knowledge gained from analysis of its genome, transcriptome and proteome can be applied to PM1-like bacteria. An understanding of the MTBE degradation pathway and its regulation will allow for optimization of MTBE bioremediation and the ability to monitor this unique process *in situ* using molecular tools.

ACKNOWLEDGMENTS

We thank Stephanie Malfatti at Lawrence Livermore National Laboratory (LLNL) for assistance with genome finishing and Miriam Land and Loren Hauser in the Genome Analysis Group at Oak Ridge National Laboratory for their assistance with genome annotation. We also thank Hussna Wakily and Victoria Katz at University of California-Davis (UCD) for their work on the random transposon mutagenesis and Janice Wu (UCD) for work on plasmid-cured PM1 strains. In addition, we thank Harry Beller (LLNL) for helpful comments on the manuscript. We acknowledge the University of California Office of the President for funding through the Campus-Laboratory Collaboration Program as well as the LLNL Laboratory-Directed Research and Development Program. This research was also supported by grant number 5 P42 ES04699-16 from the National Institute of Environmental Health Sciences (NIEHS), NIH with funding provided by U.S. EPA. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIEHS, NIH, or EPA. This work was performed under the auspices of the U.S. Department of Energy by the

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC52-06NA25396.

University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

REFERENCES

1. Afolabi, P. R., F. Mohammed, K. Amaratunga, O. Majekodunmi, S. L. Dales, R. Gill, D. Thompson, J. B. Cooper, S. P. Wood, P. M. Goodwin, and C. Anthony. 2001. Site-directed mutagenesis and X-ray crystallography of the PQQ-containing quinoprotein methanol dehydrogenase and its electron acceptor, cytochrome c_L. Biochem. **40**:9799-9809.

2. Asperger, O., A. Naumann, and H.-P. Kleber. 1984. Inducibility of cytochrome P-450 in *Acinetobacter calcoaceticus* by *n*-alkanes. Appl. Microbiol. Biotechnol. **19**:3948-4403.

3. Atkinson, R. 1995. Gas phase tropospheric chemistry of organic compounds. Volatile Organic Compounds in the Atmosphere Issues. Environ. Sci. Technol. **4:**65-89.

4. Atomi, H. 2002. Microbial enzymes involved in carbon dioxide fixation. J. Biosci. Bioengineer. **94**:497-505.

5. Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Stuhl. 1987. Current protocols in molecular biology. Wiley, New York.

6. Backert, S., and T. F. Meyer. 2006. Type IV secretion systems and their effectors in bacterial pathogenesis. Curr. Opin. Microbiol. **9**:207-217.

7. Baxter, N. J., J. Scanlan, J. De Marco, A. P. Wood, and J. C. Murrell. 2002. Duplicate copies of genes encoding methanesulfonate monooxygenase in *Marinosulfonomonas methylotropha* strain TR3 and detection of

methanesulfonate utilizers in the environment. Appl. Environ. Microbiol. **68**:289-296.

8. Braun, V. and M. Braun. 2002. Iron transport and signaling in *Escherichia coli*. FEBS Lett. **529**:78-85.

9. Bruns, M. A., J. R. Hanson, J. Mefford, and K. M. Scow. Isolate PM1 populations are dominant and novel methyl *tert*-butyl ether-degrading bacteria in compost biofilter enrichments. Environ. Microbiol. **3**:220-225.

10. Butcher, B. G., S. M. Deane, and D. E. Rawlings. 2000. The chromosomal arsenic resistance genes of *Thiobacillus ferrooxidans* have an unusual arrangement and confer increased arsenic and antimony resistance to *Escherichia coli*. Appl. Environ. Microbiol. **66**:1826-1833.

11. Byrne, A. M., R. H. Olsen. 1996. Cascade regulation of the toluene-3-monooxygenase operon (*tbuA1UBVA2C*) of *Burkholderia pickettii* PKO1: Role of the *tbuA1* promoter (P*tbuA1*) in the expression of its cognate activator, TbuT. *J. Bacteriol.* **178**:6327-6337.

12. Cafaro, V., V. Izzo, R. Scognamiglio, E. Notomista, P. Capasso, A. Casbarra, P. Pucci, and A. Di Donato. 2004. Phenol hydroxylase and toluene/*o*-xylene monooxygenase from *Pseudomonas stutzeri* OX1: Interplay between two enzymes. Appl. Environ. Microbiol. **70**:2211-2219.

13. Cervantes, C., and F. Gutierrez-Corona. 1994. Copper resistance mechanisms in bacteria and fungi. FEMS Microbiol. Rev. **14**:121-137.

14. Chain, P., J. Lamerdin, F. Larimer, W. Regala, V. Lao, M. Land, L. Hauser, A. Hooper, M. Klotz, J. Norton, L. Sayavedra-Soto, D. Arciero, N. Hommes, M. Whittaker, and D. Arp. 2003. Complete genome sequence of the ammonia-

oxidizing bacterium and obligate chemolithoautotroph *Nitrosomonas europaea*. J. Bacteriol. **185**:2759-2773.

15. Chauvaux, S., F. Chevalier, C. Le Dantec, F. Fayolle, I. Miras, F. Kunst, and P. Beguin. 2001. Cloning of a genetically unstable cytochrome P-450 gene cluster involved in degradation of the pollutant ethyl *tert*-butyl ether by *Rhodococcus ruber*. J. Bacteriol. **183**:6551-6557.

16. Chen, Y.-C., O. P. Peoples, and C. T. Walsh. 1988. *Acinetobacter* cyclohexanone monooxygenase: gene cloning and sequence determination. J. Bacteriol. **170**:781-789.

17. Chistoserdova, L., and M. E. Lidstrom. 1997. Molecular and mutational analysis of a DNA region separating two methylotrophy gene clusters in *Methylobacterium extorquens* AM1. Microbiol. **143**:1729-1736.

18. Davis-Hoover, W. J., S. A. Stavnes, J. J. Fleischman, S. C. Hunt , J. Goetz, M. Kemper, M. Roulier, K. Hristova, K. Scow, K. Knutson, W. R. Mahaffey, and D. J. Slomczynski. 2003. BTEX/MTBE bioremediation: Bionets containing Isolite, PM1, SOS or air. E-25. In: V.S. Magar and M.E. Kelley (Eds.) Proceedings of the Seventh International In Situ and On-site Bioremediation Symposium. Battelle Press, Columbus, OH.

19. Deeb, R.A., S. Nishino, J. Spain, H. Y. Hu, K. M. Scow, and L. Alvarez-Cohen. 2000. MTBE and benzene biodegradation by PM1 via two independent monooxygenase-initiated pathways. Abstracts Am. Chem. Soc. 219:ENVR 228.

20. Deeb, R. A., H.-Y. Hu, J. R. Hanson, K. M. Scow, and L. Alvarez-Cohen. 2001. Substrate interactions in BTEX and MTBE mixtures by an MTBE-degrading isolate. Environ. Sci. Technol. **35**:312-317.

21. De Marco, P., C. Pachecoa, A. Figueiredoa, and P. Moradas-Ferreira. 2004. Novel pollutant-resistant methylotrophic bacteria for use in bioremediation. FEMS Lett. **234**:75-80.

22. Dennis, J. J, and J. G. Zylstra. 1998. Plasposons: Modular self-cloning minitransposon derivatives for rapid genetic analysis of gram-negative bacterial genomes. Appl. Environ. Microbiol. **64**:2710-2715.

23. Elango, N., R. Radhakrishnan, W.A. Froland, B. J. Wallar, C. A. Earhart, J. D. Lipscomb, and D. H. Ohlendorf. 1997. Crystal structure of the hydroxylase component of methane monooxygenase from *Methylosinus trichosporium* OB3b. Protein Sci. **6**:556-568.

