Identification of a Highly Expressed Gene Cluster Likely Coding for Benzene Activation Enzymes in a Methanogenic Enrichment Culture

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 Z. 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, , Toronto, Ontario, M5G 1L7, 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 50% 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 complete circularized 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 ORM2a gene clusters of unknown function. Syntenic gene clusters were identified in two other known benzene degrader, ORM2b and *Pelotomaculum* candidate BPL, and in nine 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 may pump excess benzene out of cells to mitigate toxicity. Phylogenomic analyses place ORM2a and ORM2b within a novel genus of benzene-degrading specialists which we propose naming "*Candidatus* Benzenivorax".

41 **INTRODUCTION**

42 It has been 45 years since anaerobic benzene degradation was first documented by Ward et al. [1],
43 reversing a long-held tenet that aromatic hydrocarbons were recalcitrant under anaerobic
44 conditions. Notorious for its carcinogenicity, stringent environmental regulations, and persistence
45 in anoxic environments, benzene has been the subject of numerous studies seeking to characterize
46 its metabolism in the absence of oxygen [2, 3]. Laboratory enrichment culture studies have
47 confirmed unequivocally that anaerobic benzene degradation can be coupled to iron reduction,
48 nitrate reduction, sulfate reduction, and methanogenesis, catalyzed by a handful of specialized
49 microbial clades [2, 4]. But despite great efforts, the biochemical mechanism(s) by which benzene
50 is activated in the absence of oxygen remain elusive.

51 Initial studies using ^{13}C- or ^{14}C-benzene and occasionally $H_2{}^{18}O$ identified three possible
52 activation reactions based on the detection of labelled metabolites: (i) anaerobic hydroxylation of
53 benzene yielding phenol [5-12], (ii) carboxylation of benzene yielding benzoate [6, 8-10, 13, 14],
54 and (iii) methylation of benzene yielding toluene [10]. Neither phenol nor benzoate are diagnostic
55 of a single activation mechanism [15], and compelling evidence of benzene methylation has been
56 sparse in literature [16, 17]. From 2003-2008, hydrogen and carbon isotope fractionation ratios
57 were evaluated in active benzene-degrading enrichment cultures [18-20]. The slope of dual isotope
58 plots ($\Delta\delta\,^{2}H/\Delta\delta\,^{13}C$, a value defined as Λ) from nitrate-reducing cultures (Λ = 8 to 19, n = 5) was
59 much lower than for sulfate-reducing and methanogenic cultures surveyed (Λ = 28 to 31, n = 7),
60 supporting that at least two biochemical benzene activation mechanisms exist in nature [18].

61 By 2010, developments in DNA sequencing and analysis techniques enabled meta'omic
62 characterization of anaerobic benzene degraders for the first time [21-23]. Abu Laban *et al*. [24]
63 sequenced the metagenome of the iron-reducing enrichment culture BF and identified a protein

with homology to a known phenylphosphate carboxylase (PpcA). The authors located its corresponding putative anaerobic benzene carboxylation gene (*abcA*) on a contig affiliated with *Clostridia*. Expression of the putative *abcA* and neighboring genes (*abcD*, benzoyl-CoA ligase [*bzlA*] and a *ubiX*-like gene) was induced with benzene or benzoate but not with phenol, inferring a possible role in benzene activation [24]. Syntenic gene clusters have since been identified in the transcriptomes of several nitrate-reducing cultures [25-27], including one previously analyzed by isotopic fractionation [18, 19]. Most nitrate- and iron-reducing benzene degraders identified thus far have been affiliated with the order *Thermincolales* [25-29]; a putative benzene-carboxylating, sulfate-reducing *Thermincolales* was also recently identified in a jet fuel-contaminated aquifer via metagenomics [30]. Researchers are now attempting to verify benzene carboxylation using structural and biochemical assays, among other tools [31].

Evidence for benzene hydroxylation has only been shown for one microorganism, *Geobacter metallireducens* GS-15, via gene deletion. In 2013-2014, Zhang *et al.* [5, 23] identified a handful of genes that were upregulated during active phenol or benzene metabolism, and the deletion of key genes (Gmet0231, Gmet0232, or *ppsA*) inhibited the degradation of benzene. The gene *ppsA* encodes for the alpha subunit of phenylphosphate synthase and is required for catalyzing the first step in anaerobic phenol metabolism [32, 33]. Functional annotations suggest that the genes Gmet0231 and Gmet0232 encode for putative oxidoreductases, but their exact functions are unknown. In 2024, Bullows *et al.* [17] identified a fumarate addition gene cluster encoding for 3-hydroxybenzylsuccinate synthase and a putative methyltransferase in the genome of *Geotalea daltonii* strain FRC-32, a close relative of *Geobacter*. Expression of these genes – which are only known to catalyze the activation of alkanes and alkyl-substituted aromatic rings [34] – was upregulated during active benzene metabolism but not with toluene. Addition of

87 benzene to *G. daltonii* whole-cell lysates resulted in near-stoichiometric formation of toluene,
88 corroborating that methylation could indeed be a biochemical benzene activation mechanism [17].

89 Homologs of the known and predicted anaerobic aromatic-activating genes listed above
90 have never been detected in methanogenic benzene-degrading enrichment cultures [4, 21] nor in
91 the well-characterized sulfate-reducing enrichment culture BPL [9, 35]. Genes associated with
92 aerobic metabolism (monooxygenases) and the first step of anaerobic growth on benzoate
93 (benzoyl-CoA ligase) have also never been found in these cultures [35]. Thus far, metagenomic
94 and proteomic surveys of representative cultures have only identified the presence and expression
95 of downstream ATP-independent benzoyl-CoA biodegradation pathways during active benzene
96 degradation [21, 35]. Benzoyl-CoA is a central metabolic intermediate of all proposed anaerobic
97 benzene degradation pathways [16, 36]. Given this information and previous isotope fractionation
98 data [18, 19] indicating that the mechanism of benzene activation catalyzed by sulfate-reducing
99 and methanogenic cultures is distinct from that in nitrate-reducing cultures (where carboxylation
100 is likely), we surmise that a novel hydrocarbon activation mechanism exists that catalyzes
101 anaerobic transformation of benzene to benzoyl-CoA.

102 In this study, we integrated the information from three proteomic experiments on the
103 methanogenic benzene-degrading OR consortium. The culture, which has been maintained on
104 benzene for over 25 years [4, 10, 18, 19, 21, 37-43], harbors two closely related strains of benzene-
105 degrading *Desulfobacterota* (ORM2a and ORM2b) that make up on average 56% of total bacteria.
106 The culture also contains methanogenic archaea (*Methanothrix* and *Methanoregula*), and >100
107 other typically low abundance microorganisms with various predicted functions [39, 41, 42]. Initial
108 proteomics experiments were conducted on the OR consortium in the early 2010's [21] but results
109 were difficult to interpret due to the lack of a sound metagenome and low protein yields [21]. In

110 the years since, improvements in culture maintenance [39, 40] allowed for greater protein yields
111 and new sequencing efforts have enabled the reconstruction of two higher quality metagenomes
112 and 74 metagenome-assembled genomes (MAGs) [41]. The genome of ORM2a was also
113 successfully closed (NCBI accession no. CP113000.1), making it a valuable blueprint for
114 deciphering anaerobic benzene degradation. Here, we discovered that the most abundant proteins
115 expressed by the culture mapped to a unique cluster of syntenic genes found only in other benzene-
116 degrading obligate anaerobes and a handful of metagenomic contigs from deep hydrocarbon-
117 producing hydrothermal vents.

118

119 **MATERIALS AND METHODS**

120 **Culture maintenance**

121 The OR consortium was enriched from samples from an oil refinery in Oklahoma in 1995 [44] and
122 has been maintained in a near-identical manner ever since. Briefly, cultures are grown in a defined
123 pre-reduced anaerobic mineral medium [38, 40] and are fed benzene (~5-40 mg/L) once every 4-
124 6 weeks. The OR consortium has been subcultured many times, resulting in the emergence of five
125 distinct culture lineages [39, 42]. The "OR-b" lineage, illustrated in Figure S1, is the most well
126 characterized with a doubling time about 30 days [4, 18, 21, 39-43] and was surveyed in this study.
127 A large scale (>100 liters) bioaugmentation culture known as DGG-B [4, 40] was derived from
128 the OR-b lineage. Though the OR consortium has been amended with other substrates including
129 benzoate and toluene, only benzene has been shown to support the growth of ORM2a and ORM2b
130 [21, 37].

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

132 Between 2010-2017, the metagenomes of three OR-b subcultures designated OR-b1A, DGG1A,
133 and DGG-0 were sequenced (Figure S1). The OR-b1A (2010) metagenome was sequenced using
134 Illumina paired-end technology (2×100 bp) and was partially assembled in ABySS v.1.3.2 [45]
135 with contigs binned using varying kmer lengths (Chapter 5 of Devine [21]). Toth et al. [41]
136 assembled the metagenomes of DGG1A (2016) and DGG-0 (2017) using short-read (Illumina
137 paired-end) and long-read (PacBio) shotgun sequences, resulting in longer, high quality contigs.
138 Following binning, three draft MAGs including ORM2a were refined into closed (completed)
139 genomes. The correct assembly of each closed genome was verified by read mapping, and the
140 average sequence depth for the ORM2a genome was 1,384× [41, 46].

141 Taxonomy was assigned to all draft and complete MAGs using release 214 of the Genome
142 Taxonomy Database (GTDB) and GTDB tool kit (v.2.3.2) using the classify_wf workflow [41,
143 47, 48]. In this study, the ORM2a genome (CP113000.1) and a medium-quality MAG putatively
144 belonging to ORM2b, (ORM2b_MAG4, 3,214,903 bp across 378 contigs) were reclassified using
145 release 220 of GTDB and GTDB-Tk v.2.4.0 (discussed later). Seventy-one medium-to-high
146 quality MAGs including ORM2b_MAG4 were retained. All metagenomic assemblies and
147 complete MAGs were submitted to NCBI and/or IMG for automated gene calling and functional
148 annotation. Incomplete (noncircular) MAGs were deposited to figshare [46] and in this study were
149 annotated using RASTtk [49] in March, 2022. Data availability including accession numbers and
150 all functional annotation pipelines used is provided in Table S1.

