Identification of a Cluster of Benzene Activation Enzymes in a Strictly Anoxic Methanogenic Consortium

Courtney R. A. Toth[1], Olivia Molenda[1], Camilla Nesbø[1,2], Fei Luo[1‡], Cheryl E. Devine[1], Xu Chen[1], Kan Wu[1], Johnny Xiao[1], Shen Guo[1], Nancy Bawa[1], Robert Flick[1], and Elizabeth A. Edwards[1*]

[1]Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario, M5S 3E5, Canada.

[2]Department of Biological Sciences, University of Alberta, Edmonton, Alberta, T6G 2E9, Canada

[‡]Current Address: Liven Proteins Corporation, 200 College Street, Toronto, Ontario, M5S 3E5, Canada

[*]For correspondence, e-mail: elizabeth.edwards@utoronto.ca; Tel. (+1) 416 946 3506; Fax (+1) 416 978 8605.

Running Title: Anaerobic benzene activation enzymes

Keywords: Benzene, *Desulfobacterota* ORM2, "*Candidatus* Benzenivorax", Proteomics, Bioremediation, Biodegradation, Anaerobic Hydrocarbon Degradation, Methanogenesis

Funding and Competing Interests Statement: This study was funded by Genomic Application Partnership Program (GAPP) grants awarded to E.A.E. (Project IDs OGI-102 and OGI-173), which were supported by Genome Canada, Ontario Genomics, the Government of Ontario, Mitacs Canada, SiREM, Alberta Innovates, Federated Co-operatives Limited, and Imperial Oil. The authors declare no competing financial interest.

ABSTRACT

The Oil Refinery (OR) consortium is a model methanogenic enrichment culture for studying anaerobic benzene degradation. Over 80% of the culture's bacterial community is comprised of two closely related strains of benzene-fermenting *Desulfobacterota* (designated ORM2a and ORM2b) whose mechanism of benzene degradation is unknown. Two new metagenomes, including a fully closed metagenome-assembled genome (MAG) for ORM2a, enabled a thorough investigation of this culture's proteome. Among the proteins identified were Bam-like subunits of an ATP-independent benzoyl-CoA degradation pathway and associated downstream beta-oxidation proteins producing acetyl-CoA. The most abundant proteins identified mapped to two ORM2 gene clusters of unknown function. Syntenic gene clusters were identified in one other known benzene degrader, *Pelotomaculum* candidate BPL, as well as a handful of contigs assembled from hydrothermal vent metagenomes. Extensive searches against reference sequence and structural databases indicate that the first ("Magic") gene cluster likely catalyzes the chemically difficult benzene activation step. The second ("Nanopod") gene cluster is predicted to code for an efflux system that pumps excess benzene out of cells, mitigating some of its toxigenic effects. Phylogenomic analyses place ORM2a and ORM2b within a novel genus of benzene-degrading specialists which we propose naming "*Candidatus* Benzenivorax". We hope to engage the research community to help in confirming the roles of the proteins in the "Magic" and "Nanopod" gene clusters, and to search through their own cultures for these features.

INTRODUCTION

It has been over forty years since anaerobic benzene degradation was first discovered by Ward et al. [1], reversing a long-held tenet that aromatic hydrocarbons were recalcitrant under anaerobic conditions. Notorious for its carcinogenicity, stringent environmental regulations, and persistence in anoxic environments, benzene has been the subject of numerous studies seeking to characterize its metabolism in the absence of oxygen [2, 3]. Laboratory enrichment culture studies have confirmed unequivocally that anaerobic benzene degradation can be coupled to iron (III) reduction, nitrate reduction, sulfate reduction, and methanogenesis, catalyzed by a handful of specialized microbial clades [2, 4]. The complete biodegradation of benzene was recently shown to be independent of oxygen in a subculture of the OR consortium [5], which is also likely the case under other anaerobic electron-accepting conditions. But despite great efforts, the biochemical mechanism(s) by which unsubstituted aromatic rings are activated in the absence of oxygen have yet to be proven experimentally.

Initial studies using ^{13}C- and ^{14}C-benzene and occasionally $H_2{}^{18}O$ identified three possible activation reactions based on the detection of labelled metabolites: (i) anaerobic hydroxylation of benzene yielding phenol [6-13], (ii) carboxylation of benzene yielding benzoate [7, 9-11, 14, 15], and (iii) Friedel-Crafts-type methylation of benzene yielding toluene [11]. Neither phenol or benzoate are diagnostic of a single activation mechanism [16], and evidence of benzene methylation has since been sparse in literature [17, 18]. In 2008, hydrogen and carbon isotope fractionation ratios ($\Delta\delta^{2}H/\Delta\delta^{13}C$, a value defined as Λ) from active benzene-degrading enrichment cultures were used to look for evidence of distinct anaerobic benzene degradation pathways [19-21]. The extent of isotope fractionation in nitrate-reducing cultures (Λ = 8 to 19, n = 5) was much

66 lower than for sulfate-reducing and methanogenic cultures surveyed including OR (Λ = 28 to 31,
67 n = 7), suggesting that at least two biochemical benzene activation mechanisms exist [19].

68 By 2010, breakthrough developments in DNA sequencing and analysis techniques enabled
69 (meta)genomic characterization of anaerobic benzene degraders for the first time [22-24]. These
70 data were combined with proteomics or differential expression assays to search for elusive benzene
71 activation genes. Abu Laban et al. [25] sequenced the metagenome of the iron-reducing enrichment
72 culture BF and identified a putative anaerobic benzene carboxylation gene (*abcA*) on a *Clostridia*
73 affiliated-contig. Expression of *abcA* and neighboring genes (*abcD*, benzoyl-CoA ligase [*bzlA*]
74 and a *ubiX*-like gene) was induced with benzene but not with phenol or benzoate, hinting at a
75 possible role in benzene activation. Homologous gene clusters have since been identified in the
76 transcriptomes of several nitrate-reducing cultures [26-28] – including several previously analyzed
77 by isotopic fractionation [19, 20] – and an benzene-degrading archaeon [15]. Most nitrate and iron-
78 reducing benzene degraders identified thus far have been affiliated with the order *Thermincolales*
79 [26-30]; a putative benzene-carboxylating, sulfate-reducing *Thermincolales* was also recently
80 identified in a jet fuel-contaminated aquifer via metagenomics [31]. Researchers are now
81 attempting to verify benzene carboxylation using structural and biochemical assays, among other
82 tools [30, 32, 33].

83 Evidence for benzene hydroxylation has only been shown for one microorganism,
84 *Geobacter metallireducens* GS-15, via gene deletion. In 2013-2014, Zhang et al. [6, 24] identified
85 a handful of genes that were upregulated during active phenol or benzene metabolism, and the
86 deletion of key genes (Gmet0231, Gmet0232, or *ppsA*) inhibited the degradation of one or both
87 substrates. The gene *ppsA* encodes for the alpha subunit of phenylphosphate synthase and is
88 required for catalyzing the first step in anaerobic phenol metabolism [34, 35]. Functional

89 annotations suggest that the genes Gmet0231 and Gmet0232 encode for putative oxidoreductases,
90 but their exact functions are unknown. Homologs of Gmet0231 and Gmet0232 have been found
91 in other microorganisms not known/incapable of degrading benzene, but the adjacent localization
92 of these genes does appear to be unique to *G. metallireducens* GS-15 [24]. Further work on this
93 strain degrading benzene has not been reported. In 2024, Bullows et al. [18] identified a fumarate
94 addition gene cluster (Geob_2437-Geob_2440, encoding for 3-hydrobenzylsuccinate synthase)
95 and a putative methyltransferase (Geob_0241) in the genome of *Geotalea daltonii* strain FRC-32,
96 a close relative of *Geobacter*. All five genes were upregulated during active benzene metabolism
97 but not with toluene. Given that fumarate addition enzymes are only known to catalyze the
98 activation of alkanes and alkyl-substituted aromatic rings [17], the authors posited that benzene
99 might first be methylated by the gene product of Geob_0241 [18]. Toluene formation has not yet
100 been detected during active benzene degradation by this bacterium [18].

101 Curiously, homologs of known and predicted anaerobic aromatic-activating genes (e.g.,
102 *abcA*, *ppsA*, Gmet0231, Gmet0232, and fumarate addition genes) have never been detected in
103 methanogenic benzene-degrading enrichment cultures [4, 22] or in the well-characterized sulfate-
104 reducing enrichment culture BPL [36]. Genes associated with aerobic metabolism
105 (monooxygenases) and anaerobic growth on benzoate (benzoyl-CoA ligase) have also never been
106 found in these cultures [36]. Thus far, metagenomic and proteomic surveys of representative
107 cultures (OR and BPL) have only identified the presence and expression of putative ATP-
108 independent benzoyl-CoA biodegradation pathways during active benzene degradation [22, 36].
109 Benzoyl-CoA is a central metabolic intermediate of all proposed anaerobic benzene degradation
110 pathways [37, 38]. Given this information and previous isotope fractionation data [19, 20]
111 indicating that the mechanism of benzene activation catalyzed by the OR culture is distinct from

112 carboxylation, we surmise that a novel hydrocarbon activation mechanism or complex exists that
113 can catalyze anaerobic transformation of benzene to benzoyl-CoA.

114 In this study, we integrate the information from three rounds of proteomics experiments on
115 the methanogenic benzene-degrading OR consortium. The culture, which has been studied for over
116 25 years [4, 5, 11, 19, 20, 22, 39-44], harbors two closely related strains of benzene-degrading
117 *Desulfobacterota* (ORM2a and ORM2b), methanogenic archaea (*Methanothrix* and
118 *Methanoregula*), and >100 other typically low abundance microorganisms with various predicted
119 functions [41-43]. Proteomics was first conducted on the OR consortium in the early 2010's [22]
120 but results were difficult to interpret due to the lack of a sound metagenome and low protein counts
121 [22]. These experiments also depleted most of our culture, delaying further experimentation until
122 sufficient volumes were replenished (doubling time of ~30 days). In the years since, improvements
123 in culture maintenance [5, 41, 44] and new sequencing efforts have enabled the reconstruction of
124 two higher quality metagenomes and 74 metagenome-assembled genomes (MAGs) [42]. The
125 genome of ORM2a was successfully closed (GenBank assembly no. GCA_026625185.1, genome
126 accession no. CP113000.1), making it a potentially valuable blueprint for finally deciphering
127 anaerobic benzene degradation.

128

129 **MATERIALS AND METHODS**

130 **Culture maintenance**

131 The OR consortium was enriched from samples from an oil refinery in Oklahoma in 1995 [45] and
132 has been maintained in a near-identical manner ever since. Briefly, cultures are grown in a defined
133 pre-reduced anaerobic mineral medium [5, 40] and are fed benzene (~5-40 mg/L) once every 4-6

134 weeks. The OR consortium has been subcultured many times, resulting in the emergence of five
135 distinct culture lineages [41, 43]. The "OR-b" lineage, illustrated in Figure S1, is the most well
136 characterized [4, 5, 19, 22, 41-44] and was surveyed in this study. A large scale (>100 liters)
137 bioaugmentation culture known as DGG-B [4, 5] was derived from the OR-b lineage. Though the
138 OR consortium has been amended with other substrates (including benzoate and toluene), only
139 benzene has been shown to support the growth of ORM2a and ORM2b [22, 39].

140 **Sequencing, assembly, and annotation of OR metagenomes and MAGs**

141 Between 2010 – 2017, the metagenomes of three OR-b subcultures designated OR-b1A, DGG1A,
142 and DGG-0 were sequenced (Figure S1). Details into the DNA extraction methods, sequencing
143 platforms, and bioinformatics tools used for metagenomic assembly and binning are published
144 elsewhere [22, 42], thus only relevant information is included here. The OR-b1A (2010)
145 metagenome was sequenced using Illumina paired-end technology (2×100 base pairs, bp) and was
146 assembled in ABySS v.1.3.2 [46] using varying kmer lengths (Chapter 5 of Devine [22]). The
147 assembly process was challenging and only 23 resulting contigs were >10 kbp. Toth et al. [42]
148 assembled the metagenomes of DGG1A (2016) and DGG-0 (2017) using short-read (Illumina
149 paired-end) and long-read (PacBio) shotgun sequences, resulting in longer, high quality contigs.
150 Following binning, three draft MAGs including ORM2a were refined into closed (completed)
151 genomes [42]. The correct assembly of each closed genome was verified by read mapping. The
152 average sequence depth for the ORM2a genome was 1,371× [42, 47].