24. Elkins, M. F., and C. F. Earhart, 1989. Nucleotide sequence and regulation of the *Escherichia coli* gene for ferrienterobactin transport protein FepB. J. Bacteriol. **171:**5443-5451.

25. Ewing, B., and P. Green. 1998a. Basecalling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8**:186-194.

26. Ewing, B., L. Hillier, M. Wendl, and P. Green. 1998b. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8**:175-185.

27. Fayolle, F., J.-P. Vandecasteele, and F. Monot. 2001. Microbial degradation and fate in the environment of methyl *tert*-butyl ether and related fuel oxygenates. Appl. Microbiol. Biotechnol. **56**:339–349.

28. Felsenstein, J. 2002. PHYLIP (Phylogeny Inference Package) version 3.6a3. Department of Genome Sciences, University of Washington, Seattle, Washington.

29. Ferreira, N. L., D. Labbe, F. Monot, F. Fayolle-Guichard, and C. W. Greer. 2006. Genes involved in the methyl *tert*-butyl ether (MTBE) pathway of *Mycobacterium austroafricanum* IFP 2012. Microbiol. **152**:1361-1374.

30. Fishman, A., Y. Tao, and T. K. Wood. 2004. Physiological relevance of successive hydroxylations of toluene *para*-monooxygenase of *Ralstonia pickettii* PKO1. Biocat. Biotrans. **22**:283-289.

31. François, A., H. Mathis, D. Godefroy, P. Piveteau, F. Fayolle, and F. Monot. 2002. Biodegradation of methyl *tert*-butyl ether and other fuel oxygenates by a new strain, *Mycobacterium austroafricanum* IFP 2012. Appl. Environ. Microbiol. **68**:2754-2762.

32. Fujii, T., T. Narikawa, K. Takeda, and J. Kato. 2004. Biotransformation of various alkanes using the *Escherichia coli* expressing an alkane hydroxylase system from *Gordonia sp*. TF6. Biosci. Biotechnol. Biochem. **68**:2171-2177.

33. Gordon, D., C. Abajian, and P. Green. 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8**:195-202.

34. Hamamura, N., C. M. Yeager, and D. J. Arp, 2001. Two distinct monooxygenases for alkane oxidation in *Nocardioides* sp. strain CF8. Appl. Environ. Microbiol. **67**:4992-4998.

35. Hanson, J. R., C. E. Ackerman, and K. M. Scow. 1999. Biodegradation of methyl *tert*-butyl ether by a bacterial pure culture. Appl. Environ. Microbiol. **65**:4788-4792.

36. Hara, A., S. H. Baik, K. Syutsubo, N. Misawa, T. H. Smits, J. B. van Beilen, and S. Harayama. 2004. Cloning and functional analysis of *alkB* genes in *Alcanivorax borkumensis* SK2. Environ. Microbiol. **6**:191-197.

37. Hatzinger, P.B., K. McClay, S. Vainberg, M. Tugusheva, C. W. Condee, and R. J. Steffan. 2001. Biodegradation of methyl *tert*-butyl ether by a pure bacterial culture. Appl. Environ. Microbiol. **67**:5601-5607.

38. Higgs, P. I., P. S. Myers, and K. Postle. 1998. Interaction in the TonB-dependent energy transduction complex: ExbB and ExbD form homomultimers. J. Bacteriol. **180**:6031-6038.

39. Hristova, K.R., C. M. Lutenegger, and K. M. Scow. 2001. Detection and quantification of MTBE-degrading strain PM1 by real-time TaqMan PCR. Appl. Environ. Microbiol. **67**:5154-5160.

40. Hristova, K., B. Gebreyesus, D. Mackay, and K. M. Scow. 2003. Naturally occurring bacteria similar to the methyl *tert*-butyl ether (MTBE)-degrading strain PM1 are present in MTBE-contaminated groundwater. Appl. Environ. Microbiol. **69**:2616-2623.

41. Hugler, H., H. Huber, K. O. Stetter, and G. Fuchs. 2003. Autotrophic CO_2 fixation pathways in archaea (Crenoarchaeota). Arch. Microbiol. **179**:160-173.

42. Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson. 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. **23**:403-405.

43. Jimenez, J. I., B. Minambres, J. L. Garcia, and E. Diaz. 2002. Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. Environ. Microbiol. **4**:824-841.

44. Kahng, H.-Y., A. M. Byrne, R. H. Olsen, and J. J. Kukor. 2000. Characterization and role of *tbuX* in utilization of toluene by *Ralstonia pickettii* PKO1. J. Bacteriol. **182**:1232-1242.

45. Kalyuzhnaya, M. G., M. E. Lidstrom, and L. Chistoserdova. 2005. Analysis of gene islands involved in methanopterin-linked C1 transfer reveals new functions and provides evolutionary insights. J. Bacteriol. **187**:4607-4614.

46. Kamerbeek, N. M., M. J. H. Moonen, J. G. M. van der Ven, W. J. H. van Berkel, M. W. Fraaije, and D. B. Janssen. 2001. 4-Hydroxyacetophenone monooxygenase from *Pseudomonas fluorescens* ACB. A novel flavoprotein catalyzing Baeyer-Villiger oxidation of aromatic compounds. Eur. J. Biochem. **268**:2547-2557.

47. Kane, S. R., H. R. Beller, T. C. Legler, C. J. Koester, H. C. Pinkart, R. U. Halden, and A. M. Happel. 2001. Aerobic biodegradation of methyl *tert*-butyl ether by aquifer bacteria from leaking underground storage tank sites. Appl. Environ. Microbiol. **67**:5824-5829.

48. Kane, S. R., T. C. Legler, L. M. Balser, and K. T. O'Reilly. 2003. Aerobic biodegradation of MTBE by aquifer bacteria from LUFT sites. E-12. In: V.S. Magar and M.E. Kelley (Eds.) Proceedings of the Seventh International In Situ and On-site Bioremediation Symposium. Battelle Press, Columbus, OH.

49. Kitiyama, A., E. Suzuki, Y. Kawakami, and T. Nagamune. 1996. Gene organization and low regiospecificity in aromatic-ring hydroxylation of a benzene monooxygenase of *Pseudomonas aeruginosa* JI104. J. Ferment. Bioeng. **82**:421-425.

50. Kofoid, E., C. Rappleye, I. Stojiljkovic, and J. Roth. The 17-gene ethanolamine (*eut*) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. *J. Bacteriol.* **181**:5317-5329.

51. Kotani, T., T. Yamamoto, H. Yurimoto, Y. Sato, and N. Kato. 2003. Propane monooxygenase and NAD^+-dependent secondary alcohol dehydrogenase in propane metabolism by *Gordonia sp*. strain TY-5. J. Bacteriol. **185**:7120-7128.

52. Liu, C. Y., G. E. Speitel, Jr., and G. Georgiou. 2001. Kinetics of methyl *t*-butyl ether cometabolism at low concentrations by pure cultures of butane-degrading bacteria. Appl. Environ. Microbiol. **67**:2197–2201.

53. Malito, E., A. Alfieri, M. W. Fraaije, and A. Mattevi. 2004. Crystal structure of a Baeyer-Villiger monooxygenase. Proc. Natl. Acad. Sci. U.S.A. **101**:13157-13162.

54. Mergeay, M., S. Monchy, T. Vallaeys, V. Auquier, A. Benotmane, P. Bertin, S. Taghavi, J. Dunn, D. van der Lelie, and R. Wattiez. 2003. *Ralstonia metallidurans*, a bacterium specifically adapted to toxic metals: towards a catalogue of metal-responsive genes. FEMS Microbiol. Rev. **27**:385-410.