151 **Analysis of the ORM2a genome**

152 To assess the metabolic potential of the ORM2a genome, coding sequences predicted by the NCBI
153 Prokaryotic Genome Annotation Pipeline were uploaded to BlastKOALA for automatic KEGG

154 Orthology (KO) assignment [50], then organized into metabolic pathways using the Reconstruct
155 tool within KEGG Mapper [51, 52]. Next, the mummer2circos package in GitHub
156 (https://github.com/metagenlab/mummer2circos) [53] was used to generate a circular plot of the
157 ORM2a genome and to compare its homology with ORM2b_MAG4 (Figure 1). Two-way average
158 nucleotide identity (ANI) analysis of both genomes was performed using an online calculator
159 (http://enve-omics.ce.gatech.edu/ani/). Ori-Finder 2022 [54] was used to identify the genomes'
160 origin of replication (*oriC*). IslandViewer 4 [55] and PHASTER [56] were used to identify
161 genomic islands and prophage sequences, respectively. Functional annotations and hidden Markov
162 model assignments (from NCBI and IMG) were used to identify transposase genes. Putative
163 insertion sequences were identified using ISfinder [57]. Operon-mapper [58] was used to identify
164 operons in the ORM2a genome.

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

166 Between 2010-2018, three proteomics experiments were performed on three OR-b lineage
167 subcultures (OR-b1C, OR-b, and OR-b1A; see Figure S1). Culture information, including culture
168 volumes extracted and benzene degradation rates at the time of sampling, are provided in Table 1.
169 To avoid oxygen contamination, all containers used were incubated in an anaerobic chamber
170 (supplied with a gas mix of 10% H_2, 10% CO_2 and 80% N_2) for >24 hours prior to use, and all
171 steps were performed at 4 °C unless otherwise specified.

172 Protein extractions for Experiment #1 (OR-b1C, 2010) and Experiment #2 (OR-b, 2011)
173 were conducted by Devine [21] followed by LC-MS/MS sequencing of trypsin-digested crude
174 extracts. Detailed sample preparation methods can be found in Text S1. In Experiment #3 (OR-
175 b1A, 2018), cell pellets from two 50 mL volumes of culture (10,000 x g, 15 min) were suspended
176 in an anoxic sodium dodecyl sulfate (SDS)-containing lysis buffer (50 mM Tris-HCl [pH 7.6], 2%

177 w/v SDS, and 50 mM DDT) and sonicated at 23-30W for 10 min in pulse mode (1 s ON/1 s OFF
178 for 1 min, 1 min OFF, repeat). Proteins from one crude lysate (Experiment #3a) were purified
179 using a Microcon-30kDa centrifugal filter unit (Millipore) as outlined by Wiśniewski et al. [59],
180 then subjected to an 18-24 hr in-solution digestion with porcine trypsin at 37 °C [21]. The second
181 crude lysate (Experiment #3b) was mixed with 2× SDS gel-loading buffer (50 mM Tris-HCl [pH
182 6.8], 10% SDS [sodium dodecyl sulfate], 40% glycerol, 3 mM bromophenol blue, and 500 mM
183 DDT]) and separated on a 15% SDS-PAGE gel run for 45 min at 200 V. Visualization of the
184 resulting gel revealed a long protein smear with no distinct bands (Figure S2), which was sliced
185 into five equal sections then destained and digested with trypsin as described in Shevchenko et al.
186 [60].

187 LC-MS/MS analysis of Experiments #1 and #2 peptides [21] was performed at the
188 SickKids SPARC BioCentre using an LTQ Orbitrap mass spectrometer (Thermo Fisher
189 Scientific). Run parameters are summarized in Tang et al. [61]. LC-MS/MS analysis of Experiment
190 #3 peptides was conducted at the University of Toronto BioZone Mass Spectrometry Facility.
191 Briefly, 5 µL liquid samples of peptides were separated on a 15cm PicoTip Emitter (New
192 Objective) packed with ReproSil-Pur C18-AQ 3 µm resin (Dr. Maisch GmbH). The flow rate was
193 set to 250 nL/min and the eluents used were (A) water containing 0.1% formic acid, and (B)
194 acetonitrile containing 0.1% formic acid. The gradient started at 0% B, followed by a linear
195 gradient to 10% B over 5 min, a linear gradient to 40% B over 88 min, a linear gradient to 95% B
196 over 2 min, a hold at 95% B for 10 min, a linear gradient to 0% B over 1 min, and a final hold of
197 0% B for 14 min (total runtime of 120 min). MS detection was conducted using a Q-Exactive
198 Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a nanoelectrospray
199 ionization probe operating in positive ionization mode, with a spray voltage of 2.8 kV, capillary

200 temperature of 275°C, and S-lens radio frequency level of 55. Full MS data was acquired with an
201 m/z range of 400-2,000, mass resolution of 70,000, automatic gain control (AGC) target of
202 1.0E+06, and a maximum injection time of 30 milliseconds. MS2 was gathered using a data
203 dependant TOP10 approach with a mass resolution of 17,500, AGC target of 5.0E+04, maximum
204 injection time of 50 ms.

205 **Protein identification**

206 Raw mass spectrometry data from all three experiments were converted into mzXML using
207 MSconvert [62] then uploaded into Version 2020.11.12.1 of X! Tandem (The GPM, thegpm.org)
208 for peptide and protein discovery. Spectra were screened against a large library of protein
209 sequences (1,595,050 entries) with a mass tolerance of <0.4 Da for the fragment ions and <20 ppm
210 for the precursor ion. The library, provided in File S1 and overviewed in Table S1, contained
211 1,106,523 translated gene sequences called by IMG (three metagenomic assemblies and 3 closed
212 MAGs), NCBI (3 closed MAGs), and RAST-tk (67 draft MAGs). The database also included
213 488,411 reverse-translated (decoy) protein sequences from the DGG0 metagenome to calculate a
214 false discovery rate, and 116 entries from common human/laboratory contaminants. Peptide and
215 protein identifications were accepted if they could be established at greater than 95% and 99%
216 probability, respectively [63, 64]. Proteins sharing peptide evidence were grouped into clusters.

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

218 Protein identity was verified using protein BLAST searches against all other OR metagenomic
219 assemblies and MAGs, as well as reference (meta)genomic sequences in NCBI (nr), JGI (IMG/M),
220 and UniProt (UniProt KB reference proteomes + Swiss-Prot). Next, the sequence-based tool
221 MotifFinder [65] was used to identify conserved domains in each protein. Protein structures were
222 predicted using the AlphaFold2.ipynb tool within Colab [66, 67], then compared to homologous

223 structures and protein domains within the Dali server's Protein Data Bank [68] and DeepFRI [69].
224 All AlphaFold 2 protein structural models generated are provided in File S2. Maximum likelihood
225 trees were constructed in RAxML v.8.2.1.1 within Geneious v.8.1.9 [70, 71] to corroborate
226 predicted protein functions where possible. Protein sequences were also sent to the laboratory of
227 Prof. Christine Orengo (University College London) for independent analysis using CATH v.4.4
228 [72] and PROST [73].

229 **Phylogenomic analyses**

230 Phylogenomic analysis of the closed ORM2a genome and ORM2b_MAG4 was performed in
231 v.1.8.8 of GToTree [74]. Briefly, we retrieved all representative genomes from c__UBA8473 and
232 c__WTBG01 from GTDB v.220, as well as closely related genomes from selected
233 *Desulfobacterota*, then extracted target genes using the Proteobacteria.hmm single copy gene set
234 (119 genes) within GToTree. *In silico* translated protein sequences were aligned using MUSCLE
235 [75] and concatenated. A maximum likelihood tree was constructed from the concatenated
236 alignment using IQ-TREE v.2.3.6 [76] with the LG+F+I+R5 substitution model, inferred as best
237 model by IQ-TREE, and 100 bootstraps. For comparative purposes, a second maximum likelihood
238 tree was constructed using 16S rRNA gene sequences extracted from each representative GTDB
239 genome (if available) and from clone sequences of known/predicted anaerobic benzene degraders.
240 *Desulfocapsa sulflexigens* DSM 10523 (GCA_000341395.1) was included as an outgroup.

241

242 **RESULTS AND DISCUSSION**

243 **General features of the ORM2a genome**

244 The ORM2a genome consists of a single chromosome of 3,300,546 bp with low GC skew and no
245 definitive terminus region (Figure 1, Figure S3). General genome statistics are summarized in
246 Table S2 and supplementary details are provided in Text S2. In total, 1,431 out of 3,135 genes
247 identified in the ORM2a genome (46%) were connected to KEGG Orthology modules and could
248 be organized into metabolic pathways. Key pathways are overviewed below and in Figure 2; all
249 other complete and incomplete pathways identified are summarized in Table S3. The ORM2a
250 genome suggests limited metabolic potential to degrade carbohydrates, lipids, amino acids, and
251 nucleic acids. Pathways for electron transport and ATP generation are also limited, consisting of
252 an F-type ATPase, a Nuo-type NADH dehydrogenase [complex I] with one missing subunit
253 (*nuoG*), and a dissimilatory sulfate reduction pathway with one missing subunit (*dsrC*). Previous
254 experimental evidence suggested that the OR consortium could couple benzene oxidation to sulfate
255 reduction [4], but this is the first genomic evidence supporting this observation. BLAST searches
256 and manual inspections confirmed that the ORM2a genome does not contain homologs to any
257 known hydrocarbon activation genes or to benzoyl-CoA ligase (*bzlA*). The latter is a significant
258 distinction between microorganisms containing putative anaerobic benzene carboxylase genes
259 (*abcAD*) and those that do not [35]. However, the presence of a complete ATP-independent
260 benzoyl-CoA degradation pathway is consistent with ORM2a's involvement in benzene oxidation
261 (Figures 1 and 2).

262 Only 29 complete anabolic pathways were identified in the ORM2a genome, including
263 biosynthesis pathways for 7 amino acids and 4 vitamins (Table S3a). In contrast, 197 genes with
264 putative transport functions were identified (Table S4) – including transporters for vitamin B_{12} and
265 undefined amino acids – hinting that ORM2a imports most of its nutrients from extracellular

266 sources, perhaps from polymeric substances or from direct exchanges with other microorganisms
267 as theorized in recent microscopy and genomic investigations of the OR culture [42, 43]. The
268 ORM2a genome contains 8 operons with genes coding for predicted hydrogenases and/or formate
269 dehydrogenase (Figure 2 and Table S5), whose products most likely feed into methanogenic
270 pathways [77]. The genome also encodes for type IV pili and a Che-type chemotaxis system that
271 may be involved in bacterial motility. Overall, these findings suggest that ORM2a is likely an
272 obligate syntrophic anaerobe uniquely adapted for anaerobic benzene degradation, consistent with
273 all previous culture studies [4, 21, 39, 40].