153 Toth et al. [42] assigned taxonomy to all draft and complete MAGs using release 214 of
154 the Genome Taxonomy Database (GTDB) and GTDB tool kit (v.2.3.2) using the classify_wf
155 command [42, 48, 49]. In this study, the ORM2a genome and a medium-quality MAG putatively
156 belonging to ORM2b (MAG4_ORM2b, 3,214,903 bp across 378 contigs) were reclassified using

157 release 220 of GTDB and GTDB-tk v.2.4.0 (discussed later). All metagenomic assemblies
158 (including OR-b1A) and complete MAGs were submitted to NCBI and/or the Joint Genome
159 Institute (JGI) for automated gene calling and functional annotation. Seventy-one draft MAGs
160 including MAG4_ORM2b were deposited in figshare [47] and for this study were gene called
161 using RASTtk [50] in March, 2022. Data availability information including accession numbers
162 and functional annotation pipelines used is provided in Table S1.

163 **Analysis of the ORM2a genome**

164 Several automatic pipelines were used to visualize the organization of the ORM2a genome. Ori-
165 Finder 2022 [51] was used to identify the genomes' origin of replication (*oriC*). IslandViewer 4
166 [52] and PHASTER [53] were used to identify genomic islands and prophage sequences,
167 respectively. Functional annotations and hidden Markov model assignments (from NCBI and JGI)
168 were used to identify transposase genes. Putative insertion sequences were identified using
169 ISfinder [54]. The mummer2circos package in GitHub
170 (https://github.com/metagenlab/mummer2circos) [55] was used to generate a circular plot of the
171 ORM2a genome and to compare its homology with MAG4_ORM2b (Figure 1). Two-way average
172 nucleotide identity (ANI) analysis of both genomes was performed using an online calculator
173 (http://enve-omics.ce.gatech.edu/ani/). To assess the metabolic potential of the ORM2a genome,
174 coding sequences predicted by the NCBI Prokaryotic Genome Annotation Pipeline were uploaded
175 to BlastKOALA for automatic KEGG Orthology (KO) assignment [56], then organized into
176 metabolic pathways using the Reconstruct tool within KEGG Mapper [57, 58].

177 **Protein extraction and LC-MS/MS analysis**

178 Between 2010-2018, proteomics experiments were performed on three OR-b lineage subcultures
179 (OR-b1C, OR-b, and OR-b1A; see Figure S1). Culture information, including culture volumes

180 extracted and benzene degradation rates at the time of sampling, are provided in Table 1. To avoid
181 oxygen contamination, most protein extraction steps were performed in an anaerobic chamber
182 (Coy Laboratory Products Inc) supplied with a gas mix of 10% H_2, 10% CO_2 and 80% N_2. All
183 reagents and containers used were incubated in the anaerobic chamber for >24 hours prior to use,
184 and all steps were performed at 4 °C unless otherwise specified.

185 Experiment #1 (OR-b1C, 2010) and Experiment #2 (OR-b, 2011) were conducted by
186 Devine [22] using LC-MS/MS sequencing of porcine trypsin-digested crude extracts. Detailed
187 sample preparation methods can be found in Text S1 and in Section 6.1.3 of Devine's PhD thesis
188 [22]. In Experiment #3 (OR-b1A, 2018), cell pellets from two 50 mL volumes of culture (10,000
189 x g, 15 min, 23 °C) were suspended in a sodium dodecyl sulfate (SDS)-containing lysis buffer (50
190 mM Tris-HCl [pH 7.6], 2% w/v SDS, and 50 mM DDT) and sonicated at 23-30W for 10 min in
191 pulse mode (1 s ON/1 s OFF for 1 min, 1 min OFF, repeat). Proteins from one crude lysate
192 (Experiment #3a) were purified using a Microcon-30kDa centrifugal filter unit (Millipore) as
193 outlined by Wiśniewski et al. [59], then subjected to an 18-24 hr in-solution digestion with porcine
194 trypsin [22]. The second crude lysate (Experiment #3b) was mixed with 2× SDS gel-loading buffer
195 (50 mM Tris-HCl [pH 6.8], 10% SDS [sodium dodecyl-sulfate], 40% glycerol, 3 mM bromophenol
196 blue, and 500 mM DDT]) and separated on a 15% SDS-PAGE gel run for 45 min at 200 V.
197 Visualization of the resulting gel revealed a long protein smear with no distinct bands (Figure S2),
198 which we sliced into five equal sections then destained and digested with porcine trypsin as
199 described in Shevchenko et al. [60].

200 LC-MS/MS analysis of Experiments #1 and #2 peptides was conducted by Devine [22] at
201 the SickKids SPARC BioCentre using an LTQ Orbitrap mass spectrometer (Thermo Fisher
202 Scientific). Run parameters are summarized in Tang et al. [61]. LC-MS/MS analysis of Experiment

203 #3 peptides was conducted at the University of Toronto BioZone Mass Spectrometry Facility.
204 Briefly, peptides were separated over a 10.5 cm × 7.5 µm diameter C_{18} column over 60 min using
205 a gradient of 0% - 40% acetonitrile in 0.1% formic acid. Peptide data were acquired in the data-
206 dependent mode using a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific) with
207 a 2-Da isolation width, a 35% normalized collision energy, a 2+ charge state, a 0.250 activation
208 Q, and a 30-ms activation time.

209 **Protein identification**

210 Raw mass spectrometry data were converted into mzXML using MSconvert [62] then uploaded
211 into Version 2020.11.12.1 of X! Tandem (The GPM, thegpm.org) for peptide and protein
212 discovery. Spectra were screened against a large, redundant library of protein sequences
213 (1,595,050 entries) with a mass tolerance of <0.4 Da for the fragment ions and <20 ppm for the
214 precursor ion. The library provided in File S1 and overviewed in Table S1, contained 1,106,523
215 translated gene sequences called by IMG (three metagenomic assemblies and 3 closed MAGs),
216 NCBI (3 closed MAGs), and RAST-tk (67 draft MAGs). The database also included 488,411
217 reverse-translated (decoy) protein sequences from the DGG0 metagenome to calculate a false
218 discovery rate, and 116 entries from common human/laboratory contaminants. We are in the
219 process of depositing the library to figshare for publication. Peptide and protein identifications
220 were accepted if they could be established at greater than 95.0% and 99.0% probability,
221 respectively [63, 64]. Proteins sharing peptide evidence were grouped into clusters.

222 **Functional and structural prediction of ORM2a proteins detected in high abundances**

223 The most abundant and recurring protein clusters identified mapped to two ORM2a gene clusters
224 of mostly unknown functions (see Table 2 and Figure 1). To assess whether products of these
225 clusters were involved in anaerobic benzene degradation, we first performed protein BLAST

226 searches against all other OR metagenomic assemblies and MAGs, as well as reference
227 (meta)genomic sequences in NCBI (nr), JGI (IMG/M), and UniProt (UniProt KB reference
228 proteomes + Swiss-Prot). Next, the sequence-based tool MotifFinder
229 (https://www.genome.jp/tools/motif/) was used to identify conserved domains in each gene.
230 Protein structures were predicted using the AlphaFold2.ipynb tool within Colab [65, 66], then
231 compared to homologous structures and protein domains within the Dali server's Protein Data
232 Bank (PDB) [67] and DeepFRI [68]. All AlphaFold 2 protein structural models will be deposited
233 to figshare shortly (File S2). Maximum likelihood trees were constructed in RAxML v.8.2.1.1
234 within Geneious v.8.1.9 [69, 70] to corroborate predicted protein functions where possible.
235 Translated gene sequences were also sent to the laboratory of Prof. Christine Orengo (University
236 College London) for independent analysis using CATH v.4.4 [71] and PROST [72].

237 **Phylogenomic analyses**

238 The taxonomic classification of ORM2a and ORMb has never been fully resolved. In fact, they
239 have been reassigned at least 3 times to different classes of Desulfobacterota since 2021
240 (Deltaproteobacteria, UBA8473, and most recently to WTBG01) [4, 42]. To resolve these
241 incongruencies, phylogenomic analysis of the closed ORM2a genome and MAG4_ORM2b was
242 performed in v.1.8.8 of GToTree [73]. Briefly, we retrieved all representative genomes from
243 UBA8473 and WTBG01 from GTDB v.220, as well as closely related genomes from selected
244 Deltaproteobacteria, then extracted target genes using the Proteobacteria.hmm single copy gene
245 set within GToTree. *In silico* translated protein sequences were aligned using MUSCLE [74] and
246 concatenated. A maximum likelihood tree was constructed from the concatenated alignment using
247 IQ-TREE v.2.3.6 [75] with the LG+F+I+R5 substitution model, inferred as best model by IQ-
248 TREE, and 100 bootstraps. For comparative purposes, a second maximum likelihood tree was

249 constructed using 16S rRNA gene sequences extracted from each representative GTDB genome
250 (if available) and from clone sequences of known/predicted anaerobic benzene degraders.
251 *Desulfocapsa sulflexigens* DSM 10523 (GCA_000341395.1) was included as an outgroup.

252 RESULTS AND DISCUSSION

253 General features of the ORM2a genome

254 The ORM2a genome consists of a single chromosome 3,300,546 bp in length with low GC skew
255 and no definitive terminus region (Figure 1, Figure S3). General genome statistics are summarized
256 in Table S2. Automated pipelines identified twelve possible genomic islands, of which two are
257 likely prophage sequences and 9 contain putative transposase genes (Figure 1 and Table S3). An
258 additional 47 transposase genes and 63 insertion sequences-like structures were scattered across
259 the genome (Figure 1 and Tables S4-S5), hinting that the ORM2a genome has been shaped by
260 numerous rearrangement events. A similarly high transposase content was found in the closed
261 genome of another OR culture bacterium, "*Candidatus* Nealsonbacteria" DGGOD1a, putatively
262 involved in necromass recycling [43]. No probable CRISPR elements, virulence factors, or
263 antimicrobial resistance genes were identified.

264 Approximately half (46%) of the 3,135 genes identified in the ORM2a genome were
265 connected to KEGG Orthology modules and could be organized into metabolic pathways. Key
266 pathways are overviewed here and in Figure 2; all other (in)complete pathways identified are
267 summarized in Table S6. The ORM2a genome exhibits limited metabolic potential to degrade
268 carbohydrates, lipids, amino acids, and nucleic acids. Pathways for ATP generation are also
269 limited, consisting of an F-type ATPase, a Nuo-type NADH dehydrogenase [complex I] with one
270 missing subunit (*nuoG*), and a dissimilatory sulfate reduction pathway. Recent experimental
271 evidence suggested that ORM2a could couple benzene oxidation to sulfate reduction [4], but this

272 is the first genomic evidence supporting this observation. Curiously, products of the ORM2a
273 sulfate reduction pathway were detected in the OR proteomes despite not adding any sulfate to the
274 culture medium (Figure 2). BLAST searches and manual inspections confirmed that the ORM2a
275 genome does not contain any homologs to known/predicted hydrocarbon activation genes or to a
276 benzoyl-CoA ligase (*bzlA*). The latter is a significant distinction between microorganisms
277 containing putative anaerobic benzene carboxylase genes (*abcAD*) and those that do not [36].
278 Nevertheless, the presence of a complete ATP-independent benzoyl-CoA degradation pathway is
279 consistent with ORM2a's involvement in benzene oxidation (Figures 1 and 2).

280 Only a few complete anabolic pathways were identified in the ORM2a genome, including
281 pathways for amino acids and vitamin biosynthesis (Figure 2 and Table S6). Functional
282 annotations of >40 putative transporter genes suggest that vitamin B_{12}, tyrosine, and other
283 undefined amino acids can be imported, perhaps from direct exchanges with other OR consortia
284 members like "*Ca.* Nealsonbacteria" [43]. Functional genes putatively associated with metabolic
285 syntrophic lifestyles, including formate dehydrogenases, heterodisulfide reductases, and Fe-S
286 oxidoreductases, were scattered throughout the ORM2a genome (Figure 2). The ORM2a genome
287 also encodes for a Che-type chemotaxis system and type IV pili that may be involved in bacterial
288 motility. Overall, these findings suggest that ORM2a is an obligate syntrophic anaerobe uniquely
289 adapted for anaerobic benzene degradation, consistent with all previous culture studies [4, 5, 22,
290 41].

291 **Proteomic analyses**

292 Three rounds of proteomics experiments were conducted on OR subcultures (Table 1).
293 Experiments #1 (2010) and #2 (2011) were conducted by Devine [22] and detailed results are
294 available in Chapter 6 and Appendix L of her PhD thesis. Briefly, Devine [22] matched the mass

295 spectral data against a library of 15,173 protein sequences translated from a preliminary
296 metagenome (OR-b1A, 2010), but few unique proteins were identified (69-179 per proteome). To
297 check for any missed proteins, we re-analyzed her mass spectra against translated sequences from
298 more recent OR metagenomic assemblies and MAGs (Table S1), but only a handful of new
299 proteins were discovered (Table 1).