55. Moreels, D., L. Bastiaens, F. Ollevier, R. Merckx, L. Diels, and D. Springael. 2004. Evaluation of the intrinsic methyl *tert*-butyl ether (MTBE) biodegradation potential of hydrocarbon contaminated subsurface soils in batch microcosm systems. FEMS Microbiol. Ecol. **49**:121-128.

56. Mu, D. Y., and K. M. Scow. 1994. Effect of trichloroethylene (TCE) and toluene concentrations on TCE and toluene biodegradation and the population density of TCE and toluene degraders in soil. Appl. Environ. Microbiol. **60**:2661-2665.

57. Nakatsu, C. H., K. R. Hristova, S. Hanada, X.-Y. Meng, J. R. Hanson, K. M. Scow, and Y. Kamagata, 2006. *Methylibium petroleiphilum* PM1 gen. nov., sp.

nov., a new methyl *tert*-butyl ether (MTBE) degrading methylotroph belonging to the beta-subclass of the Proteobacteria. Int. J. Sys. Evol. Microbiol. **56**:983-989.

58. Oh, J.-I., and B. Bowien. 1998. Structural analysis of the *fds* operon encoding the NAD+-linked formate dehydrogenase of *Ralstonia eutropha.* J. Biol. Chem. **273**:26349-26360.

59. Piveteau, P., F. Fayolle, J.-P. Vandecasteele, and F. Monot. 2001. Biodegradation of *t*-butyl alcohol and related xenobiotics by a methylotrophic bacterial isolate. Appl. Microbiol. Technol. **55**:369-373.

60. Rodionov, D. A., A. G. Vitreschak, A. A. Mironov, and M. S. Gelfand. 2003. Comparative genomics of the vitamin B_{12} metabolism and regulation in prokaryotes. J. Biol. Chem. **278**:41148-41159 (2003).

61. Rohwerder, T., U. Breuer, D. Benndorf, U. Lechner, and R. H. Muller. 2006. The alkyl *tert*-butyl ether intermediate 2-hydroxyisobutyrate is degraded via a novel cobalamin-dependent mutase pathway. Appl. Environ. Microbiol. **72**:4128-4135.

62. Rosen, B.P. 2002. Biochemistry of arsenic detoxification. FEBS Lett. **529**:86-92.

63. Ryan, D., and E. Colleran. 2002. Arsenic resistance in the IncH12 plasmids. Plasmids **47**:234-240.

64. Sakai, Y., J. H. Maeng, Y. Tani, and N. Kato. 1994. Use of long-chain *n*-alkanes (C_{13}-C_{44}) by an isolate, *Acinetobacter sp*. M-1. Biosci. Biotechnol. Biochem. **58**:2128-2130.

65. Salanitro, J. P., L. A. Diaz, M. P. Williams, and H. L. Wisnieski. 1994. Isolation of a bacterial culture that degrades methyl *t*-butyl ether. Appl. Environ. Microbiol. **60**:2593-2596.

66. Shingler, V., J. Powlowski, and U. Marklund. 1992. Nucleotide sequence and functional analysis of the complete phenol/3,4- dimethylphenol catabolic pathway of *Pseudomonas sp*. strain CF600. J. Bacteriol. **174**:711-724.

67. Smith, T. J., S. E. Slade, N. P. Burton, J. C. Murrell, and H. Dalton. 2002. Improved system for protein engineering of the hydroxylase component of soluble methane monooxygenase. Appl. Environ. Microbiol. **68**:5265-5273.

68. Smith, C. A., K. T. O'Reilly, and M. R. Hyman. 2003. Characterization of the initial reactions during the cometabolic oxidation of methyl *tert*-butyl ether by propane-grown *Mycobacterium vaccae* JOB5. Appl. Environ. Microbiol. **69**:796-804.

69. Smith, C.A., and M. R. Hyman. 2004. Oxidation of methyl *tert*-butyl ether by alkane hydroxylase in dicyclopropylketone-induced and *n*-octane-grown *Pseudomonas putida* GPo1. Appl. Environ. Microbiol. **70**:4544-4550.

70. Smith, C.A., K. T. O'Reilly, and M. R. Hyman. 2004. Cometabolism of methyl *tertiary* butyl ether by *Pseudomonas mendocina* KR-1 grown on C_5 to C_8 *n*-alkanes. Appl. Environ. Microbiol. **69**:7385-7394.

71. Smith, A. E. K. Hristova, I. Wood, D. M. Mackay, E. Lory, D. Lorenzana, and K. M. Scow. 2005. Comparison of biostimulation versus bioaugmentation with bacterial strain PM1 for treatment of groundwater contaminated with methyl tertiary butyl ether (MTBE). Environ. Health Prospect. **113**:317-332.

72. Smits, T. H., S. B. Balada, B. Witholt, and J. B. van Beilen. 2002. Functional analysis of alkane hydroxylases from gram-negative and gram-positive bacteria. J. Bacteriol. **184**:1733-1742.

73. Stavnes S. A., J. Fleischman, J. Goetz, K. Hristova, S. Hunt, M. Kemper, K. Knutson, W. Mahaffee, M. Roulier, K. Scow, D. J. Slomczynski, and W. J. Davis-Hoover. 2002. MTBE bioremediation with BioNets containing Isolite, PM1, SOS or air. 2B-66. In: A. R. Gavaskar and A.S.C. Chen (eds.), Proceedings of the Third International Conference of Chlorinated and Recalcitrant Compounds. Battelle Press, Columbus, OH.

74. Steffan, R. J., K. McClay, S. Vainberg, C. W. Condee, and D. Zhang. 1997. Biodegradation of the gasoline oxygenates methyl *tert*-butyl ether, ethyl *tert*-butyl ether, and *tert*-amyl methyl ether by propane-oxidizing bacteria. Appl. Environ. Microbiol. **63**:4216-4222.

75. Tani, A., T. Ishige, Y. Sakai, and N. Kato. 2001. Gene structures and regulation of the alkane hydroxylase complex in *Acinetobacter* sp. strain M-1. J. Bacteriol. **183**:1819-23.

76. Tao, Y., A. Fishman, W. E. Bentley, and T. K. Wood. 2004. Oxidation of benzene to phenol, catechol, and 1,2,3-trihydroxybenzene by toluene-4-monooxygenase or *Pseudomonas mendocina* KR1 and toluene 3-monooxygenase of *Ralstonia picketti* PKO1. Appl. Environ. Microbiol. **70**:3814-3820.

77. Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. **28**:33-36.

78. Trevors, J. T. 1986. Plasmid curing in bacteria. FEMS Microbiol. Rev. **32**:149-157.

79. van Beilen, J. B., S. Panke, S. Lucchini, A. G. Franchini, M. Rothlisberger and B. Witholt. 2001. Analysis of *Pseudomonas putida* alkane degradation gene clusters

and flanking insertion sequences: evolution and regulation of the *alk* genes. Microbiol. **147**:1621–1630.

80. van Beilen, J. B., M. Neuenschwander, T. H. M. Smits, C. Roth, S. B. Balada, and B. Witholt. 2002. Rubredoxins involved in alkane oxidation. J. Bacteriol. **184**:1722–1732.

81. White, A. K., W. W. Metcalf. 2004. Two C-P lyase operons in *Pseudomonas stutzeri* and their roles in the oxidation of phosphonates, phosphite, and hypophosphite. J. Bacteriol. **186**:4730-4739.

82. Wilson, R. D., D. M. Mackay, and K. M. Scow. 2001. In situ MTBE biodegradation supported by diffusive oxygen release. Environ. Sci. Technol. **36**:190-199.

83. Zhulin, I. B. 2001. The superfamily of chemotaxis transducers: From physiology to genomics and back. Adv. Microbial Physiol. **45:**157-198.