274 **Proteomic analyses**

275 A total of 858 unique predicted proteins were identified from the three proteomics experiments on
276 OR subcultures (Table 1 and Supplemental Text S3). Accession numbers, predicted functional
277 annotations, amino acid sequences, and peptide/spectra counts are provided in Table S9. Peptide
278 and protein statistics are provided in Tables S10a and S10b, respectively. Proteins that mapped to
279 one or more MAGs, are putatively involved in anaerobic benzene degradation, and/or were
280 detected in high abundance in Experiment #3b (≥100 matching spectra) are highlighted in Table 2
281 and Figures 1-2. Most of the high abundance proteins identified in Experiment #3b were also
282 detected in 1 to 4 other OR proteomes (Table S9), illustrating the robustness of our experimental
283 results despite years between experiments and differences in protein extraction efficiencies.

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

285 Three-hundred and sixty-two proteins (42% of proteins identified) mapped unambiguously to the
286 closed MAG of ORM2a (Table 1), covering 11% of this bacterium's total putative proteome. We
287 also identified 33 unique proteins (4%) that mapped to a draft MAG putatively belonging to
288 ORM2b (ORM2b_MAG4, 378 contigs), and 44 proteins (5%) that mapped to both ORM2a and

289 ORM2b_MAG4 (Table 1). The draft ORM2b_MAG4 genome does not contain a 16S ribosomal
290 RNA gene sequence for comparison to a reference clone sequence of ORM2b (NCBI accession
291 no. KT025834). However, the high average nucleotide identity (89.94%) between ORM2a and
292 ORM2b_MAG4 supports its correct classification (see ring 5 of Figure 1 and Figure S4). Most
293 ORM2b_MAG4 proteins identified were highly similar (95-100% identity) to ORM2a proteins
294 described below and summarized in Tables 2 and 3. No other MAGs assembled from the OR
295 culture aligned significantly to the ORM2 MAGs.

296 One-quarter of proteins identified mapped to MAGs of methanogenic archaea (Tables 1
297 and S9), specifically acetoclastic *Methanothrix* spp. (170 proteins across 8 MAGs) and
298 hydrogenotrophic *Methanoregula* spp. (19 proteins across 3 MAGs). Most high abundance
299 archaeal proteins were affiliated with methanogenesis and mapped to more than one MAG (Table
300 2). Fifty low-abundance proteins mapped to other bacterial MAGs including "*Ca.*
301 Nealsonbacteria" DGGOD1a. The remaining proteins could not be assigned to a specific MAG
302 (171 proteins, 20%) or mapped to contaminant/decoy sequences (20 proteins, 2%).

303 **Benzoyl-CoA degradation to acetyl-CoA**

304 The OR proteomes were compared to representative sequences of known/predicted hydrocarbon
305 activation enzymes as well as to benzoyl-CoA ligase, but once again we were unable to identify
306 any homologs. Next, we searched for products of downstream aromatic degradation pathways and
307 detected a complete suite of putative benzoyl-CoA degradation enzymes mapping to the ORM2a
308 genome (Figures 1 and 2, Tables 2). Ten gene products (designated Proteins #1-10) mapped to a
309 single gene cluster and were annotated as subunits of a Bam-type benzoyl-CoA reductase
310 (BamBCDEFDHI; whose components are described in Wischgoll et al. [78] and Boll et al. [79]),
311 an FdhD-like insertase, and a cyclohexa-1,5-diene-1-carbonyl-CoA hydratase (BamR). Two high

312 abundance proteins annotated as 6-hydroxycyclohex-1-ene-1-carbonyl-CoA dehydrogenase
313 (BamQ, Protein #11) and 6-oxo-cyclohex-1-ene-carbonyl-CoA hydrolase (BamA, Protein #12)
314 mapped to other regions of the ORM2a genome (Figure 1). Multiple sequence alignments to
315 reference sequences (Table S11) and maximum likelihood phylogenetic tree analyses (Figure S5)
316 support the correct annotation of all predicted Bam proteins.

317 Nine proteins predicted to catalyze the transformation of 3-hydroxypimelyl-CoA into
318 acetyl-CoA were also identified in the OR proteomes and mapped exclusively to the genome of
319 ORM2a (Figures 1 and 2, Table 2). In Experiment #3b, higher protein abundances were recovered
320 for Proteins #13-17, involved in 3-hydroxyglutaryl-CoA transformation to glutaconyl-CoA, than
321 for downstream beta-oxidation enzymes (Proteins #18-21). This could indicate a fork in the
322 pathway (Figure 2) and warrants future investigation. Genome inspections revealed that the gene
323 coding for a non-decarboxylating glutaryl-CoA dehydrogenase (Protein #15) was located in the
324 same operon as two genes coding for bifurcating electron transfer flavoproteins (EtfAB; Proteins
325 #16-17) (Tables 2 and S12). Electron bifurcation is a mechanism of biological energy (ATP)
326 conservation that couples endergonic and exergonic redox reactions [80]. As shown in Figure 2,
327 EtfAB may couple the oxidation of NADH + H^+ with the reductions of glutaryl-CoA (exergonic)
328 and ferredoxin (endergonic), akin to the proposed reaction mechanism of EtfAB-butyryl-CoA
329 dehydrogenase complexes [80, 81].

330 **Sulfur metabolism and acetate production**

331 Several unexpected ORM2a proteins were detected in the OR proteomes, including five proteins
332 associated with the dissimilatory reduction of sulfate to hydrogen sulfide (Sat, AprAB, DsrAB)
333 despite not adding any sulfate to the culture medium (Figure 2). Ferrous sulfide (FeS) is however
334 used as a reductant [38] and could result in the formation of sulfate if oxidized. We also detected

335 four subunits (HydABDG) of a sulfhydrogenase that catalyzes the reversible reduction of
336 elemental sulfur or polysulfide to H_2S using hydrogen as an electron donor [82]. This was the only
337 hydrogenase detected in the OR proteomes that mapped to ORM2a (Figure 2 and Table S5),
338 hinting that that this organism may couple syntrophic benzene oxidation with sulfur metabolism.
339 Perhaps ORM2a cells use HydABDG to prevent H_2S buildup or to store sulfur and electrons in the
340 form of polysulfides [83]. This finding warrants a separate investigation.

341 GroEL (Protein #43) and a putative AMP-forming acetyl-CoA synthetase (Acs, #42) were
342 tied as the 5th most abundant proteins in Experiment #3b (Table 2). Acs catalyzes the ATP-
343 dependent formation of acetyl-CoA molecules from acetate and coenzyme A but given that acetyl-
344 CoA is already produced in abundance during anaerobic benzene degradation, it is plausible that
345 Protein #42 catalyzes the reverse reaction yielding acetate (a substrate for acetoclastic
346 *Methanothrix* spp.), CoA, and ATP (Figure 2). Similar reverse reactions have been proposed in
347 the metabolic pathway of syntrophic bacteria in anaerobic digesters and has been supported by
348 metatranscriptomic evidence [84]. No other acetogenic proteins were identified in the OR
349 proteomes, and the ORM2a genome does not appear to encode for other pathways that synthesize
350 acetate. Phylogenetic trees support the correct annotation of Protein #42 (Figure S6).

351 **Highly expressed ORM2a proteins from gene clusters of unknown function**

352 The three top most abundant and recurring proteins identified in the OR proteomes (Proteins #24-
353 26) mapped consecutively to an ORM2a gene cluster of unknown function (see Table 2 and
354 Figures 1 and 3). Products of several neighboring genes (Proteins #22-23, #27-30) were also
355 detected in multiple OR proteomes. This gene cluster, designated herein as the "Magic" gene
356 cluster, is situated on a small genomic island and is flanked by multiple predicted transposases and
357 pseudogenes (Figure 1). Annotations provided by NCBI, IMG, and the Protein Data Bank (PDB)

358 offered few insights into the functions of these proteins (Tables 2 and 4b) with one notable
359 exception – the tenth consecutive gene in the "Magic" gene cluster most likely codes for BamQ
360 (Protein #11) in the benzoyl-CoA degradation pathway (Figure 2). Products of a neighboring,
361 transcription-terminating Rho factor also were detected in the Experiment #3b proteome (Protein
362 #31), suggesting that expression of the "Magic" gene cluster is tightly regulated.

363 Two other highly abundant proteins (#32 and #33, ranked 18th and 4th in Experiment #3b,
364 respectively) mapped consecutively to a second ORM2a genomic island of unknown function
365 (designated the "Nanopod" gene cluster, see Table 2 and Figures 1 and 3) that includes four other
366 expressed genes. Protein #33 contains a domain of unknown function (DUF1329, WAC07727.1)
367 that has previously been identified in the nanopod structures of some hydrocarbon-degrading
368 bacteria including *Delftia acidovorans* Cs1-4 (genome accession no. CP002735.1), an aerobic
369 phenanthrene degrader [85, 86]. Nanopods are tubular structures (<100 nm in diameter) that may
370 facilitate nutrient transport or export toxic compounds from cells [86, 87]. Shetty et al. [86] found
371 that phenanthrene induced nanopod formation in *D. acidovorans* Cs1-4 but not structurally similar
372 hydrocarbons (naphthalene and toluene), and that deletion of its DUF1329 gene (DelCs14_1722)
373 impaired growth on phenanthrene. BLASTP alignments suggest that Protein #33 is distantly
374 related to the translated DelCs14_1722 gene sequence (21% amino acid identity), but the high
375 expression of this ORM2a gene cluster and its possible relation to nanopod formation is intriguing.