300 To improve detection, protein separation by gel electrophoresis prior to LC/MS analysis
301 [76] was tested in Experiment #3 (2018). Duplicate crude extracts were prepared. One was
302 immediately digested with trypsin (Experiment #3a), akin to Experiments #1 and #2, and recovered
303 similarly few proteins (Table 1). The second crude extract (Experiment #3b) was separated on a
304 15% SDS-PAGE (Figure S2), then sliced into five sections for in-gel trypsin digestion and LC-
305 MS/MS sequencing. This time, over 800 unique proteins were identified (Table 1). A molecular
306 weight assessment of the proteins identified in each gel slice supports that some separation
307 occurred during SDS-PAGE (Figure S2).

308 In total, 858 unique predicted proteins were identified in Experiments #1-3 (Table 1).
309 Corresponding gene accession numbers, predicted functional annotations, amino acid sequences,
310 and peptide/spectra counts are provided in Table S7. Peptide and protein statistics are provided in
311 Tables S8a and S8b, respectively. Proteins that were detected in high abundance in Experiment
312 #3b (≥100 matching spectra) and/or are putatively involved in anaerobic benzene degradation are
313 highlighted in Figures 1 and 2 and Table 2. Most of the abundant proteins identified in Experiment
314 #3b were detected in 1 to 4 other OR proteomes (Table 2), indicating reproducibility to our
315 experimental results.

316 **Key microorganisms identified in the OR proteomes**

317 Over one-third of all unique proteins identified (39%) mapped unambiguously to the closed MAG
318 of ORM2a, covering ~11% of this bacterium's total putative proteome. We also identified 33
319 unique proteins (4%) that mapped to a draft genome putatively belonging to ORM2b
320 (MAG4_ORM2b, 378 contigs), and 44 proteins (5%) that mapped to both ORM2a and
321 MAG4_ORM2b (Table 2). The draft MAG4_ORM2b genome does not contain a 16S ribosomal
322 RNA gene sequence for us to compare to a reference clone sequence of ORM2b (NCBI accession
323 no. KT025834). However, the high average nucleotide identity (89.94%) between ORM2a and
324 MAG4_ORM2b supports its correct classification (see ring 5 of Figure 1 and Figure S4). No other
325 MAGs assembled from the OR culture aligned significantly to the ORM2 MAGs. Higher protein
326 counts suggest that ORM2a was the dominant strain in all proteomes surveyed (Table S7), though
327 it is not necessarily the more competitive organism. The relative abundances of ORM2a and
328 ORM2b fluctuate are known considerably over time and by subculture [41].

329 One-third of proteins identified in the OR proteomes mapped to MAGs of methanogenic
330 archaea (Table S7), specifically acetoclastic *Methanothrix* spp. (170 proteins across 8 MAGs) and
331 hydrogenotrophic *Methanoregula* spp. (19 proteins across 3 MAGs). Most high abundance
332 archaeal proteins were affiliated with methanogenesis and mapped to >1 MAG (Table 2). Fifty
333 low-abundance proteins mapped to other bacterial MAGs including "*Ca.* Nealsonbacteria"
334 DGGOD1a. All other proteins could not be assigned to a specific MAG (171 proteins, 20%) or
335 mapped to contaminant/decoy sequences (20 proteins, 2%).

336 **Benzoyl-CoA degradation to acetyl-CoA**

337 The OR proteomes were aligned against representative sequences of known/predicted hydrocarbon
338 activation enzymes as well as benzoyl-CoA ligase, but once again we were unable to identify any
339 homologs (Table S7). Next, we searched for products of downstream aromatic degradation

340 pathways and detected a full complement of putative benzoyl-CoA degradation enzymes mapping
341 to the ORM2a genome (Figures 1 and 2, Tables 2). Most *bam* genes products (WAC08761.1-
342 WAC08768.1) mapped to a single gene cluster and were annotated as subunits of a Bam-type
343 tungsten-containing benzoyl-CoA reductase (BamBCDEFDHI; whose components are described
344 in Wischgoll et al. [77] and Boll et al. [78]), an FdhD-like insertase, and cyclohexa-1,5-diene-1-
345 carbonyl-CoA hydratase (BamR). A putative 6-hydroxycyclohex-1-ene-1-carbonyl-CoA
346 dehydrogenase (BamQ, WAC08375.1) and 6-oxo-cyclohex-1-ene-carbonyl-CoA hydrolase
347 (BamA, WAC07399.1) mapped to other regions of the ORM2a genome (Figure 1). A handful of
348 other predicted Bam proteins mapped to MAG4_ORM2b and to a short (7.4 kbp) ORM2 contig
349 from the OR-b1A metagenome (Table 2) [22]. We screened the MAG4_ORM2b genome for more
350 functional potential to degrade benzoyl-CoA and found a near-identical complement of *bam* genes
351 across several short contigs (data not shown). Multiple sequence alignments to reference Bam
352 sequences (Table S9) and maximum likelihood phylogenetic tree analyses (Figure S5) support the
353 correct annotation of all predicted *bam* gene products.

354 Proteins predicted to catalyze the beta-oxidation of benzoyl-CoA degradation products (3-
355 hydroxypimelyl-CoA) were also identified in the OR proteomes (Figures 1 and 2, Table 2). All
356 products detected mapped to the ORM2a genome; one protein also matched to an OR-b1A contig
357 previously assigned to ORM2 [22]. In Experiment #3b, the number of matching spectra to
358 predicted upstream beta-oxidation genes was comparable to Bam spectral counts (Table 2). Oddly,
359 downstream proteins involved in the degradation of glutaconyl-CoA to acetyl-CoA were detected
360 in far lower abundances. The cause of this phenomenon is unknown but could indicate a fork in
361 the pathway at this juncture.

362 **Highly expressed ORM2a gene clusters of unknown function**

363 The three most abundant and recurring proteins identified in the OR proteomes (WAC08368.1,
364 WAC08369.1, and WAC08370.1) mapped consecutively to an ORM2a gene cluster of unknown
365 function (Figure 1 and Table S6). Products of several neighboring genes were also detected in
366 multiple OR proteomes (Table 2 and Figures 2-3). This gene cluster, designated herein as the
367 "Magic" gene cluster, is situated on a small genomic island and is flanked by multiple predicted
368 transposases and pseudogenes (Figure 1). Products of a near-identical gene neighborhood in
369 MAG4_ORM2b were also detected in the OR proteomes (Tables 2 and 3, Figure 3), as well as
370 homologous products (>90% amino acid sequence identity) from two OR-b1A contigs previously
371 assigned to ORM2 [22]. Annotations provided by NCBI and IMG offered little insight into the
372 possible functions of these proteins (Table 2 and Figure 3) with two notable exceptions. Both
373 "Magic" gene clusters in ORM2a and ORM2b were predicted to code for BamQ (WAC08375.1
374 and MAG4_3444, respectively) and the presence of Rho transcription termination factors
375 (WAC08378.1 and MAG4_3450) suggest that expression of the "Magic" gene clusters is tightly
376 regulated.

377 Several high abundance proteins mapped to a second genomic island within the ORM2a
378 genome with similarly poor functional annotations (Figure 1 and Table 2), designated the
379 "Nanopod" gene cluster. A syntenic gene cluster was found in MAG4_ORM2b but was not
380 detected in the OR proteomes (Figure 3 and Table 2). A protein containing a domain of unknown
381 function (DUF1329, WAC07727.1) piqued our interest. DUF1329-containing proteins been
382 identified in the nanopod structures of some hydrocarbon-degrading bacteria including *Delftia*
383 *acidovorans* Cs1-4, an aerobic phenanthrene degrader [79, 80]. Nanopods are tiny tubular
384 structures (<100 nm in diameter) that may facilitate nutrient transport or export toxic compounds
385 from cells [80, 81]. Shetty et al. [80] found that phenanthrene induced nanopod formation in *D.*

386 *acidovorans* Cs1-4 but not structurally similar hydrocarbons (naphthalene and toluene), and that
387 deletion of its DUF1329 gene (Genbank nucleotide accession no. DelCs14_1722) impaired growth
388 on phenanthrene. BLASTP alignments suggest that WAC07727.1 is distantly related to the
389 translated DelCs14_1722 gene sequence (21% amino acid identity). Nevertheless, the high
390 expression of this gene cluster and its possible relation to nanopod formation is intriguing.

391 **Comparative analyses of the "Magic" and "Nanopod" gene clusters to other genomes and**
392 **proteomes**

393 To determine whether other published genomes contained syntenic regions to the "Magic" and
394 "Nanopod" gene clusters, translated sequences were searched against all publicly available
395 databases in NCBI and IMG (as of August 2024) using BLASTP. After an exhaustive search, we
396 identified only a handful of MAGs and draft bins with homologous (≥ 50% identity) genomic
397 regions (Figure 3 and Table 3). Interestingly, most hits were to *Deltaproteobacteria*-associated
398 contigs from hydrothermal vent systems [82, 83]. Hydrocarbons are important energy sources for
399 microbial communities in these ecosystems and are produced primarily by geological processes
400 [84]. Hydrothermal vent MAG Deltaproteobacteria AUK045 is notable for having syntenic contigs
401 to both the "Magic" and "Nanopod" gene clusters (Figure 3 and Table 3). Genes encoding for
402 predicted downstream benzoyl-CoA degradation and beta-oxidation pathways were also identified
403 in the Deltaproteobacteria AUK045 MAG and draft bins (Table S9), demonstrating some
404 functional potential to degrade aromatic compounds.

405 *Pelotomaculum* candidate BPL (IMG GOLD accession number Ga0073689) is the chief
406 benzene degrader in a sulfate-reducing enrichment culture from Poland [10, 36]. Dong et al. [36]
407 published proteomic data from this culture in 2017, making both comparative genomic and
408 proteomic analyses to ORM2a possible (Figure 3 and Table 3). The *Pelotomaculum* candidate BPL

409 MAG (2,986,7007 bp across 97 contigs) does not contain any likely homologs to "Nanopod" genes
410 (Table 3). However, a highly expressed *Pelotomaculum* contig (Ga0073689_145) produced eight
411 significant alignments to the "Magic" gene cluster (64-76% amino acid identity, e-values ≤ 1.0E-
412 178). Other *Pelotomaculum* proteins that mapped to contig Ga0073689_145 included a predicted
413 benzoyl-CoA reductase subunit (BamB) and an enoyl-CoA hydratase, drawing parallels to the
414 adjacent localization of related genes in the MAGs of ORM2a, ORM2b, and *Deltaproteobacteria*
415 AUK045 (Figure 3). These striking similarities cannot be a coincidence.

416 **Predicted functions of the "Magic" benzene genes**

417 Multiple sequence and structure-based protein prediction tools were used to glean more insights
418 into the molecular and biological functions of the "Magic" and "Nanopod" gene clusters [67, 71].
419 Results are summarized in Figure 4 and Tables S10-S11, and key findings are described here. Each
420 prediction tool agreed that WAC08366.1, the product of the first of 12 consecutive genes in the
421 "Magic" gene cluster, is an AMP-forming acyl-CoA synthetase (Figure 4a). This class of enzymes
422 catalyzes the ATP-dependent formation of acyl-CoA molecules from carboxylates and coenzyme
423 A [85]. Sequence-based phylogenetic analyses placed this enzyme (and homologs from other
424 benzene-degrading MAGs surveyed) in a subfamily of acyl-CoA synthetases distantly related to
425 benzoyl-CoA ligase and other functionally characterized enzymes (Figure S6). Gene Orthology
426 (GO) term search results using DeepFRI [68] suggest that WAC08366.1 can bind to organic cyclic
427 compounds and such as benzene or perhaps its activation product. In all MAGs surveyed (Figure
428 3), genes coding for (homologs) of WAC08366.1 and WAC08367.1 were always co-localized,
429 suggesting that they may represent subunits of the same enzyme. We could not assign a definitive
430 function to WAC08367.1, but some evidence of organic cyclic compound binding as well as metal
431 ion binding was found. WAC08374.1 contains a HotDog fold domain and shares some structural

432 similarity with 4-hydroxybenzoyl-CoA thioesterase (Figure 4c) and (3*R*)-hydroxyacetyl-ACP
433 dehydratase subunit HadA. WAC08373.1 appears to be another Had-like subunit.

434 Products of two downstream "Magic" genes, WAC08371.1 and WAC08372.1, are likely
435 subunits of a class III CoA-transferase enzyme complex. This class of enzymes facilitates the
436 transfer of acyl-CoA compounds between substrates such as benzylsuccinate and succinyl-CoA
437 [87], which are intermediates of the anaerobic toluene degradation pathway (Figure 4c). It is
438 curious that "Magic" gene clusters appear to code for two predicted mechanisms of CoA addition,
439 one via a ligation reaction (WAC08367.1) and one using a CoA-transferase mechanisms
440 (WAC08371.1 and WAC08372.1). How this relates to anaerobic benzene transformation to
441 benzoyl-CoA is unknown.