1

TABLE 1. General features of the *Methylibium petroleiphilum* PM1 genome[1]

Feature	Chromosome	Megaplasmid
Total bases	4,044,225	599,444
CDS, total	3831	646
CDS, percent	92.4	86.3
Average CDS size, bp	1059	951
RNA operons (16S-23S-5S)	1	0
tRNAs	42	27
GC percent	69.20	65.98
Repeated IS elements	25	11
Alpha-proteobacterial KEGG hits	364 (9.5%)	31 (4.8%)
Beta-proteobacterial KEGG hits	2210 (57.7%)	122 (18.9%)
Gamma-proteobacterial KEGG hits	589 (15.4%)	101 (15.6%)
No hits in KEGG genome DB	378 (9.9%)	308 (47.7%)
Conserved hypothetical	684 (17.9%)	59 (9.1%)
Hypothetical (Unique)	582 (15.2%)	382 (59.1%)
Total proteins with EC assignment	682	

2 [1] The BLASTP threshold e-score was 1e-05.

3

1 **TABLE 2**. *M. petroleiphilum* PM1 genes putatively involved in metabolism of methanol.

Gene ID	Gene Name	COG no.	Predicted Role
PQQ synthesis cluster			
MpeA3829	*pqqA*		Pyrroloquinoline quinone biosynthesis protein A
MpeA2585	*pqqB*		Pyrroloquinoline quinone biosynthesis protein B
MpeA2586	*pqqC*	COG5424	Pyrroloquinoline quinone biosynthesis protein C
MpeA2587	*pqqD*		Pyrroloquinoline quinone biosynthesis protein D
MpeA2588	*pqqE*	COG0535	Pyrroloquinoline quinone biosynthesis protein E
C1 transfer cluster 1			
MpeA2606	*mptG*	COG1907	synthase, methanopterin biosynthesis
MpeA2607	*mptD*		methylene tetrahydromethanopterin dehydrogenase
MpeA2608		COG2232	ATP-dependent carboligase
MpeA2609	*mch*	COG3252	methenyl-H_4MPT cyclohydrolase
MpeA2610		COG0189?	putative H_4MPT biosynthesis protein
MpeA2611		COG1767	triphosphoribosyl-dephosphoCoA-synthetase
MpeA2612	*fae*	COG1795	formaldehyde-activating enzyme
MpeA2613	*folB*	COG1539	Dihydroneopterin aldolase
MpeA2614		COG3284	putative transcriptional activator
MpeA2615		COG1036	similar to Archael flavoprotein
MpeA2616	*folP*	COG0294	Dihydropteroate synthase
MpeA2617			involved in H_4MPT -linked formaldehyde oxidation
MpeA2618		COG1891	uncharacterized protein in Archaea, involved in denovo synthesis of NAD in prokaryotes & eukaryotes
MpeA2619		COG2054	uncharacterized Archael kinase
MpeA2620		COG1548	putative transcriptional regulator/kinase
MpeA2621		COG1821	putative ATP utilizing enzyme
MpeA2622		COG1411	involved in H_4MPT-linked formaldehyde oxidation
MpeA2623	*fhcB*	COG1029	formyltransferase/hydrolase complex, beta subunit
MpeA2624	*fhcA*	COG1229	formyltransferase/hydrolase complex, alpha subunit
MpeA2625	*fhcD*	COG2037	formyltransferase/hydrolase complex, delta subunit
MpeA2626	*fhcC*	COG2218	formyltransferase/hydrolase complex, gamma subunit
MpeA2627	*fae*	COG1795	putative formaldehyde-activating enzyme, present in Archaea
Primary oxidation Methylamine DH cluster			
MpeA2651		COG4977	Transcriptional regulator
MpeA2652	*cytC*	COG2010	Cytochrome C
MpeA2653	*mauA*		methylamine dehydrogenase small subunit
MpeA2654	*mauD*		methylamine utilization protein
C1 transfer +C1 assimilation cluster 2			
MpeA3251		COG1560	layroyl acyltransferase
MpeA3254	*mcl*		methyltrophic malyl-CoA lyase
MpeA3255	*ppc*	COG2352	methylotrophic phosphoenolpyruvate carboxylase
MpeA3256	*mtkB*	COG0074	malate thiokinase
MpeA3257	*mtkA*		malate thiokinase
MpeA3258	*fch*	COG3404	methenyl tetrahydrofolate cyclohydrolase
MpeA3259	*mtdA*	COG0300	methylene-H_4MPT/ methylene-H_4F dehydrogenase
MpeA3260			hydroxypyruvate reductase
MpeA3261	*sgaA*	COG0075	serine glyoxylate aminotransferase
MpeA3262	*glyA*	COG0112	serine hydroxymethyltransferase
MpeA3263	*gck*	COG2379	glycerate kinase
MpeA3264	*ftfL*	COG2759	formyl-tetrahydrofolate ligase
MpeA3265	*folC*	COG0285	dihydrofolate synthetase

TABLE 2. (continued)

Gene ID	Gene Name	COG no.	Predicted Role
Primary oxidation - Mxa cluster			
MpeA3273	*mxaL*	COG1764	MxaL protein
MpeA3274	*mxaK*		MxaK protein
MpeA3275	*mxaC*	COG2425	MxaC protein
MpeA3276	*mxaA*		MxaA protein
MpeA3277	*mxaS*	COG1721	MxaS protein
MpeA3278	*mxaR*	COG0714	putative MxaR - involved in the regulation of formation of active MDH
MpeA3391		COG1521	Putative transcriptional regulator
MpeA3393	*mxaF?*	COG4993	methanol dehydrogenase large subunit homolog
MpeA3394	*cytC*	COG2010	cytochrome C-555
MpeA3395	*mxaJ*		MxaJ like protein
Primary oxidation - Formate dehydrogenase cluster			
NAD-linked molybdenum dependent formate dehydrogenase			
MpeA3706		COG2998	ABC type tungstate transport system permease
MpeA3708	*fdh2A*	COG1905	formate dehydrogenase gamma subunit
MpeA3709	*fdh2B*	COG1894	formate dehydrogenase beta subunit
MpeA3710	*fdh2C*	COG3383	formate dehydrogenase alpha subunit
Tungsten-containing formate dehydrogenase			
MpeA0337	*fdh1A*	COG3383	formate dehydrogenase alpha subunit
MpeA0338	*fdh1B*	COG1894	formate dehydrogenase beta subunit
Cytochrome-linked formate dehydrogenase			
MpeA1170	*fdh3A*	COG0243	formate dehydrogenase alpha subunit
MpeA1171	*fdh3B*	COG0437	formate dehydrogenase beta subunit
MpeA1173	*fdh3C*	COG2864	formate dehydrogenase gamma subunit
C1 assimilation			
MpeA1014	*mcmB*	COG1884	Methylmalonyl-CoA mutase, beta subunit
MpeA1014	*mcmA*		Methylmalonyl-CoA mutase, alpha subunit
MpeA1015	*meaB*	COG1703	GTPase/ATPase, essential for methylmalonyl-CoA
MpeA1016	*pccB*		mutase reaction
MpeA1017	*pccA*	COG4770	Propionyl-CoA carboxylase, alpha subunit
MpeA1018	*epm*		Methylmalonyl-CoA epimerase
Succinate dehydrogenase cluster			
MpeA2165	*gltA*	COG0372	Citrate (Si)-synthase
MpeA2166		COG2938	
MpeA2167	*sdhB*	COG0479	Succinate dehydrogenase/fumarate reductase Fe-S protein subunit
MpeA2168	*sdhA*	COG1053	Succinate dehydrogenase/fumarate reductase
MpeA2169	*sdhD*	COG2142	Transmembrane succinate dehydrogenase flavoprotein subunit
MpeA2170	*sdhC*	COG2009	Succinate dehydrogenase/fumarate reductase
MpeA2172	*mdh*	COG0039	malate dehydrogenase cytochrome b subunit hydrophobic membrane anchor subunit

TABLE 3. *M. petroleiphilum* PM1 genes putatively involved in metabolism of aromatic hydrocarbons denoting the predicted role of the gene product and percent similarity with well-characterized homologs.