376 **Comparative analyses of the "Magic" and "Nanopod" gene clusters to other genomes and**
377 **proteomes**

378 To determine whether other genomes contain syntenic regions to the "Magic" and "Nanopod" gene
379 clusters, translated sequences were searched against all reference genomes in NCBI and IMG (as
380 of August 2024) using BLASTP. Four syntenic contigs were identified (Figure 3 and Table 3). We

381 also identified six short (2-3 gene) contigs containing highly similar genes (51-76% amino acid
382 identity) to the "Magic" gene cluster (Table S11). Interestingly, 9 of the 10 contigs identified were
383 affiliated to *Deltaproteobacteria* spp. from hydrothermal vent systems [88, 89]. Hydrocarbons are
384 important energy sources for microbial communities in these ecosystems and are produced
385 primarily by geological processes [90]. Hydrothermal vent MAG *Deltaproteobacteria* AUK045 is
386 notable for having syntenic contigs to both the "Magic" and "Nanopod" gene clusters (Figure 3
387 and Table 3). Genes encoding for predicted downstream benzoyl-CoA degradation and beta-
388 oxidation pathways were also identified in AUK045 and the two draft *Deltaproteobacteria* bins
389 (Table S11), demonstrating some functional potential to degrade aromatic compounds.

390 *Pelotomaculum* candidate BPL (IMG GOLD accession number Ga0073689) is the chief
391 benzene degrader in a sulfate-reducing enrichment culture from Poland [9, 35]. Dong et al. [35]
392 published proteomic data from this culture in 2017, making both comparative genomic and
393 proteomic analyses to ORM2a possible (Figure 3 and Table 3). The *Pelotomaculum* candidate BPL
394 MAG (2,986,7007 bp across 97 contigs) does not contain any homologs of "Nanopod" genes
395 (Table 3). However, a highly expressed *Pelotomaculum* contig (Ga0073689_145) produced eight
396 significant alignments to the "Magic" gene cluster (64-76% amino acid identity, e-values ≤ 1.0E-
397 178). Other *Pelotomaculum* proteins encoded on contig Ga0073689_145 included a predicted
398 benzoyl-CoA reductase subunit (BamB) and an enoyl-CoA hydratase, drawing parallels to the
399 gene neighborhood in the genomes of ORM2a, ORM2b, and *Deltaproteobacteria* AUK045
400 (Figure 3). These striking similarities are unlikely to be a coincidence.

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

402 Multiple sequence and structure-based protein prediction tools were used to identify operons in
403 the "Magic" gene cluster and to glean more insights into the molecular and biological functions of

404 each encoded gene. Search results are summarized in Figure 4 and Tables S11-S12, and key
405 findings are described here. We also attempted to construct phylogenetic trees for each protein
406 coded by these gene clusters, but most were too distantly related to reference enzymes (with
407 experimentally confirmed functions) to produce meaningful data.

408 Five operons were predicted in the "Magic" gene cluster, each coding for 2-3 proteins
409 (Figure 4a). Interestingly, genes in "Magic" Operons #1, #2 and #3 were co-localized in every
410 MAG and bin surveyed (Figure 3) and may represent subunits of distinct enzyme complexes. Each
411 protein prediction tool agreed that Protein #22 is most likely a novel AMP-forming acyl-CoA
412 synthetase (Figure 4b). Phylogenetic analyses place Protein #22 (and its homologs from other
413 benzene degraders surveyed) in a subfamily of acyl-CoA synthetases distinct from functionally
414 characterized enzymes including benzoyl-CoA ligase and acetyl-CoA synthetase (Figures 4c and
415 S6). Gene Orthology (GO) term search results using DeepFRI [69] suggest that Protein #22 can
416 bind to organic cyclic compounds (Table S11b), possibly benzene or its activation product. Given
417 that the same cofactors are predicted to participate in Proteins #22 and #42 reactions (e.g.,
418 ATP/AMP, coenzyme A), perhaps these reactions are coupled together and allow one reaction's
419 energy to drive the other (Figure 2).

420 Genes coding for homologs of Proteins #22 and #23 were always co-localized, as were
421 genes coding for homologs of Proteins #27 and #28 (Figure 4a) Although we could not confidently
422 assign a function to Protein #23, DeepFRI search results suggest that the protein may bind to
423 organic cyclic compounds as well as metal ions. Proteins #27 and #28 likely represent subunits of
424 a class III CoA-transferase enzyme complex (Figures 4b and 4f). This class of enzymes catalyzes
425 the transfer of coenzyme A groups from acyl-CoA donors to carboxylic acid acceptors, and are
426 known to be involved in the metabolism diverse substrates (e.g., oxalate, carnitine, toluene, and

427 bile acid) [91]. Despite some shared functional properties to other CoA-transferase classes, they
428 are not closely related (<25% amino acid identity) [92]. Search results from AphaFold2 [66, 67]
429 and DALI [68] determined that Protein #30 is structurally most similar to (*3R*)-hydroxyacyl-ACP
430 (acyl carrier protein) dehydratase subunit HadA (Table 3b), involved in fatty acid beta-oxidation
431 [93]. Sequence analyses in MotifFinder [65] and PROST [73] indicate that Protein #29 could
432 represent a second acyl-ACP or acyl-CoA dehydratase subunit (Tables 3b and S12).

433 Proteins #24, #25 and #26 – which were the three most abundant proteins identified in the
434 OR proteomes (Table 2) – are especially intriguing. Most prediction tools annotated Proteins #24
435 and #26 as flavoproteins with possible L-lactate or 4-cresol oxidoreductase activities (Figure 4b
436 and 4e). DeepFRI predictions suggest that both proteins can bind cyclic organic compounds, and
437 that Protein #24 may also be nucleotide and co-enzyme binding (Table S11b). The sequence and
438 structure of Protein #25 are most similar to heterodisulfide reductase subunits HdrD and HdrA,
439 respectively, implicated in methanogenesis (Figures 4b and 4d). Heterodisulfide reductases (HDR)
440 are a functionally diverse but largely uncharacterized class of enzymes found in both archaea and
441 bacteria [94]. Phylogenetic analyses place this protein in an uncharacterized subfamily of HDRs
442 (Figure S7-S8). We note here that several other HDR coding genes were identified in the ORM2a
443 genome (Figure 2 and Table S5) but their products were not detected in the proteome data. Protein
444 #25 may also belong to the radical S-adenosylmethionine (SAM) superfamily [95] as it contains
445 two characteristic cysteine-rich regions (CxxxCxxC) near its N-terminus (Figure S8). Members of
446 the SAM superfamily use $[4Fe\text{-}4S]^{+}$ clusters to reductively cleave the C5′–sulfonium bond of
447 SAM, triggering radical-based chemistry in hundreds of distinct chemical reactions including
448 methylation and fumarate addition [96, 97]. Perhaps Proteins #24-26 work together to couple
449 benzene activation, a likely highly endergonic reaction, to a unique flavin-based electron

450 bifurcation mechanism [80]. Ferredoxin and a disulfide-containing cofactor could potentially serve
451 as electron acceptors; hydrogen or perhaps benzene could serve as electron donors. ATP may fuel
452 the reaction complex that must also involve a transfer of CoA. We have attempted many times to
453 devise a working model for this reaction but concede that the underlying biochemistry is (for now)
454 far from understood.

455 **Predicted functions of the "Nanopod" gene cluster**

456 Protein prediction tools suggest that the "Nanopod" gene cluster codes for a
457 transmembrane complex (Figure 2, Tables S12-S13). Protein #32 most closely resembles a porin,
458 involved in passive substrate transport, whereas Protein #37 (a predicted RND superfamily efflux
459 protein) appears to be involved in active transport. GO term searches in PROST [73] and DeepFRI
460 [69] suggest that Protein #37 and possibly Protein #34 (an uncharacterized cytoplasmic protein)
461 can bind to organic cyclic structures (Tables S11a and S11b). Most tools agreed that Protein #35
462 is a universal stress factor. Membrane stress can result from partitioning of hydrophobic
463 compounds such as benzene into lipid bilayers [98]. Collectively, it is plausible that "Nanopod"
464 gene cluster codes for a benzene tolerance mechanism that exports excess benzene from ORM2
465 cells (Figure 2).

466 **Phylogenomic placement of the ORM2a and ORM2b_MAG4 genomes**

467 A phylogenomic tree (Figure 5a) placed the ORM2a and ORM2b genomes within an
468 uncharacterized clade of *Desulfobacterota* distinct from other classes previously assigned to these
469 organisms (including *Syntrophales*, and candidate GTDB classes WTBG01 and UBA8473) as well
470 as *Deltaproteobacteria* AUK045. The clade – likely representing a new class of *Desulfobacterota*
471 – contains three other representative MAGs from environmental samples, including one from yet
472 another hydrothermal vent system (GCA_021160085.1) [88]. These MAGs contain some

473 functional potential to degrade benzoyl-CoA and downstream intermediates (Table S11) but they
474 do not appear capable of degrading benzene (Table 3).

475 When compared to a complementary 16S rRNA-based phylogenetic tree (Figure 5b),
476 similar clustering patterns can be clearly seen. It also becomes obvious that ORM2a and ORM2b
477 belong to a monophyletic group of benzene-degrading specialists [4]. These organisms, which
478 were enriched from geographically diverse origins, share ≥99% 16S rRNA gene sequence identity
479 and have only ever been shown to degrade benzene under sulfate-reducing and/or methanogenic
480 conditions [38, 99, 100]. Accounting for new phylogenomic insights presented in this study, there
481 is more than sufficient evidence to recognize these organisms as a novel genus within the
482 *Desulfobacterota*. We propose naming this genus "*Candidatus* Benzenivorax" (ben.ze.ni.i.vor'ax;
483 N.L. neut. n. benzenum benzene, L. adj. vorax voracious or devouring, N.L. neut. n. Benzenivorax,
484 benzene-devouring). We are in the process of registering the proposed taxonomy in SeqCode.

485 **Concluding remarks**

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

496 benzene degradation, and likely in its elusive, energetically difficult activation step. The
497 "Nanopod" gene clusters found in some (ORM2a, ORM2b_MAG4, *Deltaproteobacteria*
498 AUK045) but not all benzene degraders surveyed is more likely involved in benzene tolerance. As
499 most MAGs surveyed were incomplete, homologous "Nanopod" genes could have been missed.
500 It is telling that the only other genomes we could find with matching regions to the "Magic" and
501 "Nanopods" gene cluster were to other strict anaerobes with niche specializations in benzene
502 degradation. Both the "Magic" and "Nanopod" gene clusters appear to be located on genomic
503 islands (Figure 1) suggesting they were acquired through gene transfer. Now that these unique
504 gene clusters have been identified, we hope that the research community can help confirm their
505 functions and to search through other cultures and datasets for homologous features.