442 Proteins WAC08368.1, WAC08369.1, and WAC08370.1 are especially intriguing and may
443 represent subunits of a fourth enzyme complex. WAC08368.1 shares several predicted molecular
444 binding functions with WAC08367.1, suggesting that the product of one enzyme could serve as a
445 substrate for the other. WAC08370.1 is structurally similar to 4-cresol dehydrogenase (Figure 4d),
446 hinting at another possible role in aromatic metabolism [89]. Both proteins can likely bind to FAD.
447 Multiple protein prediction tools agreed that WAC08369.1 is [4Fe-4S]-containing protein. It is the
448 only "Magic" protein predicted to be membrane-associated and may be capable of heterodisulfide
449 reduction and (de)hydrogenase activity . The N-terminus of WAC08369.1 also contains two
450 cysteine-rich regions (CxxxCxxC, see Figure S7), which are characteristic features of radical S-
451 adenosylmethionine (SAM) enzymes [90, 91]. This enzyme superfamily uses $[4Fe\text{-}4S]^{+}$ clusters
452 to reductively cleave the C5′–sulfonium bond of SAM, triggering radical-based chemistry in
453 hundreds of distinct chemical reactions including methylation and fumarate addition (Figure 4e)
454 [92, 93]. WAC08368.1, WAC08369.1, and WAC08370.1 may work together to couple benzene

455 activation, a highly endergonic reaction, using an uncharacterized flavin-based electron
456 bifurcation mechanism [88]. Ferredoxin and a disulfide-containing cofactor could potentially serve
457 as electron acceptors; hydrogen or perhaps benzene itself could serve as an electron donor source.
458 We've attempted many times to devise a working model for this reaction but must concede that
459 the underlying biochemistry is (for now) beyond our understanding.

460 Protein prediction tools support that the "Nanopod" gene cluster codes for a
461 transmembrane complex (Figure 2, Tables S10-S11). WAC07726.1 closely resembles a porin,
462 involved in passive transport, whereas WAC07731.1 (a predicted RND superfamily efflux protein)
463 appears to be involved in active transport. GO term searches in PROST [72] and DeepFRI [68]
464 indicate WAC07731.1 and possibly WAC07728.1 (an uncharacterized cytoplasmic protein) can
465 bind to organic cyclic structures. WAC07728.1 has an unusually wide array of possible metabolic
466 and biological functions (Table S11b). Several tools agreed that WAC07729.1 is a universal stress
467 factor. Membrane stress can result from partitioning of hydrophobic compounds such as benzene
468 into lipid bilayers [94]. Collectively, it is plausible that "Nanopod" gene cluster codes for a
469 benzene tolerance mechanism that exports excess benzene from ORM2 cells (Figure 2).

470 **Phylogenomic placement of the ORM2a and MAG4_ORM2b genomes**

471 Maximum likelihood phylogenetic analyses of concatenated conserved single copy protein
472 sequences were performed to resolve inconsistencies in the classification of ORM2a and ORM2b
473 [4, 41, 42]. The resulting consensus tree (Figure 5a) placed the genomes within an uncharacterized
474 clade of *Desulfobacterota* distinct from other classes previously assigned to these organisms
475 (including *Syntrophales*, and candidate GTDB classes WTBG01 and UBA8473) as well as
476 *Deltaproteobacteria* AUK045. The clade – likely representing a new class of *Desulfobacterota* –
477 contains three other representative MAGs from environmental samples, including one from yet

478 another hydrothermal vent system (GCA_021160085.1) [82]. These MAGs contain some
479 functional potential to degrade benzoyl-CoA and downstream intermediates (Table S9) but they
480 do not appear to degrade benzene (Table 3).

481 When compared to a complementary 16S rRNA-based phylogenetic tree (Figure 5b),
482 similar cladding patterns can be clearly seen. It also becomes obvious that ORM2a and ORM2b
483 belong to a monophyletic group of benzene-degrading specialists [4]. These organisms, which
484 were enriched from geographically diverse origins, share ≥99% 16S rRNA gene sequence identity
485 and have only ever been shown to degrade benzene under sulfate-reducing and/or methanogenic
486 conditions [40, 95, 96]. Accounting for new phylogenomic insights presented in this study, there
487 is more than sufficient evidence to recognize these organisms as a novel genus within the
488 *Desulfobacterota*. We propose naming this genus "*Candidatus* Benzenivorax" (ben.zeni.i.vor'ax;
489 N.L. neut. n. benzenum benzene, L. adj. vorax voracious or devouring, N.L. neut. n. Benzenivorax,
490 benzene-devouring).

491 **Summary and next steps**

492 Many attempts have been made to elucidate mechanisms of anaerobic benzene activation.
493 Improvements in sequencing accuracy, protein detection, and *in silico* protein structure
494 identification are now enabling a deeper dive into these mysterious processes. Because benzene
495 is the only substrate supporting growth of organisms like "*Ca.* Benzenivorax" strains ORM2a and
496 ORM2b, it is not possible to carry out differential growth experiments. Rather, we are constrained
497 to simply analyze proteins during active benzene metabolism. Nevertheless, as this study has
498 shown, analyses of expressed proteins over multiple years and from distinct labs have converged
499 on two remarkable and unique gene clusters, also found in recent MAGs from deep hydrothermal
500 vents. There is little doubt that the "Magic" gene cluster in ORM2, *Pelotomaculum* candidate BPL,

501 and specific hydrothermal vent organisms participate in anaerobic benzene degradation, and likely
502 in its elusive, chemically difficult activation step. The "Nanopod" gene clusters found in some
503 (ORM2a, MAG4_ORM2b, *Deltaproteobacteria* AUK045) but not all benzene degraders surveyed
504 is more likely involved in benzene tolerance. As most MAGs surveyed were incomplete,
505 homologous "Nanopod" genes could have been missed. It is telling that the only other genomes
506 we could find with matching regions to the "Magic" and "Nanopods" gene cluster were to other
507 strict anaerobes with niche specializations in benzene degradation. Regardless of their origins,
508 their phylogenies cluster very closely (Figure 5). Both the "Magic" and "Nanopod" gene clusters
509 appear to be located on genomic islands (Figure 1) suggesting they were acquired through gene
510 transfer. Now that these unique gene clusters have been identified, we hope that the research
511 community can help confirm their roles and to search through other cultures and datasets for
512 homologous features.

513

514 **ACKNOWLEDGEMENTS**

515 This study was made possible by the contributions of many dedicated students and research staff
516 over the past 15 years. Special recognition is given to Dr. Harry Beller (US Environmental
517 Protection Agency), Dr. Johann Heider (Philipps University of Marburg), Dr. Christine Orengo
518 (University College London), and Joel Roca (University College London) for their invaluable
519 guidance on characterizing proteins of unknown function.

REFERENCES

1. Ward DM, Atlas RM, Boehm PD *et al.* Microbial biodegradation and chemical evolution of oil from the Amoco Spill. *Ambio*. 1980;**9**:277-83 https://doi.org/http://www.jstor.org/stable/4312605
2. Vogt C, Kleinsteuber S, Richnow HH. Anaerobic benzene degradation by bacteria. *Microb Biotechnol*. 2011;**4**:710-24 https://doi.org/10.1111/j.1751-7915.2011.00260.x
3. Meckenstock RU, Boll M, Mouttaki H *et al.* Anaerobic degradation of benzene and polycyclic aromatic hydrocarbons. *J Mol Microbiol Biotechnol*. 2016;**26**:92-118 https://doi.org/10.1159/000441358
4. Toth CRA, Luo F, Bawa N *et al.* Anaerobic benzene biodegradation linked to the growth of highly specific bacterial clades. *Environ Sci Technol*. 2021;**55**:7970-80 https://doi.org/10.1021/acs.est.1c00508
5. Guo S, Toth CRA, Luo F *et al.* Transient oxygen exposure causes profound and lasting changes to a benzene-degrading methanogenic community. *Environ Sci Technol*. 2022;**20**:13036-45 https://doi.org/10.1101/2022.04.11.487956
6. Zhang T, Tremblay PL, Chaurasia AK *et al.* Anaerobic benzene oxidation via phenol in *Geobacter metallireducens*. *Appl Environ Microbiol*. 2013;**79**:7800-06 https://doi.org/10.1128/AEM.03134-13
7. Caldwell ME, Suflita JM. Detection of phenol and benzoate as intermediates of anaerobic benzene biodegradation under different terminal electron-accepting conditions. *Environ Sci Technol*. 2000;**34**:1216-20 https://doi.org/https://doi.org/10.1021/es990849j
8. Vogel TM, Grbić-Galić D. Incorporation of oxygen from water into toluene and benzene during anaerobic fermentative transformation. *Appl Environ Microbiol*. 1986;**52**:200-02
9. Kunapuli U, Griebler C, Beller HR *et al.* Identification of intermediates formed during anaerobic benzene degradation by an iron-reducing enrichment culture. *Environ Microbiol*. 2008;**10**:1703-12 https://doi.org/10.1111/j.1462-2920.2008.01588.x
10. Abu Laban N, Selesi D, Jobelius C *et al.* Anaerobic benzene degradation by Gram-positive sulfate-reducing bacteria. *FEMS Microbiol Ecol*. 2009;**68**:300-11 https://doi.org/10.1111/j.1574-6941.2009.00672.x
11. Ulrich AC, Beller HR, Edwards EA. Metabolites detected during biodegradation of $^{13}C_6$-benzene in nitrate-reducing and methanogenic enrichment cultures. *Environ Sci Technol*. 2005;**39**:6681-91 https://doi.org/https://doi.org/10.1021/es050294u
12. Weiner JM, Lovley DR. Rapid benzene degradation in methanogenic sediments from a petroleum-contaminated aquifer. *Appl Environ Microbiol*. 1998;**64**:1937-39 https://doi.org/https://doi.org/10.1128/aem.64.5.1937-1939.1998
13. Botton S, Parsons JR. Degradation of btx by dissimilatory iron-reducing cultures. *Biodegradation*. 2007;**18**:371-81 https://doi.org/10.1007/s10532-006-9071-9
14. Phelps CD, Zhang X, Young LY. Use of stable isotopes to identify benzoate as a metabolite of benzene degradation in a sulphidogenic consortium. *Environ Microbiol*. 2001;**3**:600-03 https://doi.org/https://doi.org/10.1046/j.1462-2920.2001.00228.x
15. Holmes DE, Risso C, Smith JA *et al.* Anaerobic oxidation of benzene by the hyperthermophilic archaeon *Ferroglobus placidus*. *Appl Environ Microbiol*. 2011;**77**:5926-33 https://doi.org/10.1128/AEM.05452-11
16. Gieg LM, Toth CRA Signature metabolite analysis to determine in situ anaerobic hydrocarbon biodegradation. *Anaerobic Utilization of Hydrocarbons, Oils, and Lipids*, 361-90

17. Boll M, Estelmann S, Heider J Anaerobic degradation of hydrocarbons: Mechanisms of hydrocarbon activation in the absence of oxygen. In: Boll M (ed.). *Anaerobic Utilization of Hydrocarbons, Handbook of Hydrocarbon and Lipid Microbiology*, Cham: Springer.
18. Bullows JE, Kanak A, Shedrick L *et al.* Anaerobic benzene oxidation in *Geotalea daltonii* involves activation by methylation and is regulated by the transition state regulator AbrB. *Environ Microbiol*. 2024;**90**:e00856-24
19. Mancini SA, Devine CE, Elsner M *et al.* Isotopic evidence suggests different initial reaction mechanisms for anaerobic benzene biodegradation. *Environ Sci Technol*. 2008;**42**:8290-96
20. Mancini SA, Ulrich AC, Lacrampe-Couloume G *et al.* Carbon and hydrogen isotopic fractionation during anaerobic biodegradation of benzene. *Appl Environ Microbiol* 2003;**69**:191-98 https://doi.org/10.1128/aem.69.1.191-198.2003
21. Fischer A, Herklotz I, Herrmann S *et al.* Combined carbon and hydrogen isotope fractionation investigations for elucidating benzene biodegradation pathways. *Environ Sci Technol*. 2008;**42**:4356-63
22. Devine CE. Identification of key organisms, genes and pathways in benzene-degrading methanogenic cultures. Chemical Engineering and Applied Chemistry, University of Toronto, 2013
23. Taubert M, Vogt C, Wubet T *et al.* Protein-SIP enables time-resolved analysis of the carbon flux in a sulfate-reducing, benzene-degrading microbial consortium. *ISME J*. 2012;**6**:2291-301 https://doi.org/10.1038/ismej.2012.68
24. Zhang T, Tremblay PL, Chaurasia AK *et al.* Identification of genes specifically required for the anaerobic metabolism of benzene in *Geobacter metallireducens*. *Front Microbiol*. 2014;**5**:245 https://doi.org/10.3389/fmicb.2014.00245
25. Abu Laban N, Selesi D, Rattei T *et al.* Identification of enzymes involved in anaerobic benzene degradation by a strictly anaerobic iron-reducing enrichment culture. *Environ Microbiol*. 2010;**12**:2783-96 https://doi.org/10.1111/j.1462-2920.2010.02248.x
26. Luo F, Gitiafroz R, Devine CE *et al.* Metatranscriptome of an anaerobic benzene-degrading, nitrate-reducing enrichment culture reveals involvement of carboxylation in benzene ring activation. *Appl Environ Microbiol*. 2014;**80**:4095-107 https://doi.org/10.1128/AEM.00717-14
27. Atashgahi S, Hornung B, van der Waals MJ *et al.* A benzene-degrading nitrate-reducing microbial consortium displays aerobic and anaerobic benzene degradation pathways. *Sci Rep*. 2018;**8**:4490 https://doi.org/10.1038/s41598-018-22617-x
28. Melkonian C, Fillinger L, Atashgahi S *et al.* High biodiversity in a benzene-degrading nitrate-reducing culture is sustained by a few primary consumers. *Commun Biol*. 2021;**4**:530 https://doi.org/10.1038/s42003-021-01948-y
29. Xiao JZ, Nesbø CL, Molenda O *et al.* Draft genome of an anaerobic nitrate-reducing, benzene-degrading member of the order *Thermincolales*. *Microbiol Resour Announc*. 2024;**13**:e00295-24
30. Xiao JZ, Nesbø CL, Molenda O *et al.* Metagenomic and genomic sequences from a nitrate-reducing benzene-degrading enrichment culture. *Microbiol Resour Announc*. 2024;**13**:e00294-24
31. Hidalgo KJ, Centurion VB, Lemos LN *et al.* Disentangling the microbial genomic traits associated with aromatic hydrocarbon degradation in a jet fuel-contaminated aquifer. *Biodegradation*. 2025;**36**:7