Operon	Gene ID	Gene Name	Predicted Role	% Similarity with PKO1[1]
tbu				
Operon I	MpeA0811		two-component response regulator	
	MpeA0812		two-component sensor histidine kinase	
	MpeA0813		hypothetical	
	MpeA0814	*bmoA/tbuA1*	toluene monooxygenase alpha subunit	84
	MpeA0815	*bmoB/tbuU*	toluene monooxygenase gamma subunit	81
	MpeA0816	*bmoC/tbuB*	toluene monooxygenase ferredoxin	74
	MpeA0817	*bmoD1/tbuV*	toluene monooxygenase activator	76
	MpeA0818	*tbuA2*	toluene monooxygenase beta subunit	82
	MpeA0819	*tbuC*	toluene monooxygenase oxidoreductase	64
	MpeA0820		probable Zn-alcohol dehydrogenase	
	MpeA0821	*tbuX*	toluene facilitator	65
tbu				
Operon II	MpeA2536		two-component response regulator	
	MpeA2537		two-component sensor histidine kinase	
	MpeA2538	*tnp*		
	MpeA2539	*bmoA/tbuA1*	toluene monooxygenase alpha subunit C-terminal end	81
	MpeA2540	*bmoB/tbuU*	toluene monooxygenase gamma subunit	76
	MpeA2541	*bmoC/tbuB*	toluene monooxygenase ferredoxin	76
	MpeA2542	*bmoD1/tbuV*	toluene monooxygenase activator	77
	MpeA2543	*tbuA2*	toluene monooxygenase beta subunit	79
	MpeA2544	*tbuC*	toluene monooxygenase oxidoreductase N-terminal end	69
	MpeA2545	*tbuC*	toluene monooxygenase oxidoreductase C-terminal end	60
dmp				**% Similarity with CF600[2]**
Operon I	MpeA2286	*dmpR*	regulator (HylR family)	65
	MpeA2285	*dmpK*	phenol hydrolase assembly	60
	MpeA2284	*dmpL*	phenol hydrolase beta subunit	63
	MpeA2283	*dmpM*	phenol hydrolase activator	76
	MpeA2282	*dmpN*	phenol hydrolase alpha subunit	77
	MpeA2281	*dmpO*	phenol hydrolase gamma subunit	63
	MpeA2280	*dmpP*	phenol hydrolase reductase	73
	MpeA2279	*aphT*	regulator (LysR family)	
	MpeA2278	*dmpQ*	catechol-2,3-dioxygenase ferredoxin	62
	MpeA2277	*dmpB*	catechol-2,3-dioxygenase	70
	MpeA2276	*aphY*	conserved in some dmp-like operons	
	MpeA2275	*dmpC*	2-hydroxymuconic semialdehyde dehydrogenase	83
	MpeA2274	*dmpD*	2-hydroxymuconic semialdehyde hydrolase	81
	MpeA2273	*dmpE*	2-hydroxypent-2,4-dienoate hydratase	78
	MpeA2272	*dmpH*	4-oxalocrotonate decarboxylase	81
	MpeA2267	*dmpF*	acetaldehyde dehydrogenase (acylating)	69

	MpeA2266	*dmpG*	4-hydroxy-2-isovalerate aldolase	70
	MpeA2265	*dmpI*	4-oxalocrotonate isomerase	73
***dmp* Operon II**	MpeA3328	*dmpC*	2-hydroxymuconic semialdehyde dehydrogenase N-terminal end	72
	MpeA3327	*tnp*	transposase	
	MpeA3326	*dmpC*	2-hydroxymuconic semialdehyde dehydrogenase C-terminal end	83
	MpeA3325	*dmpE*	2-hydroxypent-2,4-dienoate hydratase	73
	MpeA3324	*dmpH*	4-oxalocrotonate decarboxylase	72
	MpeA3323	*dmpI*	4-oxalocrotonate isomerase	66
	MpeA3322	*dmpF*	acetaldehyde dehydrogenase (acylating)	67
	MpeA3321	*dmpG*	4-hydroxy-2-isovalerate aldolase	73
	MpeA3313	*aphT*	regulator (LysR family)	
	MpeA3312	*dmpQ*	catechol-2,3-dioxygenase ferredoxin	63
	MpeA3311	*dmpB*	catechol-2,3-dioxygenase	68
	MpeA3310	*dmpR*	regulator (HylR family)	61
	MpeA3309	*dmpK*	phenol hydrolase assembly	64
	MpeA3308	*dmpL*	phenol hydrolase beta subunit	67
	MpeA3307	*dmpM*	phenol hydrolase activator	76
	MpeA3306	*dmpN*	phenol hydrolase alpha subunit	74
	MpeA3305	*dmpO*	phenol hydrolase gamma subunit	57
	MpeA3304	*tnp*	transposase	

[1] Percent similarity with homologs in *Ralstonia pickettii* PKO1 (Tao et al., 2004; Accession nos. AY541701, AF100782, and U72645).

[2] Percent similarity with homologs in *Pseudomonas sp*. strain CF600 (Shingler et al., 1992; Accession nos. X60657, X60835, X52805, X33263, X60276, and X68033).

TABLE 4. Genomic differences between the plasmid of PM1 with those from isolates MG4 and 312.

Genome Location[a]	DNA Change[b,c]	Gene	pPM1	MG4	312
3921	Δ (1337 bp)	ISMca2 tnp tnp OCP	Not present	Present	Present
168369	Δ (1321 bp)	ISMca2 tnp tnp OCP	Not present	Present	Present
194159	A → G	Upstream of MpeB219	A	G	G
267528	T → G	Upstream of MpeB319	T	G	T
315627	T → G	Upstream of MpeB365	T	G	G
465830	C → G	*cobA* (Mpe515)	His176	Asp176	Asp176
504193	Δ (529 bp)	TIS 1021 tnp protein	Not present	Present	Present
536177	Δ (1370 bp)	ISMca2 tnp tnp OCP	Not present	Present	Present
556849	Δ (1229 bp)	MpeB604-B605 lipase/esterase, tnp-IS5 family	Not present	Present	Present
559059	T → G	*alkB* (MpeB606)	Phe264	Cys264	Cys264
Total SNPs				5	4
Missing genes				9	9

[a] Location in PM1 genome. SNP = single nucleotide polymorphism.

[b] Deletion (Δ) or SNP.

[c] Change relative to PM1 sequence. Predicted amino acid changes are noted where appropriate.