506

507 **ACKNOWLEDGEMENTS**

508 This study was made possible by the contributions of many dedicated students and research staff
509 over the past 15 years. Special recognition is given to Dr. Harry Beller (Lawrence Berkeley
510 National Laboratory), Dr. Johann Heider (Philipps University of Marburg), Dr. Christine Orengo
511 (University College London), and Joel Roca (University College London) for their invaluable
512 guidance on characterizing proteins of unknown function.

513 **REFERENCES**

514 1. Ward DM, Atlas RM, Boehm PD *et al.* Microbial biodegradation and chemical evolution
515 of oil from the Amoco spill. *Ambio*. 1980;**9**:277-83
516 2. Vogt C, Kleinsteuber S, Richnow HH. Anaerobic benzene degradation by bacteria. *Microb*
517 *Biotechnol*. 2011;**4**:710-24
518 3. Meckenstock RU, Boll M, Mouttaki H *et al.* Anaerobic degradation of benzene and
519 polycyclic aromatic hydrocarbons. *J Mol Microbiol Biotechnol*. 2016;**26**:92-118
520 4. Toth CRA, Luo F, Bawa N *et al.* Anaerobic benzene biodegradation linked to the growth
521 of highly specific bacterial clades. *Environ Sci Technol*. 2021;**55**:7970-80
522 5. Zhang T, Tremblay PL, Chaurasia AK *et al.* Anaerobic benzene oxidation via phenol in
523 *Geobacter metallireducens*. *Appl Environ Microbiol*. 2013;**79**:7800-06
524 6. Caldwell ME, Suflita JM. Detection of phenol and benzoate as intermediates of anaerobic
525 benzene biodegradation under different terminal electron-accepting conditions. *Environ*
526 *Sci Technol*. 2000;**34**:1216-20
527 7. Vogel TM, Grbić-Galić D. Incorporation of oxygen from water into toluene and benzene
528 during anaerobic fermentative transformation. *Appl Environ Microbiol*. 1986;**52**:200-02
529 8. Kunapuli U, Griebler C, Beller HR *et al.* Identification of intermediates formed during
530 anaerobic benzene degradation by an iron-reducing enrichment culture. *Environ Microbiol*.
531 2008;**10**:1703-12
532 9. Abu Laban N, Selesi D, Jobelius C *et al.* Anaerobic benzene degradation by Gram-positive
533 sulfate-reducing bacteria. *FEMS Microbiol Ecol*. 2009;**68**:300-11
534 10. Ulrich AC, Beller HR, Edwards EA. Metabolites detected during biodegradation of $^{13}C_6$-
535 benzene in nitrate-reducing and methanogenic enrichment cultures. *Environ Sci Technol*.
536 2005;**39**:6681-91
537 11. Weiner JM, Lovley DR. Rapid benzene degradation in methanogenic sediments from a
538 petroleum-contaminated aquifer. *Appl Environ Microbiol*. 1998;**64**:1937-39
539 12. Botton S, Parsons JR. Degradation of BTX by dissimilatory iron-reducing cultures.
540 *Biodegradation*. 2007;**18**:371-81
541 13. Phelps CD, Zhang X, Young LY. Use of stable isotopes to identify benzoate as a metabolite
542 of benzene degradation in a sulphidogenic consortium. *Environ Microbiol*. 2001;**3**:600-03
543 14. Holmes DE, Risso C, Smith JA *et al.* Anaerobic oxidation of benzene by the
544 hyperthermophilic archaeon *Ferroglobus placidus*. *Appl Environ Microbiol*.
545 2011;**77**:5926-33
546 15. Gieg LM, Toth CRA. Signature metabolite analysis to determine in situ anaerobic
547 hydrocarbon biodegradation. In: Boll M (ed.) *Anaerobic Utilization of Hydrocarbons, Oils,*
548 *and Lipids*. Handbook of Hydrocarbon and Lipid Microbiology. 2020. Springer, Cham, pp
549 361-90
550 16. Boll M, Estelmann S, Heider J. Catabolic pathways and enzymes involved in the anaerobic
551 degradation of monocyclic aromatic compounds. In: Boll M (ed.). *Anaerobic Utilization*
552 *of Hydrocarbons, Oils, and Lipids.* Handbook of Hydrocarbon and Lipid Microbiology.
553 2020. Springer, Cham, pp 85-133
554 17. Bullows JE, Kanak A, Shedrick L *et al.* Anaerobic benzene oxidation in *Geotalea daltonii*
555 involves activation by methylation and is regulated by the transition state regulator AbrB.
556 *Environ Microbiol*. 2024;**90**:e00856-24

557 18. Mancini SA, Devine CE, Elsner M *et al.* Isotopic evidence suggests different initial
558 reaction mechanisms for anaerobic benzene biodegradation. *Environ Sci Technol*.
559 2008;**42**:8290-96
560 19. Mancini SA, Ulrich AC, Lacrampe-Couloume G *et al.* Carbon and hydrogen isotopic
561 fractionation during anaerobic biodegradation of benzene. *Appl Environ Microbiol*.
562 2003;**69**:191-98
563 20. Fischer A, Herklotz I, Herrmann S *et al.* Combined carbon and hydrogen isotope
564 fractionation investigations for elucidating benzene biodegradation pathways. *Environ Sci*
565 *Technol*. 2008;**42**:4356-63
566 21. Devine CE. Identification of key organisms, genes and pathways in benzene-degrading
567 methanogenic cultures. 2013. PhD dissertation, University of Toronto
568 22. Taubert M, Vogt C, Wubet T *et al.* Protein-SIP enables time-resolved analysis of the carbon
569 flux in a sulfate-reducing, benzene-degrading microbial consortium. *ISME J*. 2012;**6**:2291-
570 301
571 23. Zhang T, Tremblay PL, Chaurasia AK *et al.* Identification of genes specifically required
572 for the anaerobic metabolism of benzene in *Geobacter metallireducens*. *Front Microbiol*.
573 2014;**5**:245
574 24. Abu Laban N, Selesi D, Rattei T *et al.* Identification of enzymes involved in anaerobic
575 benzene degradation by a strictly anaerobic iron-reducing enrichment culture. *Environ*
576 *Microbiol*. 2010;**12**:2783-96
577 25. Luo F, Gitiafroz R, Devine CE *et al.* Metatranscriptome of an anaerobic benzene-
578 degrading, nitrate-reducing enrichment culture reveals involvement of carboxylation in
579 benzene ring activation. *Appl Environ Microbiol*. 2014;**80**:4095-107
580 26. Atashgahi S, Hornung B, van der Waals MJ *et al.* A benzene-degrading nitrate-reducing
581 microbial consortium displays aerobic and anaerobic benzene degradation pathways. *Sci*
582 *Rep*. 2018;**8**:4490
583 27. Melkonian C, Fillinger L, Atashgahi S *et al.* High biodiversity in a benzene-degrading
584 nitrate-reducing culture is sustained by a few primary consumers. *Commun Biol*.
585 2021;**4**:530
586 28. Xiao JZ, Nesbø CL, Molenda O *et al.* Draft genome of an anaerobic nitrate-reducing,
587 benzene-degrading member of the order *Thermincolales*. *Microbiol Resour Announc*.
588 2024;**13**:e00295-24
589 29. Xiao JZ, Nesbø CL, Molenda O *et al.* Metagenomic and genomic sequences from a nitrate-
590 reducing benzene-degrading enrichment culture. *Microbiol Resour Announc*.
591 2024;**13**:e00294-24
592 30. Hidalgo KJ, Centurion VB, Lemos LN *et al.* Disentangling the microbial genomic traits
593 associated with aromatic hydrocarbon degradation in a jet fuel-contaminated aquifer.
594 *Biodegradation*. 2025;**36**:7
595 31. Roberts GW, Leys D. Structural insights into UbiD reversible decarboxylation. *Curr Opin*
596 *Struct Biol*. 2022;**75**:102432
597 32. Schleinitz KM, Schmeling S, Jehmlich N *et al.* Phenol degradation in the strictly anaerobic
598 iron-reducing bacterium *Geobacter metallireducens* GS-15. *Appl Environ Microbiol*.
599 2009;**75**:3912-9
600 33. Keller AH, Schleinitz KM, Starke R *et al.* Metagenome-based metabolic reconstruction
601 reveals the ecophysiological function of *Epsilonproteobacteria* in a hydrocarbon-
602 contaminated sulfidic aquifer. *Front Microbiol*. 2015;**6**:1396

603 34. Boll M, Estelmann S, Heider J. Anaerobic degradation of hydrocarbons: Mechanisms of
604 hydrocarbon activation in the absence of oxygen. In: Boll M (ed.). *Anaerobic Utilization*
605 *of Hydrocarbons, Handbook of Hydrocarbon and Lipid Microbiology*. 2020. Springer,
606 Cham, pp. 3-29
607 35. Dong X, Dröge J, von Toerne C *et al.* Reconstructing metabolic pathways of a member of
608 the genus *Pelotomaculum* suggesting its potential to oxidize benzene to carbon dioxide
609 with direct reduction of sulfate. *FEMS Microbiol Ecol*. 2017;**93**:fiw254
610 36. Harwood CS, Burchhardt G, Herrmann H *et al.* Anaerobic metabolism of aromatic
611 compounds via the benzoyl-CoA pathway. *FEMS Microbiol Rev*. 1999;**22**:439-58
612 37. Luo F. Characterization of the microbial community composition and benzene activation
613 mechanisms in anaerobic benzene-degrading enrichment cultures. 2012. PhD dissertation,
614 University of Toronto
615 38. Ulrich AC, Edwards EA. Physiological and molecular characterization of anaerobic
616 benzene-degrading mixed cultures. *Environ Microbiol*. 2003;**5**:92-102
617 39. Luo F, Devine CE, Edwards EA. Cultivating microbial dark matter in benzene-degrading
618 methanogenic consortia. *Environ Microbiol*. 2016;**18**:2923-36
619 40. Guo S, Toth CRA, Luo F *et al.* Transient oxygen exposure causes profound and lasting
620 changes to a benzene-degrading methanogenic community. *Environ Sci Technol*.
621 2022;**20**:13036-45
622 41. Toth CRA, Molenda O, Nesbø C *et al.* Metagenomic and genomic sequences from a
623 methanogenic benzene-degrading consortium. *Microbiol Resour Announc*.
624 2023;**22**:e01342-22
625 42. Chen X, Molenda O, Brown CT *et al.* "*Candidatus* Nealsonbacteria" are likely biomass
626 recycling ectosymbionts of methanogenic archaea in a stable benzene-degrading
627 enrichment culture. *Appl Environ Microbiol*. 2023;**89**:e00025-23
628 43. Chen X, Toth CRA, Guo S *et al.* Visualization of syntrophic benzene-fermenting
629 *Desulfobacterota* ORM2 in a methanogenic enrichment culture using fluorescence in situ
630 hybridization. *Environ Sci Technol*. 2025;**59**:591-602
631 44. Nales M, Butler BJ, Edwards EA. Anaerobic benzene biodegradation: A microcosm
632 survey. *Bioremediat J*. 1998;**2**:125-44
633 45. Simpson JT, Wong K, Jackman SD *et al.* Abyss: A parallel assembler for short read
634 sequence data. *Genome Res*. 2009;**19**:1117-23
635 46. Toth CRA, Molenda O, Nesbo C *et al.* FASTA files and relevant statistics for draft MAGs
636 generated from a methanogenic benzene-degrading enrichment culture. *figshare*.
637 2022;**Dataset**
638 47. Chaumeil PA, Mussig AJ, Hugenholtz P *et al.* GTDB-Tk: A toolkit to classify genomes
639 with the genome taxonomy database. *Bioinformatics*. 2019
640 48. Parks DH, Chuvochina M, Rinke C *et al.* GTDB: An ongoing census of bacterial and
641 archaeal diversity through a phylogenetically consistent, rank normalized and complete
642 genome-based taxonomy. *Nucleic Acids Res*. 2021;**50**:D785-D94
643 49. Brettin T, Davis JJ, Disz T *et al. RASTtk*: A modular and extensible implementation of the
644 RAST algorithm for building custom annotation pipelines and annotating batches of
645 genomes. *Sci Rep*. 2015;**5**:8365
646 50. Kanehisa M, Sato Y, Morishima K. BlastKOALA and ghostKOALA: KEGG tools for
647 functional characterization of genome and metagenome sequences. *J Mol Biol*.
648 2016;**428**:726-31

51. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci*. 2020;**29**:28-35
52. Kanehisa M, Sato Y, Kawashima M. KEGG mapping tools for uncovering hidden features in biological data. *Protein Sci*. 2022;**31**:47-53
53. Krzywinski M, Schein J, Birol I *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res*. 2009;**19**:1639-45
54. Dong M-J, Luo H, Gao F. Ori-Finder 2022: A comprehensive web server for prediction and analysis of bacterial replication origins. *Genom Proteom Bioinform*. 2022;**20**:1207-13
55. Bertelli C, Laird MR, Williams KP *et al.* IslandViewer 4: Expanded predicted of genomic islands for larger-scale datasets. *Nucleic Acids Res*. 2017;**45**:W30-W35
56. Arndt D, Grant J, Marcu A *et al.* PHASTER: A bettter, faster version of the PHAST phage search tool. *Nucleic Acids Res*. 2016;**44**:W16-W21
57. Siguier P, Perochon J, Lestrade L *et al.* ISfinder: The reference centre for bacterial insertion sequences. *Nucleic Acids Res*. 2006;**34**:D32-6
58. Taboada B, Estrada K, Ciria R *et al.* Operon-Mapper: A web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*. 2018;**34**:4118-20
59. Wiśniewski JR, Zougman A, Nagaraj N *et al.* Universal sample preparation method for proteome analysis. *Nat Methods*. 2009;**6**:359-62
60. Shevchenko A, Tomas H, Havlis J *et al.* In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc*. 2006;**1**:2856-60
61. Tang S, Chan WWM, Fletcher KE *et al.* Functional characterization of reductive dehalogenases by using blue native polyacrylamide gel electrophoresis. *Appl Environ Microbiol.* 2013;**79**:974-81
62. Kessner D, Chambers M, Burke R *et al.* ProteoWizard: Open source software for rapid proteomics tools development. *Bioinformatics*. 2008;**24**:2534-36
63. Keller A, Nesvizhskii AI, Kolker E *et al.* Empirical statistical model to estimate the accuracy of peptide identifiations made by MS/MS and database search. *Anal Chem*. 2002;**74**:5383-92
64. Nesvizhskii AI, Keller A, Kolker E *et al.* A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem*. 2003;**75**:4646-58
65. Leung WHP, Chi Tam W, Chang BCH *et al.* Effects of search pattern variations in motif discovery algorithm: Motiffinder. *IFAC Proc Vol*. 2003;**36**:501-06
66. Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;**596**:583-89
67. Mirdita M, Schütze K, Moriwaki Y *et al.* ColabFold: Making protein folding accessible to all. *Nat Methods*. 2022;**19**:679-82
68. Holm L. Dali server: Structural unification of protein families. *Nucleic Acids Res*. 2022;**50**:W210-W15
69. Gligorijević V, Renfrew PD, Kosciolek T *et al.* Structure-based protein function prediction using graph convolutional networks. *Nat Commun*. 2021;**12**:3168
70. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;**30**:1312-13
71. 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

694 72. Waman VP, Bordin N, Lau A *et al.* Cath v4.4: Major expansion of CATH by experimental
695 and predicted structural data. *Nucleic Acids Res*. 2024:gkae1087
696 73. Kilinc M, Jia K, Jernigan RL. Improved global protein homolog detection with major gains
697 in function identification. *PNAS*. 2023;**120**:e2211823120
698 74. Lee MD. GtoTree: A user-friendly workflow for phylogenomics. *Bioinformatics*.
699 2019;**35**:4162-64
700 75. Edgar RC. Muscle: Multiple sequence alignment with high accuracy and high throughput.
701 *Nucleic Acids Res*. 2004;**32**:1792-7
702 76. Nguyen L-T, Schmidt HA, von Haeseler A *et al.* IQ-tree: A fast and effective stochastic
703 algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2014;**3**:268-74
704 77. Morris BE, Henneberger R, Huber H *et al.* Microbial syntrophy: Interaction for the
705 common good. *FEMS Microbiol Rev*. 2013;**37**:384-406
706 78. Wischgoll S, Heintz D, Peters F *et al.* Gene clusters involved in anaerobic benzoate
707 degradation of *Geobacter metallireducens*. *Mol Microbiol*. 2005;**58**:1238-52
708 79. Boll M, Einsle O, Ermler U *et al.* Structure and function of the unusual tungsten enzymes
709 acetylene hydratase and class II benzoyl-coenzyme A reductase. *J Mol Microbiol*
710 *Biotechnol*. 2016;**26**:119-37
711 80. Buckel W, Thauer RK. Flavin-based electron bifurcation, ferredoxin, flavodoxin, and
712 anaerobic respiration with protons (Ech) or NAD^+ (Rnf) as electron acceptors: A historical
713 review. *Front Microbiol*. 2018;**9**:401
714 81. Demmer JK, Chowdhury NP, Selmer T *et al.* The semiquinone swing in the bifurcating
715 electron transferring flavoprotein/butyryl-
716 82. Ma K, Weiss R, Adams MWW. Characterization of hydrogenase from the
717 hyperthermophilic archaeon *yrococcus furiosus* and assessment of its role in sulfur
718 reduction. *J Bacteriol*. 2000;**182**:1864-71
719 83. Berg JS, Schwedt A, Kreutzmann A *et al.* Polysulfides as intermediations in the oxidation
720 of sulfide to sulfate by *eggiatoa* spp. *Appl Environ Microbiol*. 2014;**80**:e02852-13
721 84. Hao L, Michaelsen TY, Singleton CM *et al.* Novel syntrophic bacteria in full-scale
722 anaerobic digesters revealed by genome-centric metatranscriptomics. *ISME J*.
723 2020;**14**:906-18
724 85. Shetty A, Chen S, Tocheva EI *et al.* Nanopods: A new bacterial structure ans mechanism
725 for deployment of outer membrane vesicles. *PLoS ONE*. 2011;**6**:e20725
726 86. Shetty A, Hickey WJ. Effects of outer membrane vesicle formation, surface-layer
727 production and nanopod development on the metabolism of phenanthrene by *Delftia*
728 *acidovorans* Cs1-4. *PLoS One*. 2014;**9**:e92143
729 87. Baumgarten T, Sperling S, Seifert J *et al.* Membrane vesicle formation as a multiple-stress
730 response mechanism enhances *Pseudomonas putida* DOT-T1E cell surface hydrophobicity
731 and biofilm formation. *Appl Environ Microbiol*. 201;**78**:6217-24
732 88. Speth DR, Yu FB, Connon SA *et al.* Microbial communities of Auka hydothermal
733 sediments shed light on vent biogeography and the evolutionary history of thermophily.
734 *ISME J*. 2022;**16**:1750-64
735 89. Dombrowski N, Teske AP, Baker BJ. Expansive microbial metabolic versatility and
736 biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat Commun*.
737 2018;**9**:4999

90. Campbell KA. Hydrocarbon seep and hydrothermal vent paleoenvironments and paleontology: Past developments and future research directions. *Palaeogeogr Palaeoclimatol Palaeoecol*. 2006;**232**:362-407
91. Heider J. A new family of CoA-transferases. *FEBS Lett*. 2001;**509**:345-49
92. Hackmann TJ. Redefining the coenzyme A transferase superfamily with a large set of manually annotated proteins. *Protein Sci*. 2022;**7**:864-81
93. Sacco E, Suarez Covarrubias A, O'Hare HM *et al.* The missing piece of the type II fatty acid synthase system from *Mycobacterium tuberculosis*. *PNAS*. 2007;**104**:14628-33
94. Yan Z, Wang M, Ferry JG. A ferredoxin and $F_{420}H_2$-dependent, electron-bifurcating, heterodisulfide reductase with homologs in the domains *Bacteria* and *Archaea*. *mBio*. 2017;**8**:e02285-16
95. Booker SJ, Lloyd CT. Twenty years of radical SAM! The genesis of the superfamily. *ACS Bio & Med Chem Au*. 2022;**2**:538-655
96. Nicolet Y. Structure-function relationships of radical SAM enzymes. *Nat Catal*. 2020;**3**:337-50
97. Broderick JB, Duffus BR, Duschene KS *et al.* Radical *S*-adenosylmethionine enzymes. *Chem Rev*. 2014;**114**:4229-317
98. Ramos JL, Duque E, Gallegos MT *et al.* Mechanisms of solvent tolerance in Gram-negative bacteria. *Ann Rev Microbiol*. 2002;**56**:473-768
99. 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
100. 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

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 proteins (≥100 matching spectra in Experiment #3b, highlighted in bold) and proteins of interest recovered with high certainty (≥99% confidence). Proteins shown mapped to one or more MAG reconstructed from OR metagenomes. Proteins are grouped by predicted biochemical functions and ranked in abundance based on total spectra counts; proteins tied in rank abundance are denoted with (=) symbols. Other proteins identified are listed in Table S9.