612 32. Roberts GW, Leys D. Structural insights into UbiD reversible decarboxylation. *Curr Opin*
613 *Struct Biol*. 2022;**75**:102432
614 33. Heker I, Haberhauer G, Meckenstock RU. Naphthalene carboxylation in the sulfate-
615 reducing enrichment culture N47 is proposed to proceed via 1,3-dipolar cycloaddition to
616 the cofactor prenylated flavin mononucleotide. *Appl Environ Microbiol*. 2023;**89**:e01927-
617 22
618 34. Schleinitz KM, Schmeling S, Jehmlich N *et al.* Phenol degradation in the strictly anaerobic
619 iron-reducing bacterium *Geobacter metallireducens* gs-15. *Appl Environ Microbiol*.
620 2009;**75**:3912-9 https://doi.org/10.1128/AEM.01525-08
621 35. Keller AH, Schleinitz KM, Starke R *et al.* Metagenome-based metabolic reconstruction
622 reveals the ecophysiological function of *Epsilonproteobacteria* in a hydrocarbon-
623 contaminated sulfidic aquifer. *Front Microbiol*. 2015;**6**:1396
624 https://doi.org/10.3389/fmicb.2015.01396
625 36. Dong X, Dröge J, von Toerne C *et al.* Reconstructing metabolic pathways of a member of
626 the genus *Pelotomaculum* suggesting its potential to oxidize benzene to carbon dioxide
627 with direct reduction of sulfate. *FEMS Microbiol Ecol*. 2017;**93**:fiw254
628 https://doi.org/10.1093/femsec/fiw254
629 37. Boll M, Estelmann S, Heider J Catabolic pathways and enzymes involved in the anaerobic
630 degradation of monocyclic aromatic compounds In: Boll M (ed.). *Anaerobic*
631 *Utilization of Hydrocarbons, Oils, and Lipids. Handbook of Hydrocarbon and Lipid*
632 *microbiology*, Cham: Springer. 85-133
633 38. Harwood CS, Burchhardt G, Herrmann H *et al.* Anaerobic metabolism of aromatic
634 compounds via the benzoyl-CoA pathway. *FEMS Microbiol Rev*. 1999;**22**:439-58
635 39. Luo F. Characterization of the microbial community composition and benzene activation
636 mechanisms in anaerobic benzene-degrading enrichment cultures. Chemical Engineering
637 and Applied Sciences, University of Toronto, 2016
638 40. Ulrich AC, Edwards EA. Physiological and molecular characterization of anaerobic
639 benzene-degrading mixed cultures. *Environ Microbiol*. 2003;**5**:92-102
640 41. Luo F, Devine CE, Edwards EA. Cultivating microbial dark matter in benzene-degrading
641 methanogenic consortia. *Environ Microbiol*. 2016;**18**:2923-36
642 https://doi.org/10.1111/1462-2920.13121
643 42. Toth CRA, Molenda O, Nesbø C *et al.* Metagenomic and genomic sequences from a
644 methanogenic benzene-degrading consortium. *Microbiol Resour Announcs*.
645 2023;**22**:e01342-22
646 43. Chen X, Molenda O, Brown CT *et al.* "*Candidatus* Nealsonbacteria" are likely biomass
647 recycling ectosymbionts of methanogenic archaea in a stable benzene-degrading
648 enrichment culture. *Appl Environ Microbiol*. 2023;**89**:e00025-23
649 44. Chen X, Toth CRA, Guo S *et al.* Visualization of syntrophic benzene-fermenting
650 *Desulfobacterota* ORM2 in a methanogenic enrichment culture using fluorescence *in situ*
651 hybridization. *BioRxiv*. 2023:2023.07.16.549207
652 45. Nales M, Butler BJ, Edwards EA. Anaerobic benzene biodegradation: A microcosm
653 survey. *Bioremediat J*. 1998;**2**:125-44 https://doi.org/10.1080/10889869891214268
654 46. Simpson JT, Wong K, Jackman SD *et al.* Abyss: A parallel assembler for short read
655 sequence data. *Genome Res*. 2009;**19**:1117-23 https://doi.org/10.1101/gr.089532.108

656 47. Toth CRA, Molenda O, Nesbo C *et al.* FASTA files and relevant statistics for draft MAGs
657 generated from a methanogenic benzene-degrading enrichment culture. *figshare*.
658 2022;**Dataset**
659 48. Chaumeil PA, Mussig AJ, Hugenholtz P *et al.* GTDB-tk: A toolkit to classify genomes
660 with the genome taxonomy database. *Bioinformatics*. 2019
661 https://doi.org/10.1093/bioinformatics/btz848
662 49. Parks DH, Chuvochina M, Rinke C *et al.* Gtdb: An ongoing census of bacterial and archaeal
663 diversity through a phylogenetically consistent, rank normalized and complete genome-
664 based taxonomy. *Nucleic Acids Res*. 2021;**50**:D785-D94
665 50. Brettin T, Davis JJ, Disz T *et al. Rasttk*: A modular and extensible implementation of the
666 RAST algorithm for building custom annotation pipelines and annotating batches of
667 genomes. *Sci Rep*. 2015;**5**:8365
668 51. Dong M-J, Luo H, Gao F. Ori-finder 2022: A comprehensive web server for prediction and
669 analysis of bacterial replication origins. *Genom Proteom Bioinform*. 2022
670 https://doi.org/https://doi.org/10.1016/j.gpb.2022.10.002
671 52. Bertelli C, Laird MR, Williams KP *et al.* IslandViewer 4: Expanded predicted of genomic
672 islands for larger-scale datasets. *Nucleic Acids Res*. 2017;**45**:W30-W35
673 53. Arndt D, Grant J, Marcu A *et al.* PHASTER: A bettter, faster version of the PHAST phage
674 search tool. *Nucleic Acids Res*. 2016;**44**:W16-W21
675 54. Siguier P, Perochon J, Lestrade L *et al.* ISfinder: The reference centre for bacterial insertion
676 sequences. *Nucleic Acids Res*. 2006;**34**:D32-6 https://doi.org/10.1093/nar/gkj014
677 55. Krzywinski M, Schein J, Birol I *et al.* Circos: An information aesthetic for comparative
678 genomics. *Genome Res*. 2009;**19**
679 56. Kanehisa M, Sato Y, Morishima K. Blastkoala and ghostkoala: KEGG tools for functional
680 characterization of genome and metagenome sequences. *J Mol Biol*. 2016;**428**:726-31
681 57. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein
682 sequences. *Protein Sci*. 2020;**29**:28-35
683 58. Kanehisa M, Sato Y, Kawashima M. KEGG mapping tools for uncovering hidden features
684 in biological data. *Protein Sci*. 2022;**31**:47-53
685 59. Wiśniewski JR, Zougman A, Nagaraj N *et al.* Universal sample preparation method for
686 proteome analysis. *Nature Methods*. 2009;**6**:359-62
687 60. Shevchenko A, Tomas H, Havlis J *et al.* In-gel digestion for mass spectrometric
688 characterization of proteins and proteomes. *Nat Protoc*. 2006;**1**:2856-60
689 61. Tang S, Chan WWM, Fletcher KE *et al.* Functional characterization of reductive
690 dehalogenases by using blue native polyacrylamide gel electrophoresis. *Appl Environ*
691 *Microbiol*. 2013;**79**:01873-12 https://doi.org/10.1128/AEM.01873-12
692 62. Kessner D, Chambers M, Burke R *et al.* Proteowizard: Open source software for rapid
693 proteomics tools development. *Bioinformatics*. 2008;**24**:2534-36
694 https://doi.org/10.1093/bioinformatics/btn323
695 63. Keller A, Nesvizhskii AI, Kolker E *et al.* Empirical statistical model to estimate the
696 accuracy of peptide identifiations made by MS/MS and database search. *Anal Chem*.
697 2002;**74**:5383-92
698 64. Nesvizhskii AI, Keller A, Kolker E *et al.* A statistical model for identifying proteins by
699 tandem mass spectrometry. *Analytical Chemistry*. 2003;**75**:4646-58
700 65. Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with
701 Alphafold. *Nature*. 2021;**596**:583-89

66. Mirdita M, Schütze K, Moriwaki Y *et al.* CoLabfold: Making protein folding accessible to all. *Nat Methods*. 2022;**19**:679-82
67. Holm L. Dali server: Structural unification of protein families. *Nucleic Acids Res*. 2022;**50**:W210-W15
68. Gligorijević V, Renfrew PD, Kosciolek T *et al.* Structure-based protein function prediction using graph convolutional networks. *Nat Commun*. 2021;**12**:3168
69. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;**30**:1312-13 https://doi.org/10.1093/bioinformatics/btu033
70. Kearse M, Moir R, Wilson A *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;**28**:1647-49 https://doi.org/10.1093/bioinformatics/bts199
71. Waman VP, Bordin N, Lau A *et al.* CATH v4.4: Major expansion of cath by experimental and predicted structural data. *Nucleic Acids Res*. 2024:gkae1087
72. Kilinc M, Jia K, Jernigan RL. Improved global protein homolog detection with major gains in function identification. *PNAS*. 2023;**120**:e2211823120
73. Lee MD. Gtotree: A user-friendly workflow for phylogenomics. *Bioinformatics*. 2019;**35**:4162-64 https://doi.org/10.1093/bioinformatics/btz188
74. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;**32**:1792-7 https://doi.org/10.1093/nar/gkh340
75. Nguyen L-T, Schmidt HA, von Haeseler A *et al.* IQ-tree: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2014;**3**:268-74
76. Dzieciatkowska M, Hill R, Hansen KC. GeLC-MS/MS analysis of complex protein mixtures. *Methods Mol Biol*. 2014;**1156**:53-66
77. Wischgoll S, Heintz D, Peters F *et al.* Gene clusters involved in anaerobic benzoate degradation of *Geobacter metallireducens*. *Mol Microbiol*. 2005;**58**:1238-52 https://doi.org/10.1111/j.1365-2958.2005.04909.x
78. Boll M, Einsle O, Ermler U *et al.* Structure and function of the unusual tungsten enzymes acetylene hydratase and class II benzoyl-coenzyme A reductase. *J Mol Microbiol Biotechnol*. 2016;**26**:119-37
79. Shetty A, Chen S, Tocheva EI *et al.* Nanopods: A new bacterial structure ans mechanism for deployment of outer membrane vesicles. *PLoS ONE*. 2011;**6**:e20725
80. Shetty A, Hickey WJ. Effects of outer membrane vesicle formation, surface-layer production and nanopod development on the metabolism of phenanthrene by *Delftia acidovorans* Cs1-4. *PLoS One*. 2014;**9**:e92143
81. Baumgarten T, Sperling S, Seifert J *et al.* Membrane vesicle formation as a multiple-stress response mechanism enhances *Pseudomonas putida* DOT-T1E cell surface hydrophobicity and biofilm formation. *Appl Environ Microbiol*. 201;**78**:6217-24
82. Speth DR, Yu FB, Connon SA *et al.* Microbial communities of Auka hydothermal sediments shed light on vent biogeography and the evolutionary history of thermophily. *ISME J*. 2022;**16**:1750-64
83. Dombrowski N, Teske AP, Baker BJ. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat Commun*. 2018;**9**:4999