Figure Legends

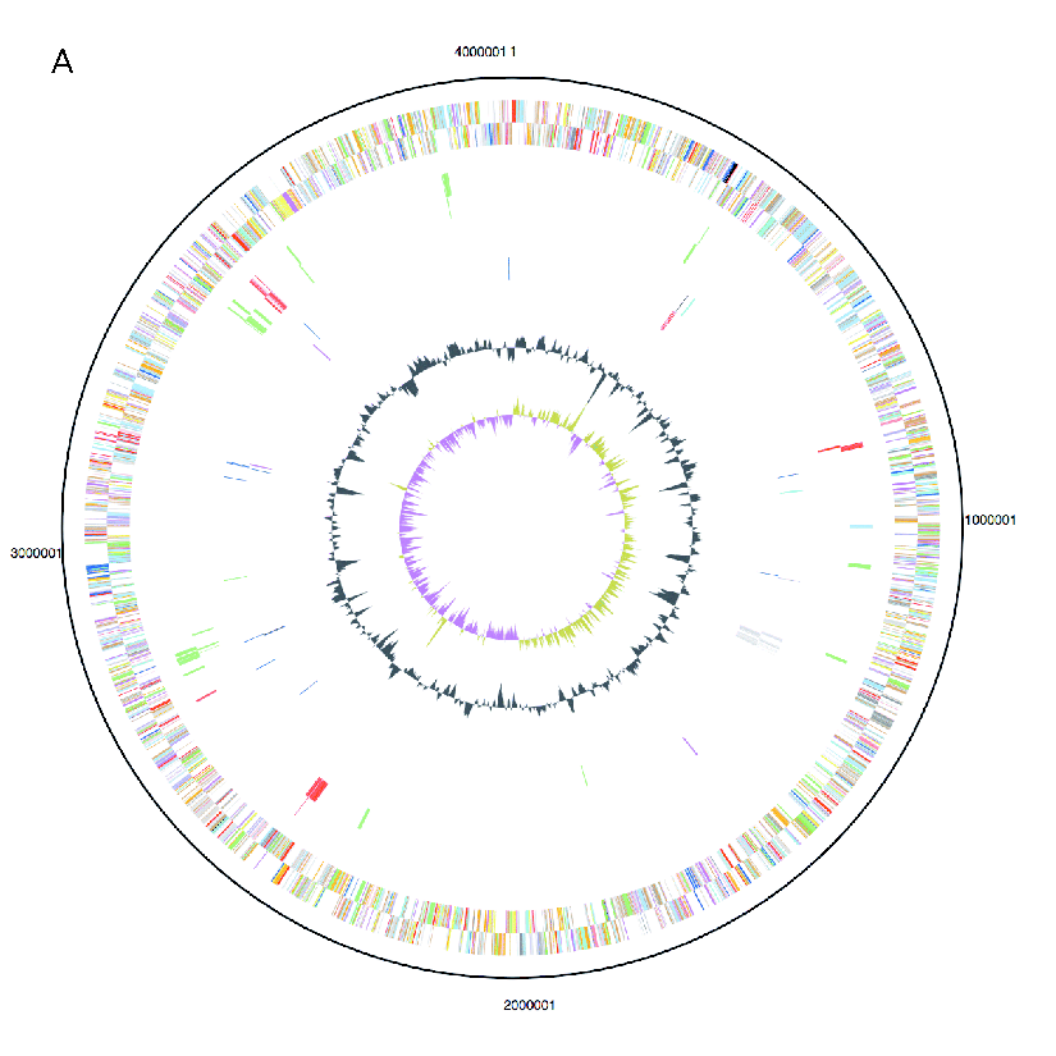
Figure 1. Circular maps of the chromosome (**a**) and megaplasmid pPM1 (**b**) of *M. petroleiphilum* PM1. The outer rings 1 and 2 show all CDSs colored by functional category (dark gray = hypothetical proteins; light gray = conserved hypothetical and unknown function; brown = general function prediction; red = replication and repair; green = energy metabolism; blue = carbon and carbohydrate metabolism; cyan = lipid metabolism; magenta = transcription; yellow = translation; orange = amino acid metabolism; pink = metabolism of cofactors and vitamins; light red = purine and pyrimidine metabolism; lavender = signal transduction; sky blue = cellular processes; pale green = structural RNAs). Genes coding for major metabolic features (green, = methylotrophy; red = aromatic hydrocarbon degradation; light blue = alkane degradation) are shown on rings 3 and 4. Large repeat regions are indicated on rings 5 and 6. The larger repeated regions colored light gray (29 kb repeat) and dark gray (40 kb repeat) are described in the text, while the other colors represent repeated IS elements, ISmp1-8. Ring 7 shows the deviation from the average G+C and the innermost ring 8 shows the GC skew (G-C)/(G+C). The plasmid and chromosome are not drawn to scale.

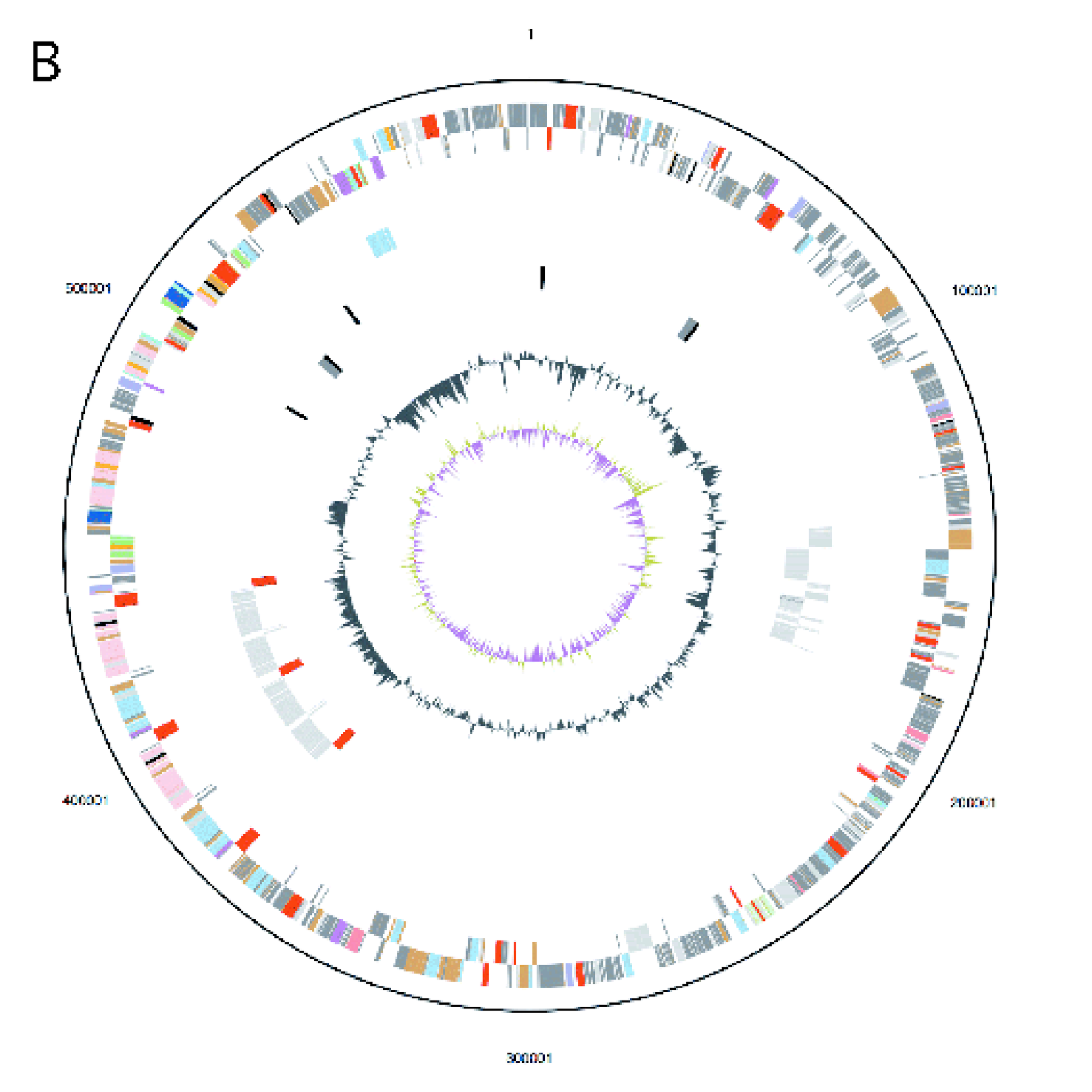
Figure 2. Schematic diagram of a *M. petroleiphilum* PM1 cell showing structural features and cellular processes including predicted methylotrophy and fuel hydrocarbon degradation pathways. Two adjacent arrows imply multiple steps. Abbreviations: MTBE, methyl *tert*-butyl ether; HMTBE, hydroxymethyl *tert*-butyl ether; TBF, *tert*-butyl formate; TBA, *tert*-butyl alcohol; HIBA, hydroxyisobutyric acid. Systems involved in metal resistance and transport, secretion, motility, chemotaxis and electron transport are shown within or associated with the cell membrane. Cofactors are labeled green and electron donors are labeled red in the cytosol. Figure is not drawn to scale.