Protein #	Predicted function (NCBI Prokaryotic Genome Annotation Pipeline version 6.3 and best BLASTP hits)	Protein spectra count (Exp. #3b)	Protein abundance rank	Matching MAG(s)	Protein accession #
Benzoyl-CoA degradation to 3-hydroxypimelyl-CoA					
1	Benzoyl-CoA reductase subunit BamB	80	=26	ORM2a and ORM2b	WAC08761.1, ORM2b_MAG4_01005
2	Benzoyl-CoA reductase subunit BamC	15	=154	ORM2a	WAC08762.1, ORM2b_MAG4_01004
3	Benzoyl-CoA reductase subunit BamD	61	=43	ORM2a	WAC08763.1
4	Formate dehydrogenase accessory sulfurtransferase FdhD	6	=287	ORM2a	WAC09094.1
5	Benzoyl-CoA reductase subunit BamE	68	32	ORM2a	WAC09095.1
6	Benzoyl-CoA reductase subunit BamF	18	=135	ORM2a	WAC08764.1
7	Benzoyl-CoA reductase subunit BamG	6	=287	ORM2a	WAC08765.1
8	Benzoyl-CoA reductase subunit BamH	91	23	ORM2a and ORM2b	WAC08766.1, ORM2b_MAG4_04225
9	Benzoyl-CoA reductase subunit BamI	9	=227	ORM2a	WAC08767.1
10	Cyclohexa-1,5-diene-1-carbonyl-CoA hydratase BamR	49	59	ORM2a	WAC08768.1
11	**6-hydroxycyclohex-1-ene-1-carbonyl-CoA dehydrogenase BamQ**	**103**	**20**	**ORM2a**	**WAC08375.1**
12	**6-oxo-cyclohex-1-ene-carbonyl-CoA hydrolase BamA**	**143**	**13**	**ORM2a**	**WAC07399.1**
3-hydroxypimelyl-CoA degradation to acetyl-CoA					
13	**3-hydroxyacyl-CoA dehydrogenase**	**148**	**9**	**ORM2a**	**WAC08792.1**
14	Thiolase family protein	69	31	ORM2a	WAC08791.1
15	Glutaryl-CoA dehydrogenase (non-decarboxylating)	58	=49	ORM2a	WAC06987.1
16	Electron transfer flavoprotein subunit beta/FixA family protein	39	=69	ORM2a	WAC06988.1
17	Electron transfer flavoprotein subunit alpha/FixB family protein	63	=39	ORM2a	WAC06989.1
18	Glutaconyl-CoA decarboxylase subunit alpha	5	=332	ORM2a	WAC08228.1
19	Enoyl-CoA hydratase-related protein	3	=458	ORM2a	WAC07318.1
20	3-hydroxybutyryl-CoA dehydrogenase	9	=227	ORM2a	WAC09154.1
21	Acetyl-CoA C-acetyltransferase	3	=458	ORM2a	WAC07319.1
Highly expressed "Magic" gene cluster putatively involved in benzene activation					
22	AMP-binding protein	74	29	ORM2a and ORM2b	WAC08366.1, ORM2b_MAG4_03434
23	MaoC family dehydratase	57	52	ORM2a	WAC08367.1
24	**FAD-binding oxidoreductase**	**172**	**3**	**ORM2a**	**WAC08368.1**
25	**(Fe-S)-binding protein**	**219**	**2**	**ORM2a**	**WAC08369.1**
26	**FAD-binding oxidoreductase**	**265**	**1**	**ORM2a**	**WAC08370.1**
27	Type III CoA transferase	45	61	ORM2a	WAC08371.1
28	Type III CoA transferase	58	=49	ORM2a and ORM2b	WAC08372.1, ORM2b_MAG4_03441
29	Hypothetical protein	19	=128	ORM2a and ORM2b	WAC08373.1, ORM2b_MAG4_03442
30	MaoC family dehydratase	50	=57	ORM2a	WAC08374.1
31	Transcription termination factor Rho	25	=100	ORM2a	WAC08378.1
Highly expressed "Nanopod" gene cluster putatively involved in benzene tolerance					
32	**Hypothetical protein**	**106**	**=18**	**ORM2a**	**WAC07726.1**
33	**DUF1329 domain-containing protein**	**158**	**4**	**ORM2a**	**WAC07727.1, ORM2b_MAG4_01802**
34	Universal stress protein	6	=287	ORM2a	WAC07728.1
35	Hypothetical protein	7	=269	ORM2a	WAC07729.1
36	YCF48-related protein	65	=36	ORM2a	WAC07730.1
37	MMPL family transporter	60	=45	ORM2a	WAC07731.1
Archaeal proteins associated with methanogenesis					
38	**Acetate-CoA synthetase Acs**	**142**	**14**	***Methanothrix* MAG20**	**MAG20_01236**
39	**Coenzyme-B sulfoethylthiotransferase McrB**	**149**	**8**	***Methanothrix* MAG21**	**MAG21_01157**
40	**Coenzyme-B sulfoethylthiotransferase McrB**	**154**	**7**	***Methanothrix* MAG11, MAG20 and MAG28**	**MAG11_02361, MAG20_02648, MAG28_00841**
41	**Acetate-CoA synthetase Acs**	**125**	**16**	***Methanothrix***	**MAG11_00500**
Other proteins					
42	**Acetate-CoA synthetase Acs**	**156**	**=5**	**ORM2a**	**WAC06830.1**
43	**Chaperonin GroL**	**156**	**=5**	**ORM2a**	**WAC08671.1**

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. Protein accession numbers and best BLASTP hit search results for other ORM2a proteins of interest and other reference genomes are provided in Table S11.

	Protein accession no. of best BLASTP hit								
Protein cluster #	**ORM2a**	**ORM2b_MAG4**	***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**
	This study	This study	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	**100%**	**96%**	**76%**	**72%**	**71%**	**71%**	25%	24%	27%
24	**100%**	**95%**	**76%**	**73%**		**73%**			
25	**100%**	**97%**	**74%**	**76%**				22%	
26	**100%**	**97%**	**70%**	**69%**	**66%**	28%	26%	29%	29%
27	**100%**	**94%**	**66%**	**62%**	**61%**			24%	
28	**100%**	**98%**	**74%**	**65%**	**72%**	**69%**			28%
29	**100%**	**98%**	**69%**	**60%**	**62%**	**62%**			29%
30	**100%**	**90%**		**61%**	51%				
31	**100%**	**90%**		**60%**	52%				
11	**100%**	**94%**	**64%**	27%		57%	21%	**62%**	
Highly expressed "Nanopod" gene cluster putatively involved in benzene tolerance									
32	**100%**	**84%**		44%		26%	30%	23%	29%
33	**100%**	**94%**		**60%**	32%	36%	31%	23%	33%
34	**100%**	**82%**	32%	43%	31%	29%		28%	31%
35	**100%**	**92%**							
36	**100%**	**88%**		46%			31%	30%	30%
37	**100%**	**92%**		51%	33%		35%	36%	34%