84. Campbell KA. Hydrocarbon seep and hydrothermal vent paleoenvironments and paleontology: Past developments and future research directions. *Palaeogeogr Palaeoclimatol Palaeoecol*. 2006;**232**:362-407
85. Starai VJ, Escalante-Semerena JC. Acetyl-coenzyme A synthetase (AMP forming). *Cell Mol Life Sci*. 2004;**61**:2020-30
86. Heider J. A new family of CoA-transferases. *FEBS Lett*. 2001;**509**:345-49
87. Carmona M, Zamarro MT, Blazquez B *et al.* Anaerobic catabolism of aromatic compounds: A genetic and genomic view. *Microbiol Mol Biol Rev*. 2009;**73**:71-133 https://doi.org/10.1128/MMBR.00021-08
88. Booker SJ, Lloyd CT. Twenty years of radical SAM! The genesis of the superfamily. *ACS Bio & Med Chem Au*. 2022;**2**:538-655
89. Frey PA, Hegeman AD, Ruzicka FJ. The radical SAM superfamily. *Crit Rev Biochem Mol Biol*. 2008;**43**:63-88
90. Nicolet Y. Structure-function relationships of radical SAM enzymes. *Nat Catal*. 2020;**3**:337-50
91. Broderick JB, Duffus BR, Duschene KS *et al.* Radical *S*-adenosylmethionine enzymes. *Chem Rev*. 2014;**114**:4229-317
92. Buckel W, Thauer RK. Flavin-based electron bifurcation, ferredoxin, flavodoxin, and anaerobic respiration with protons (Ech) or nad^+ (Rnf) as electron acceptors: A historical review. *Front Microbiol*. 2018;**9**:401
93. Ramos JL, Duque E, Gallegos MT *et al.* Mechanisms of solvent tolerance in Gram-negative bacteria. *Ann Rev Microbiol*. 2002;**56**:473-768
94. Noguchi M, Kurisu F, Kasuga I *et al.* Time-resolved DNA stable isotope probing links *Desulfobacterales*- and *Coriobacteriaceae*-related bacteria to anaerobic degradation of benzene under methanogenic conditions. *Microbes Environ*. 2014;**29**:191-99 https://doi.org/10.1264/jsme2.ME13104
95. Sakai N, Kurisu F, Yagi O *et al.* Identification of putative benzene-degrading bacteria in methanogenic enrichment cultures. *J Biosci Bioeng*. 2009;**108**:501-07 https://doi.org/10.1016/j.jbiosc.2009.06.005
96. Thoden JB, Zhuang Z, Dunaway-Mariano D *et al.* The structure of 4-hydroxybenzoyl-CoA thioesterase from *Arthrobacter* sp. Strain SU. *J Biol Chem*. 2003;**278**:709-16
97. Heider J, Szaleniec M, Martins BM *et al.* Structure and function of benzylsuccinate synthase and related fumarate-adding glycyl radical enzymes. *J Mol Microbiol Biotechnol*. 2016;**26**:29-44 https://doi.org/10.1159/000441656

Table 1. OR proteomic experiments completed in this study, key methodology differences, and overview of results. One protein sample was analyzed in Experiment #1, whereas technical duplicate protein samples (a and b) were analyzed in Experiments #2 and #3. All samples were extracted for proteins using sonication methods.

Experiment # (Year)	OR subculture used	Benzene degradation rate (mg/L/day)	Culture volume extracted (mL)	Protein purification methodology	Protein separation methodology	Total # of spectra detected	No. of unique predicted proteins identified	No. of unique predicted proteins assigned to designated MAG (s) (% of unique proteins identified)				
								ORM2a	MAG4_ ORM2b	Both ORM2a and MAG4_ ORM2b	Archaeal MAGs	Other Bacterial MAGs
1 (2010)	OR-b1C	0.3	175	Phenol-methanol/ammonium acetate extraction	None	701	120	29 (24%)	14 (12%)	0 (0%)	37 (31%)	2 (2%)
2a (2011)	OR-1b	0.4	175	Phenol-methanol/ammonium acetate extraction	None	228	49	20 (41%)	3 (6%)	1 (2%)	13 (27%)	1 (2%)
2b (2011)	OR-1b	0.4	175	Phenol-methanol/ammonium acetate extraction	None	457	74	27 (36%)	5 (7%)	0 (0%)	21 (28%)	1 (1%)
3a (2018)	OR-b1A	0.7	50	Filter-aided sample preparation	None	47	24	12 (50%)	0 (0%)	1 (4%)	6 (23%)	2 (8%)
3b (2018)	OR-b1A	0.7	50	Filter-aided sample preparation	15% SDS-PAGE	10,938	822	350 (43%)	25 (3%)	43 (5%)	184 (22%)	48 (6%)
							Average	**39%**	**6%**	**2%**	**27%**	**4%**
							Total # Unique Proteins identified	**362**	**33**	**44**	**190**	**50**

Table 2. Summary of high abundance protein clusters (100+ spectra in Experiment #3b) and proteins of interest recovered with high certainty (>99% confidence) from OR proteomes. Protein clusters detected in ≥3 proteomes are highlighted in blue. The inset heat map denotes the sum of matching MS/MS spectra recovered from Experiment #3b, where higher abundance proteins are indicated with darker shades of red. Other protein clusters identified are provided in Table S7.

Protein Cluster #	Predicted function (NCBI Prokaryotic Genome Annotation Pipeline version 6.3 and best BLASTP hits)	# of proteomes with ≥1 matching spectra	Sum of matching spectra in Exp. #3b	Matching protein accession no.(s)		
				ORM2a (NCBI)	Other MAG(s) (File S1)	Unbinned contigs (IMG)
Benzoyl-CoA degradation to 3-hydroxypimelyl-CoA						
1	Benzoyl-CoA reductase subunit BamB	2	80	WAC08761.1	ORM2b_MAG4_01005	
2	Benzoyl-CoA reductase subunit BamC	3	15	WAC08762.1	ORM2b_MAG4_01004	
3	Benzoyl-CoA reductase subunit BamD	1	61	WAC08763.1		
4	Formate dehydrogenase accessory sulfurtransferase FdhD	1	6	WAC09094.1		
5	Benzoyl-CoA reductase subunit BamE	1	68	WAC09095.1		
6	Benzoyl-CoA reductase subunit BamF	1	18	WAC08764.1		
7	Benzoyl-CoA reductase subunit BamG	1	6	WAC08765.1		
8	Benzoyl-CoA reductase subunit BamH	1	91	WAC08766.1	ORM2b_MAG4_04225	
9	Benzoyl-CoA reductase subunit BamI	1	9	WAC08767.1		
10	Cyclohexa-1,5-diene-1-carbonyl-CoA hydratase BamR	1	49	WAC08768.1		BDMCk91_1074863
11	6-hydroxycyclohex-1-ene-1-carbonyl-CoA dehydrogenase BamQ	3	103	WAC08375.1		
12	6-oxo-cyclohex-1-ene-carbonyl-CoA hydrolase BamA	4	143	WAC07399.1		
3-hydroxypimelyl-CoA degradation to acetyl-CoA						
13	3-hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase family protein	4	148	WAC08792.1		
14	Thiolase family protein	4	69	WAC08791.1		
15	Glutaryl-CoA dehydrogenase (non-decarboxylating)	4	58	WAC06987.1		
16	Electron transfer flavoprotein subunit beta/FixA family protein	2	39	WAC06988.1		
17	Electron transfer flavoprotein subunit alpha/FixB family protein	4	63	WAC06989.1		
18	Glutaconyl-CoA decarboxylase subunit alpha	1	5	WAC08228.1		BDMCk91_1065314
19	Enoyl-CoA hydratase-related protein	1	3	WAC07318.1		
20	3-hydroxybutyryl-CoA dehydrogenase	1	9	WAC09154.1		
21	Acetyl-CoA C-acetyltransferase	1	3	WAC07319.1		
Highly expressed "Magic" gene cluster putatively involved in benzene activation						
22	AMP-binding protein	1	74	WAC08366.1	ORM2b_MAG4_03434	
23	MaoC family dehydratase	1	57	WAC08367.1		
24	FAD-binding oxidoreductase	4	172	WAC08368.1		
25	(Fe-S)-binding protein	5	219	WAC08369.1		BDMCk91_1046562
26	FAD-binding oxidoreductase	5	265	WAC08370.1		BDMCk91_1006392
27	Type III CoA transferase	2	45	WAC08371.1		
28	Type III CoA transferase	3	58	WAC08372.1	ORM2b_MAG4_03441	BDMCk91_1053181
29	Hypothetical protein	1	19	WAC08373.1	ORM2b_MAG4_03442	
30	MaoC family dehydratase	4	50	WAC08374.1		
31	Transcription termination factor Rho	1	25	WAC08378.1		
Highly expressed "Nanopod" gene cluster putatively involved in benzene tolerance						
32	Hypothetical protein	4	106	WAC07726.1		
33	DUF1329 domain-containing protein	3	158	WAC07727.1	ORM2b_MAG4_01802	BDMCk91_1026391
34	Universal stress protein	1	6	WAC07728.1		
35	Hypothetical protein	1	7	WAC07729.1		
36	YCF48-related protein	3	65	WAC07730.1		
37	MMPL family transporter	2	60	WAC07731.1		
Archaeal proteins associated with methanogenesis						
38	Acetate--CoA ligase	5	142		MAG20_01236	
39	Coenzyme-B sulfoethylthiotransferase McrB	2	149		MAG21_01157	
40	Coenzyme-B sulfoethylthiotransferase McrA	2	147			Ga0206388_11752373
41	Coenzyme-B sulfoethylthiotransferase McrB	1	154		MAG11_02361, MAG20_02648, MAG28_00841	
42	Coenzyme-B sulfoethylthiotransferase McrA	1	146			Ga0310375_10020214
43	Coenzyme-B sulfoethylthiotransferase McrB	1	144			Ga0206388_10804781
44	Acetate--CoA ligase	1	125		MAG11_00500	
Other proteins						
45	Acetate--CoA ligase Acs	4	156	WAC06830.1		
46	Chaperonin GroL	4	156	WAC08671.1		
47	(Fe-S)-binding protein	5	127			BDMCk91_1046562
48	(Fe-S)-binding protein	2	121			Ga0206388_11077202
49	Chaperonin GroEL	2	106			BDMCk91_1074241

Table 3. BLASTP search results for ORM2a "Magic" and "Nanopod" proteins against MAG4_ORM2b and publicly available sequence databases in GenBank and JGI (as of August, 2024). NCBI accession numbers and percent amino acid identities are shown for best BLASTP hits to select MAGs and bins producing significant alignments (e-value < 1.0E-05, coverage ≥ 50%) to ORM2a proteins. The inset heat map also denotes the percent identity of aligned sequences, where more homologous sequences are indicated with darker shades of red. Best BLASTP hit search results for other ORM2a proteins of interest and other reference genomes are provided in Table S9.

		Protein accession no. of best BLASTP hit							
Protein cluster #	**ORM2a protein accession no. (NCBI)**	**MAG4_ORM2b**	***Pelotomaculum* candiate BPL MAG**	**Hydrothermal vent MAG AUK045**	**Hydrothermal vent bin B94_G9**	**Hydrothermal vent bin B7_G9**	**Hydrothermal vent MAG, Pescadero Basin**	**Wastewater MAG, Germany**	**Freshwater sediment MAG, Oklahoma**
		Toth et al. 2022 [47]	Ga0073689 (IMG)	GCA_021163815.1	GCA_003646755.1	GCA_003646785.1	GCA_021160085.1	GCA_017996625.1	GCA_016935335.1
Highly expressed "Magic" gene cluster putatively involved in benzene activation									
23	WAC08366.1	MAG4_03434 (96%)	Ga0073689_14510 (76%)	MCD6262100.1 (72%)	RLB10792.1 (71%)	RLB12343.1 (71%)	MCD6353789.1 (25%)	MBP7321654.1 (24%)	MBN2467806.1 (27%)
24	WAC08367.1	MAG4_03435 (95%)	Ga0073689_14513 (76%)	MCD6262101.1 (73%)		RLB12342.1 (73%)			
25	WAC08368.1	MAG4_03436 (97%)	Ga0073689_1454 (74%)	MCD6261936.1 (76%)				MBP7322676.1 (22%)	
26	WAC08369.1	MAG4_03437 (97%)	Ga0073689_1455 (70%)	MCD6261937.1 (69%)	RLB11855.1 (66%)	RLB11378.1 (28%)	MCD6352761.1 (26%)	MBP7323401.1 (29%)	MBN2467515.1 (29%)
27	WAC08370.1	MAG4_03438 (94%)	Ga0073689_1456 (66%)	MCD6261938.1 (62%)	RLB11856.1 (61%)			MBP7321894.1 (24%)	
28	WAC08371.1	MAG4_03440 (98%)	Ga0073689_14514 (74%)	MCD6262102.1 (65%)	RLB11323.1 (72%)	RLB12341.1 (69%)			MBN2468817.1 (28%)
29	WAC08372.1	MAG4_03441 (98%)	Ga0073689_14515 (69%)	MCD6262103.1 (60%)	RLB11322.1 (62%)	RLB12340.1 (62%)			MBN2468817.1 (29%)
30	WAC08373.1	MAG4_03442 (90%)		MCD6262763.1 (61%)	RLB08760.1 (51%)				
31	WAC08374.1	MAG4_03443 (90%)		MCD6262762.1 (60%)	RLB08759.1 (52%)				
11	WAC08375.1	MAG4_03444 (94%)	Ga0073689_10360 (64%)	MCD6262187.1 (27%)		RLB11401.1 (57%)	MCD6352639.1 (21%)	MBP7322437.1 (62%)	
Highly expressed "Nanopod" gene cluster putatively involved in benzene tolerance									
32	WAC07726.1	MAG4_01804 (84%)		MCD6261932.1 (44%)		RLB12413.1 (26%)	MCD6353230.1 (30%)	MBP7322355.1 (23%)	MBN2468008.1 (29%)
33	WAC07727.1	MAG4_01802 (94%)		MCD6261931.1 (60%)	RLB08538.1 (32%)	RLB12412.1 (36%)	MCD6353229.1 (31%)	MBP7323620.1 (23%)	MBN2468009.1 (33%)
34	WAC07728.1	MAG4_01798 (82%)	Ga0073689_14410 (32%)	MCD6261930.1 (43%)	RLB11670.1 (31%)	RLB11952.1 (29%)		MBP7321527.1 (28%)	MBN2466607.1 (31%)
35	WAC07729.1	MAG4_01797 (92%)							
36	WAC07730.1	MAG4_01796 (88%)		MCD6261929.1 (46%)			MCD6352593.1 (31%)	MBP7323820.1 (30%)	MBN2468006.1 (30%)
37	WAC07731.1	MAG4_01795 (92%)		MCD6261928.1 (51%)	RLB10794.1 (33%)		MCD6352592.1 (35%)	MBP7323819.1 (36%)	MBN2468007.1 (34%)