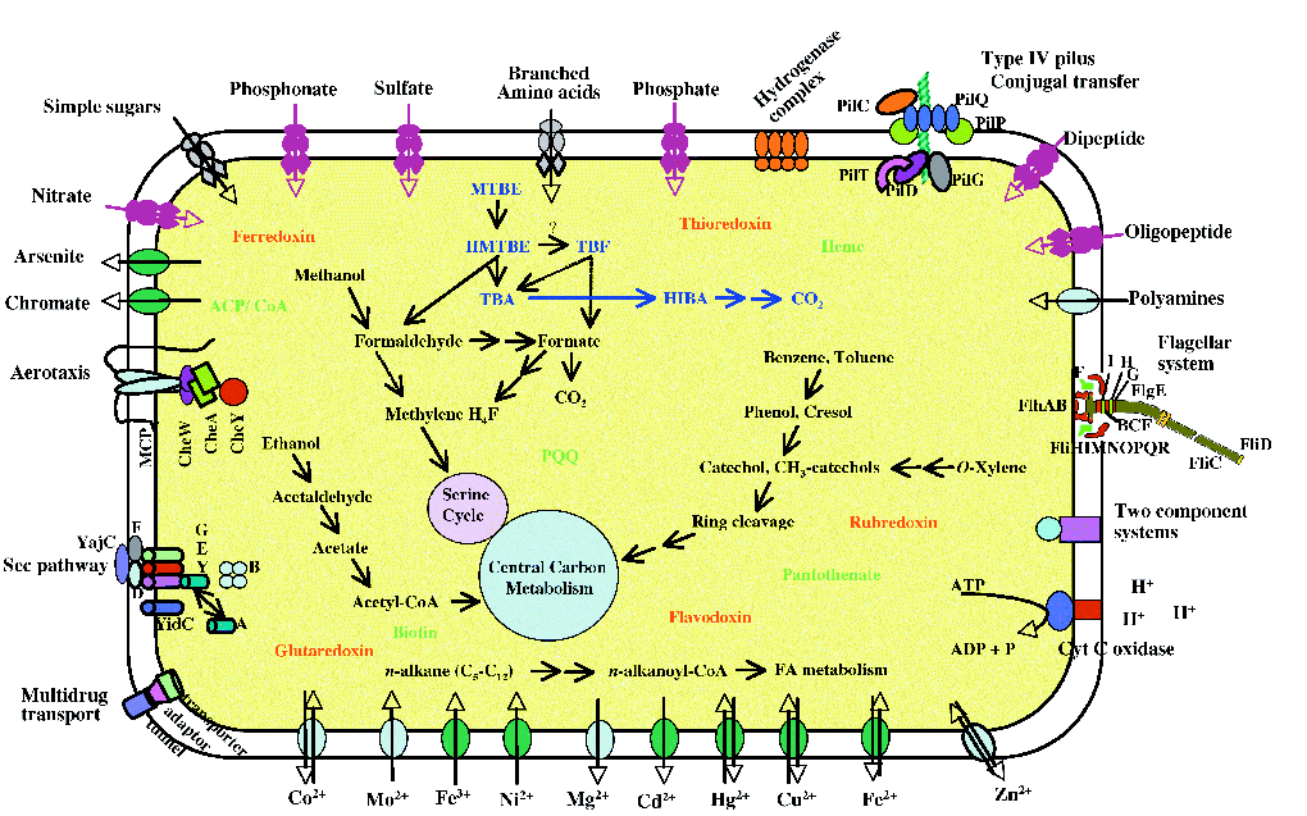
Figure 3. The phylogenetic tree of MpeA3393, which putatively codes for the large subunit of methanol dehydrogenase. The protein product from MpeA3393 is most similar to glucose dehydrogenase [GluDH] from *B. fungorum* (81% similar; Acc. No.

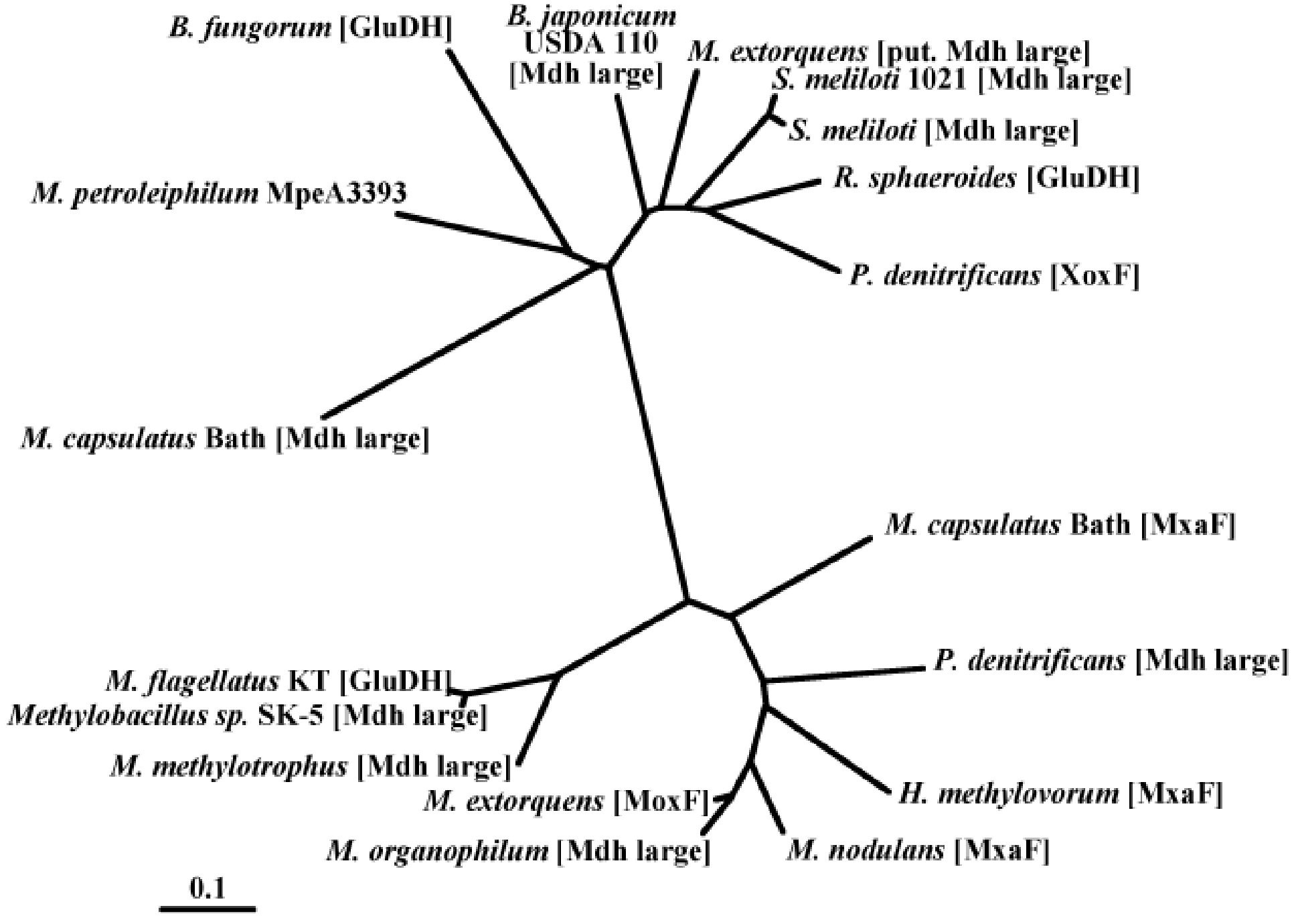
ZP_00283396) and methanol dehydrogenase large subunit [Mdh_large] from *M. capsulatus* Bath (75% similar; Acc. No. AAU90462). Gene names are in square brackets.