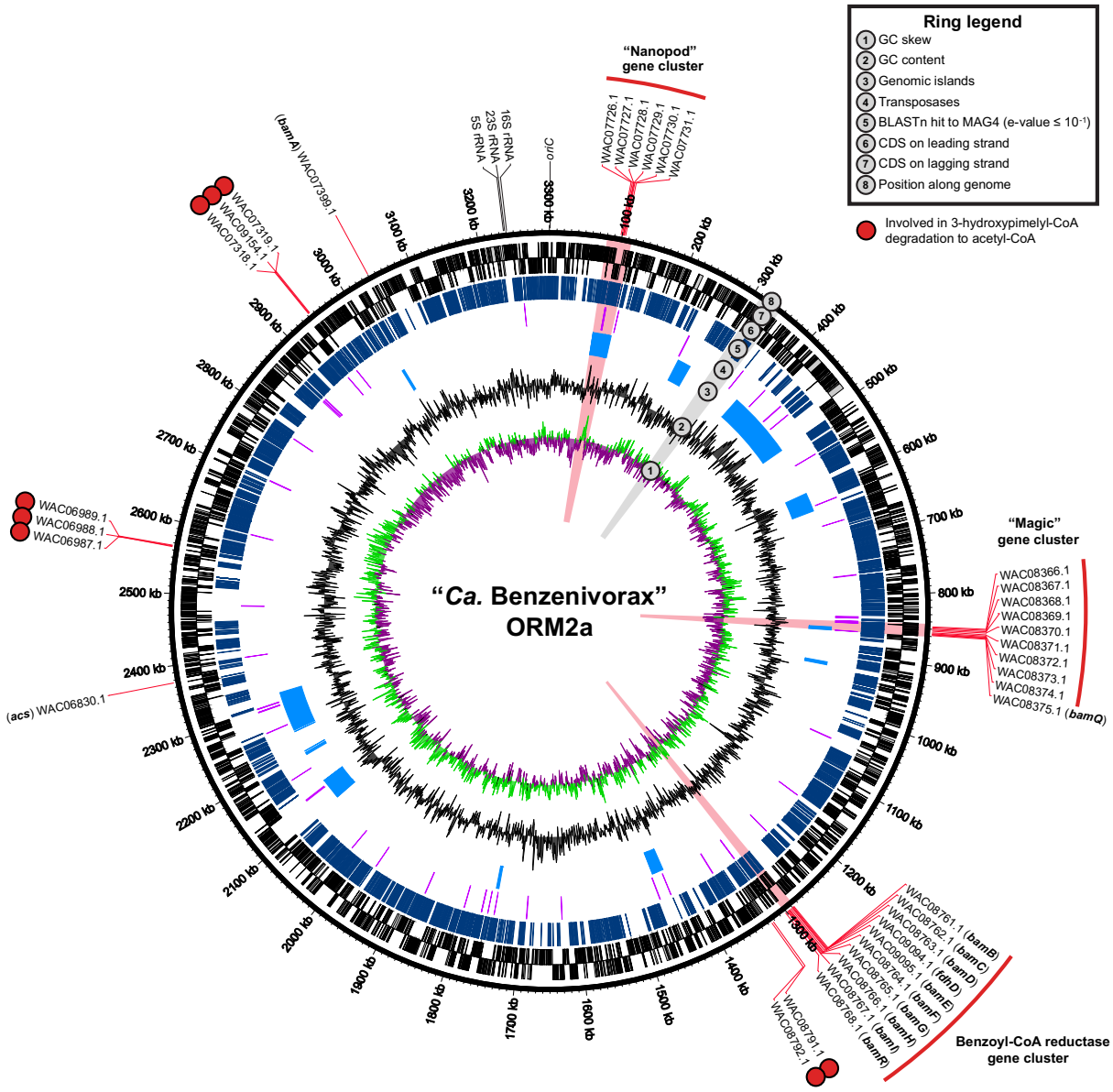

Figure 1. The complete ORM2a genome. Genomic neighbourhoods and genes of interest whose products were recovered with high certainty (≥99% confidence) from the OR proteomes are highlighted in red and are overviewed in Table 2. Genes whose products are involved in the degradation of 3-hydroxypimelyl-CoA to acetyl-CoA are marked with red circles. Protein accession numbers are shown; corresponding ORM2a gene locus tags are provided in Table S7. The positions of the origin of replication *(oriC)* and ribosomal RNA genes are also shown in this figure.

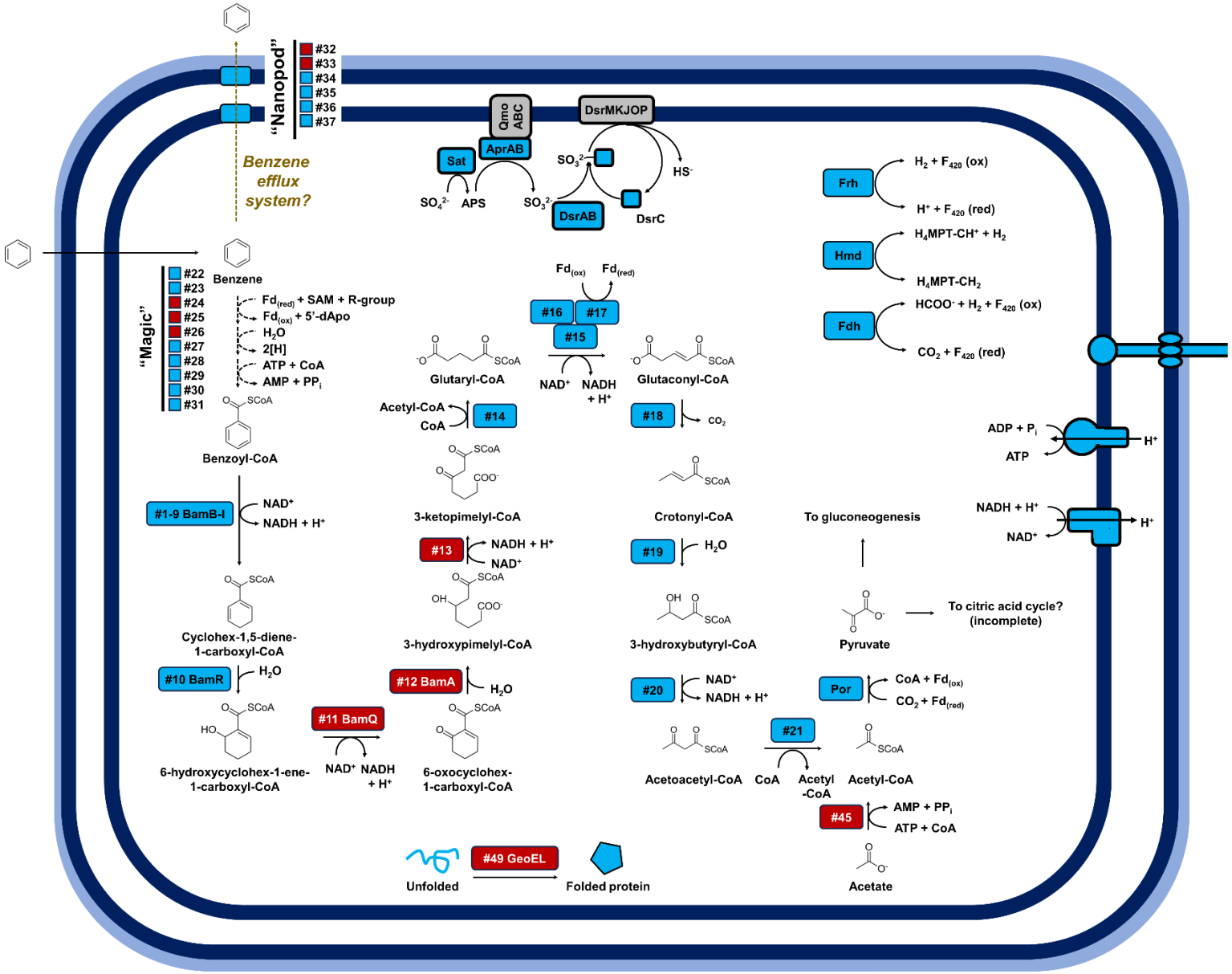
"Nanopod"
#32
#33
#34
#35
#36
#37
Benzene efflux system?
Qmo ABC
DsrMKJOP
AprAB
Sat
SO_4^{2-}
APS
SO_3^{2-}
SO_3^{2-}
DsrAB
DsrC
HS^-
Frh
$H_2 + F_{420}$ (ox)
$H^+ + F_{420}$ (red)
Hmd
$H_4MPT\text{-}CH^+ + H_2$
$H_4MPT\text{-}CH_2$
Fdh
$HCOO^- + H_2 + F_{420}$ (ox)
$CO_2 + F_{420}$ (red)
"Magic"
#22
#23
#24
#25
#26
#27
#28
#29
#30
#31
Benzene
$Fd_{(red)}$ + SAM + R-group
$Fd_{(ox)}$ + 5'-dApo
H_2O
2[H]
ATP + CoA
AMP + PP_i
Benzoyl-CoA
#1-9 BamB-I
NAD^+
$NADH + H^+$
Cyclohex-1,5-diene-1-carboxyl-CoA
#10 BamR
H_2O
6-hydroxycyclohex-1-ene-1-carboxyl-CoA
#11 BamQ
NAD^+
NADH + H^+
#12 BamA
H_2O
6-oxocyclohex-1-carboxyl-CoA
#13
$NADH + H^+$
NAD^+
3-hydroxypimelyl-CoA
#14
Acetyl-CoA
CoA
3-ketopimelyl-CoA
Glutaryl-CoA
$Fd_{(ox)}$
$Fd_{(red)}$
#16
#17
#15
NAD^+
NADH + H^+
Glutaconyl-CoA
#18
CO_2
Crotonyl-CoA
#19
H_2O
3-hydroxybutyryl-CoA
#20
NAD^+
$NADH + H^+$
Acetoacetyl-CoA
#21
CoA
Acetyl -CoA
Acetyl-CoA
#45
AMP + PP_i
ATP + CoA
Acetate
Por
CoA + $Fd_{(ox)}$
$CO_2 + Fd_{(red)}$
Pyruvate
To citric acid cycle? (incomplete)
To gluconeogenesis
ADP + P_i
ATP
H^+
NADH + H^+
NAD^+
H^+
Unfolded
#49 GeoEL
Folded protein
Known enzymatic reaction
Hypothetical/proposed enzymatic reaction
High abundance protein (≥100 matching spectra)
Low abundance protein (<99 matching spectra)
Not detected

Figure 2. Summary of key metabolic pathways identified in the ORM2a genome. The total number of spectra mapping to specific genes in Experiment #3b was used to estimate protein abundances. Other complete and incomplete metabolic pathways identified in the ORM2a genome, including matching spectrum counts for associated genes, are summarized in Table S6. Annotations for numbered proteins are provided in Table 2. Abbreviations; SAM, *S*-adenosyl methionone; 5'-dApo, 5'-deoxyadenosine; Fd, ferredoxin; Sat, sulfate adenylyltransferase; Apr, adenylyl-sulfate reductase; Dsr, dissimilatory sulfite reductase; Frh, [NiFe] hydrogenase; Hmd, 5,10-Methenyltetrahydromethanopterin hydrogenase; Fdh, formate dehydrogenase; F_{420}, coenzyme F420; H_4MPT, tetrahydromethanopterin; Por, pyruvate synthase.