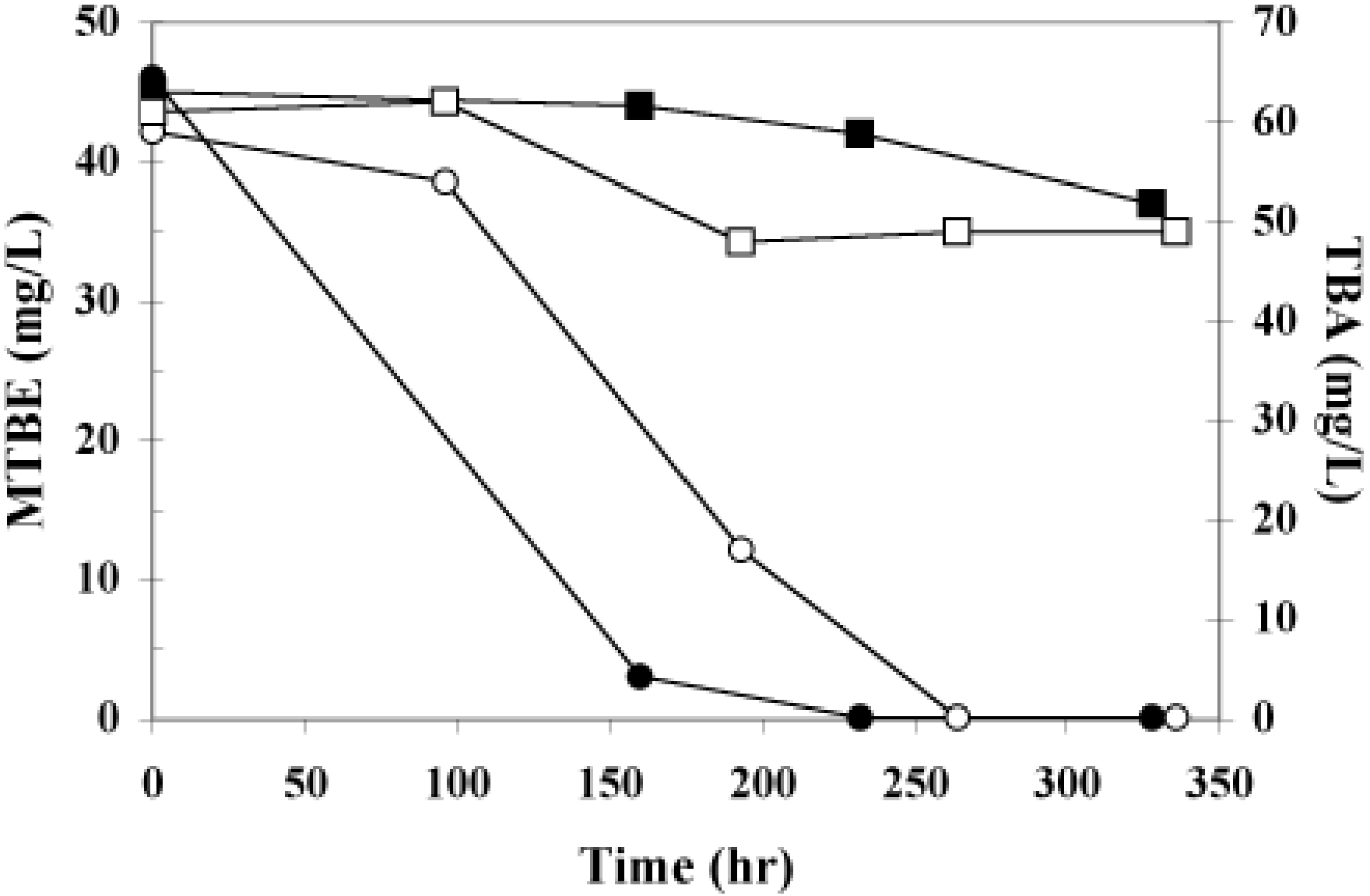
Figure 4. MTBE and TBA degradation by *M. petroleiphilum* strain PM1 by the parent strain MP0005 that carries the <SmO> marker on the megaplasmid (MTBE - ●, TBA - ○) and the megaplasmid-free mutant MP0007 (MTBE - ■, TBA - □).


A
4000001 1
1000001
2000001
3000001


B
1
100001
200001
300001
400001
500001


Simple sugars
Phosphonate
Sulfate
Branched Amino acids
Phosphate
Hydrogenase complex
Type IV pilus
Conjugal transfer
PilC
PilQ
PilP
PilT
PilD
PilG
Dipeptide
Nitrate
Arsenite
Chromate
Aerotaxis
MCP
CheW
CheA
CheY
Ferredoxin
MTBE
HMTBE
TBF
TBA
HIBA
CO_2
Thioredoxin
Heme
Oligopeptide
Polyamines
Methanol
ACP/ CoA
Formaldehyde
Formate
CO_2
Methylene H_4F
Benzene, Toluene
Phenol, Cresol
Catechol, CH_3-catechols
O-Xylene
Flagellar system
FlhAB
FliF
FliI
FliH
FliG
FlgE
BCF
FliHIMNOPQR
FliD
FliC
Ethanol
Acetaldehyde
Acetate
Acetyl-CoA
PQQ
Serine Cycle
Central Carbon Metabolism
Ring cleavage
Rubredoxin
Pantothenate
Flavodoxin
Biotin
Glutaredoxin
Two component systems
YajC
Sec pathway
F
G
E
Y
B
D
A
MdC
ATP
ADP + P
H^+
Cyt C oxidase
n-alkane (C_8-C_{12})
n-alkanoyl-CoA
FA metabolism
Multidrug transport
transporter
adaptor
tunnel
Co^{2+}
Mo^{2+}
Fe^{3+}
Ni^{2+}
Mg^{2+}
Cd^{2+}
Hg^{2+}
Cu^{2+}
Fe^{2+}
Zn^{2+}


B. fungorum [GluDH]
B. japonicum USDA 110 [Mdh large]
M. extorquens [put. Mdh large]
S. meliloti 1021 [Mdh large]
S. meliloti [Mdh large]
R. sphaeroides [GluDH]
M. petroleiphilum MpeA3393
P. denitrificans [XoxF]
M. capsulatus Bath [Mdh large]
M. capsulatus Bath [MxaF]
P. denitrificans [Mdh large]
M. flagellatus KT [GluDH]
Methylobacillus sp. SK-5 [Mdh large]
M. methylotrophus [Mdh large]
M. extorquens [MoxF]
H. methylovorum [MxaF]
M. organophilum [Mdh large]
M. nodulans [MxaF]
0.1


MTBE (mg/L)
TBA (mg/L)
Time (hr)
0
10
20
30
40
50
60
70
0
50
100
150
200
250
300
350