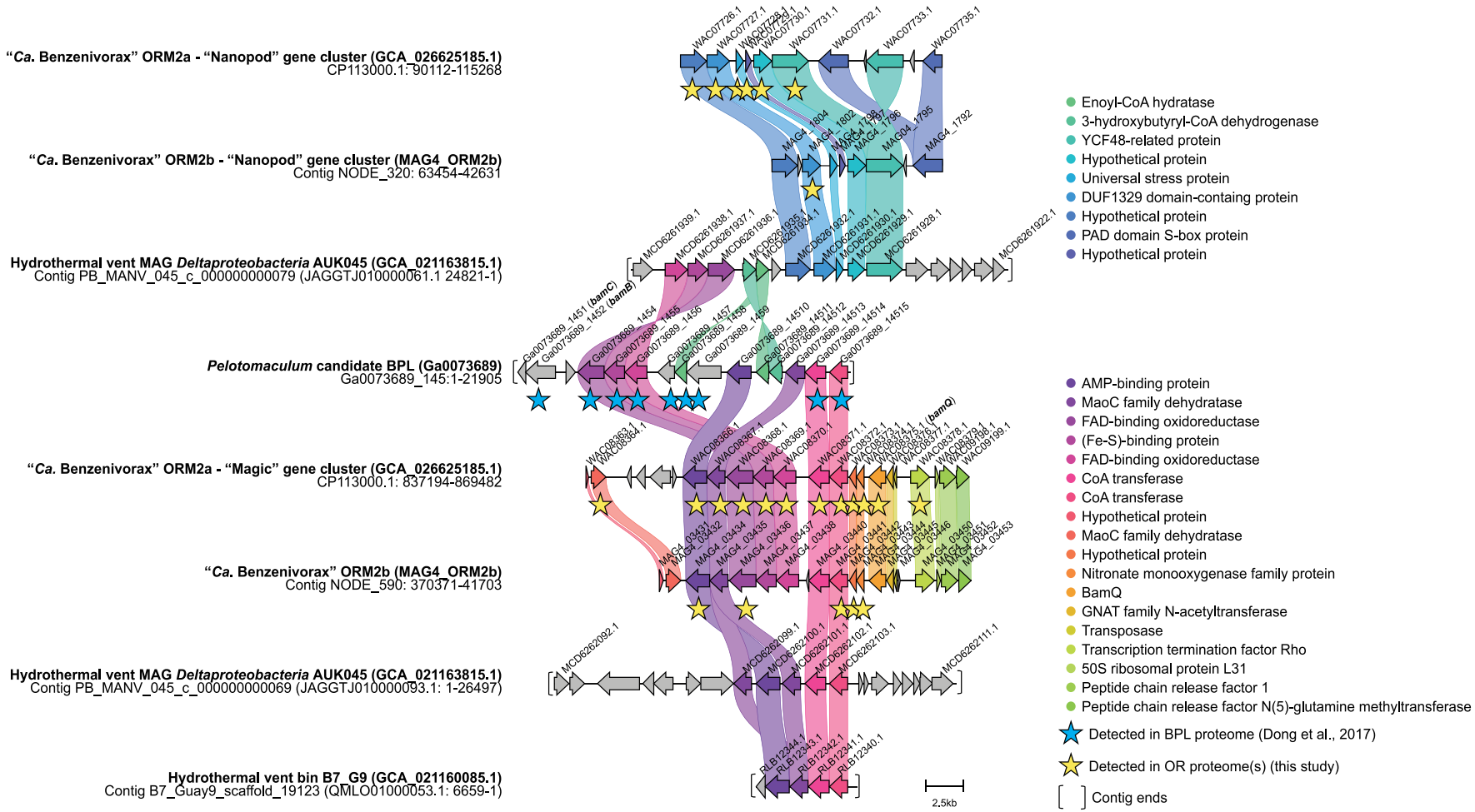

Figure 3. Genomic neighborhoods of all publicly available MAGs in NCBI and IMG producing significant alignments to the "Magic" and "Nanopod" gene clusters in ORM2a and MAG4_ORM2b. Homologous genes are color-coded by function; sequence alignment statistics for relevant proteins are shown in Table 3; additional alignment statistics are provided in Table S9. Corresponding products detected in the proteomes of the OR consortium and the enrichment culture BPL are marked with yellow or blue stars, respectively. The ends of a contig are capped with brackets.

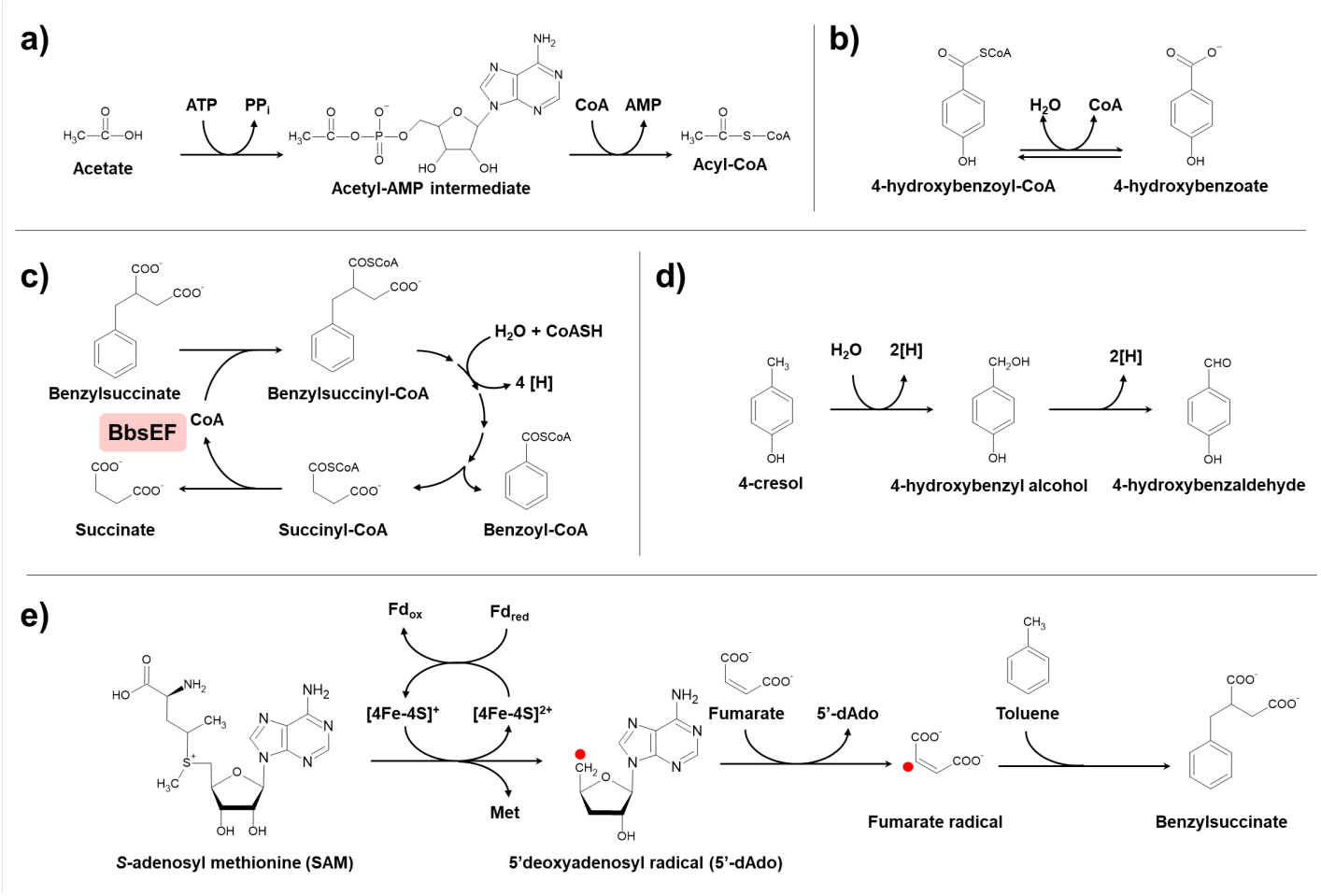

Figure 4. Known catabolic pathways that may be analogous to reactions catalyzed by products of the "Magic" gene cluster. Examples shown are for a) acetyl-coenzyme A synthetase (AMP forming) [85], b) 4-hydroxybenzoyl-CoA thioesterase [97], c) succinyl-CoA:(*R*)-benzylsuccinate CoA transferase (BbsEF) [98], d) 4-cresol dehydrogenase [89], and e) benzylsuccinate synthase [98].

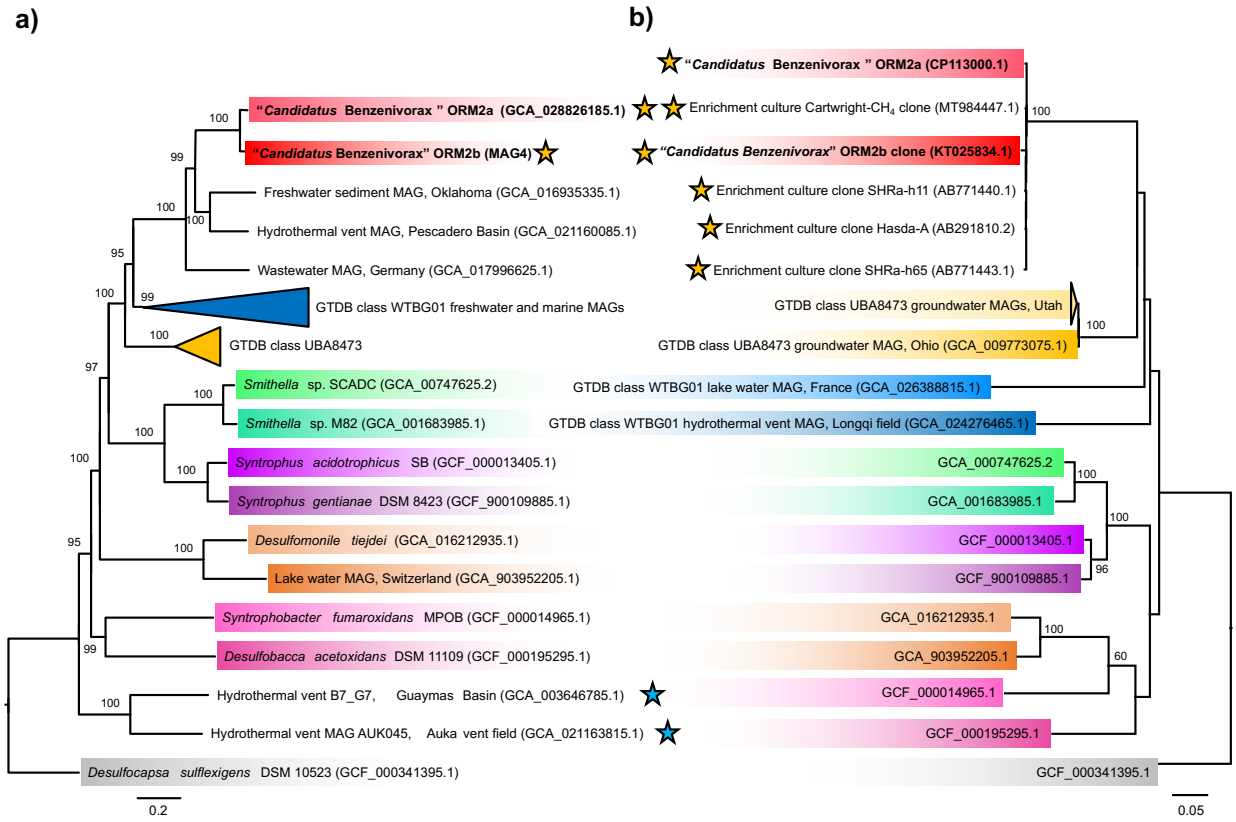

Figure 5. Taxonomic classification of ORM2a and reference *Desulfobacterota* species based on a) genome and b) 16S rRNA gene maximum likelihood phylogenetic trees. Sequences from this study are highlighted in bold. Sequences belonging to known and predicted benzene-degrading anaerobes are marked with yellow stars; putative anaerobic benzene degraders (based on functional gene evidence presented in this study) are denoted with blue stars. *Desulfocapsa sulflexigens* was used to root both trees. Bootstrap values <60% are not shown.