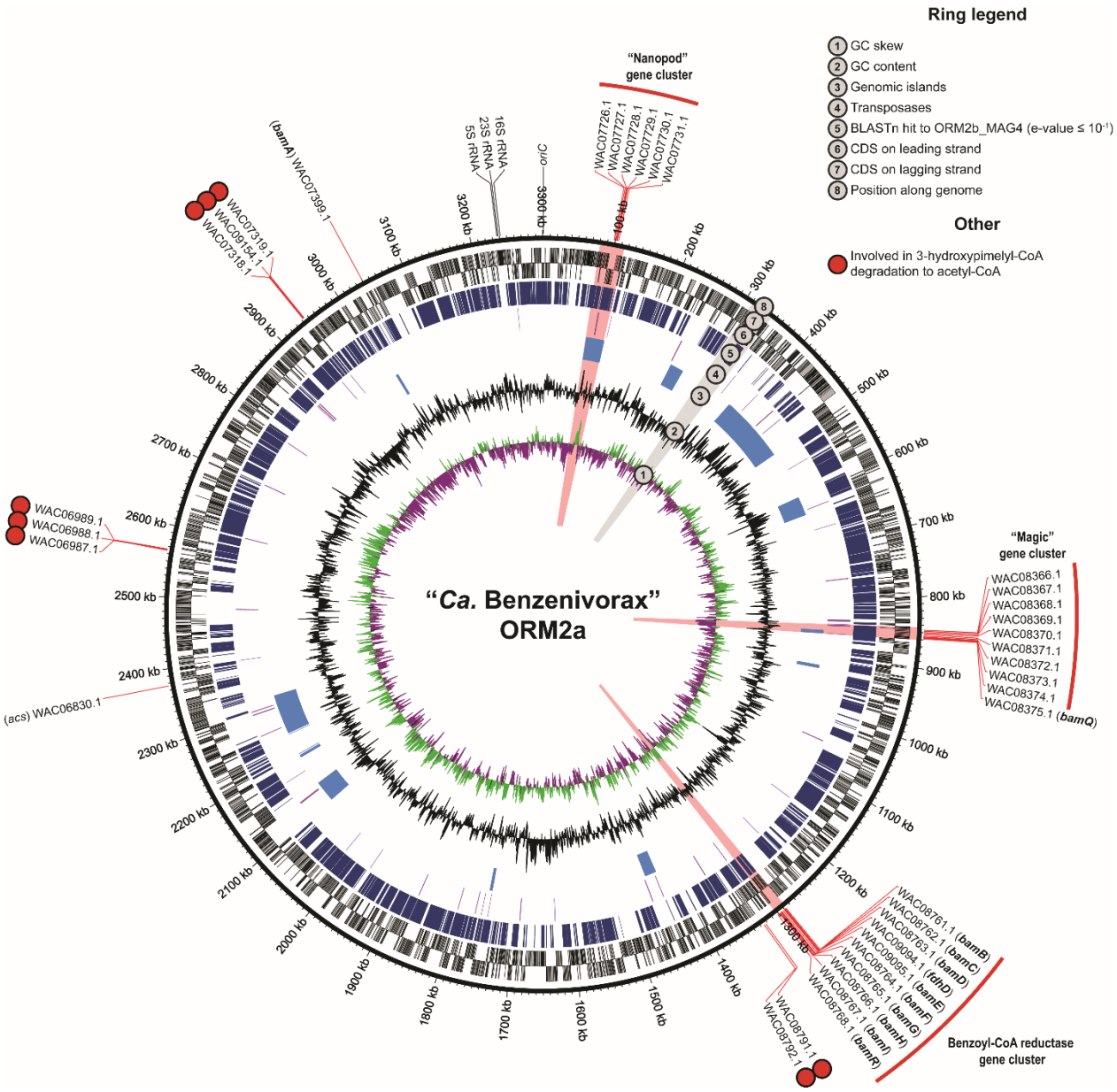
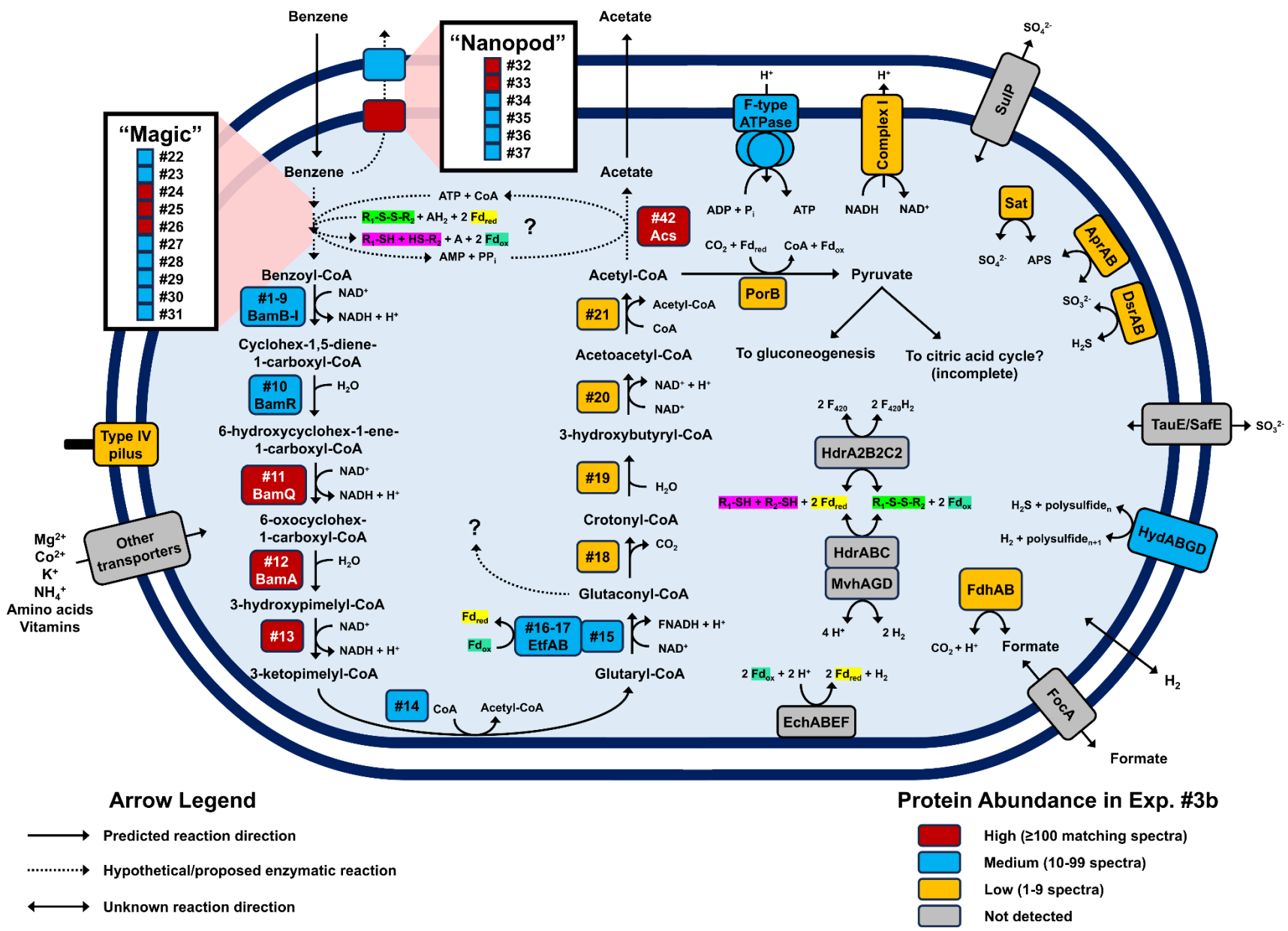


Figure 1. The complete ORM2a genome. Genomic neighbourhoods and genes of interest whose products were recovered with high certainty (≥99% confidence) from the OR proteomes are highlighted in red and are overviewed in Table 2. Genes whose products are involved in the degradation of 3-hydroxypimelyl-CoA to acetyl-CoA are marked with red circles. Protein accession numbers are shown; corresponding ORM2a gene locus tags are provided in Table S9. The positions of the origin of replication *(oriC)* and ribosomal RNA genes are also shown in this figure.



Arrow Legend

- Predicted reaction direction
- Hypothetical/proposed enzymatic reaction
- Unknown reaction direction

Protein Abundance in Exp. #3b

- High (≥100 matching spectra)
- Medium (10-99 spectra)
- Low (1-9 spectra)
- Not detected

Figure 2. Summary of key metabolic pathways identified in the ORM2a proteome. Protein spectral counts from Experiment #3b were used to estimate protein abundances. Other complete and incomplete metabolic pathways identified in the ORM2a genome are summarized in Table S3. Annotations for numbered proteins are provided in Table 2. Abbreviations (clockwise from top-left); R-S-S-R, heterodisulfide; ACS, acetyl-CoA synthetase; Fd, ferredoxin; PorB, pyruvate synthase; SulP, sulfate permease; Sat, sulfate adenylyltransferase; AprAB, adenylyl-sulfate reductase; dissimilatory sulfite reductase; TauE/SafE, sulfite exporter; HydABGD, sulfhydrogenase; FdhAB, formate dehydrogenase; FocA, formate-nitrite transporter; HdrA2B2C2, F_{420}-reducing heterodisulfide reductase; HdrABC, heterodisulfide reductase; Mvh, F_{420}-non-reducing hydrogenase; EchABEF, energy converting hydrogenase; EtfAB, electron-transferring flavoprotein oxidoreductase.

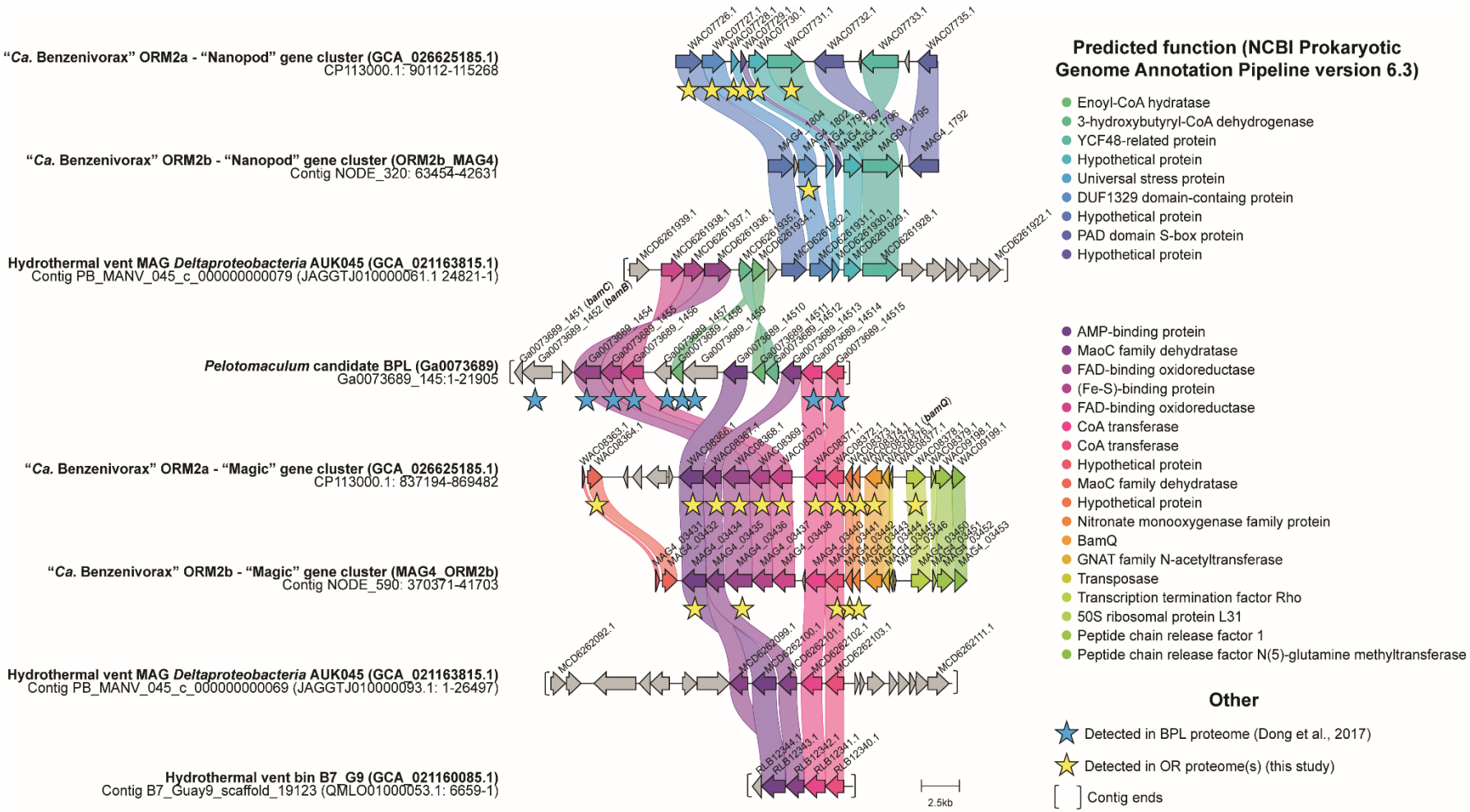
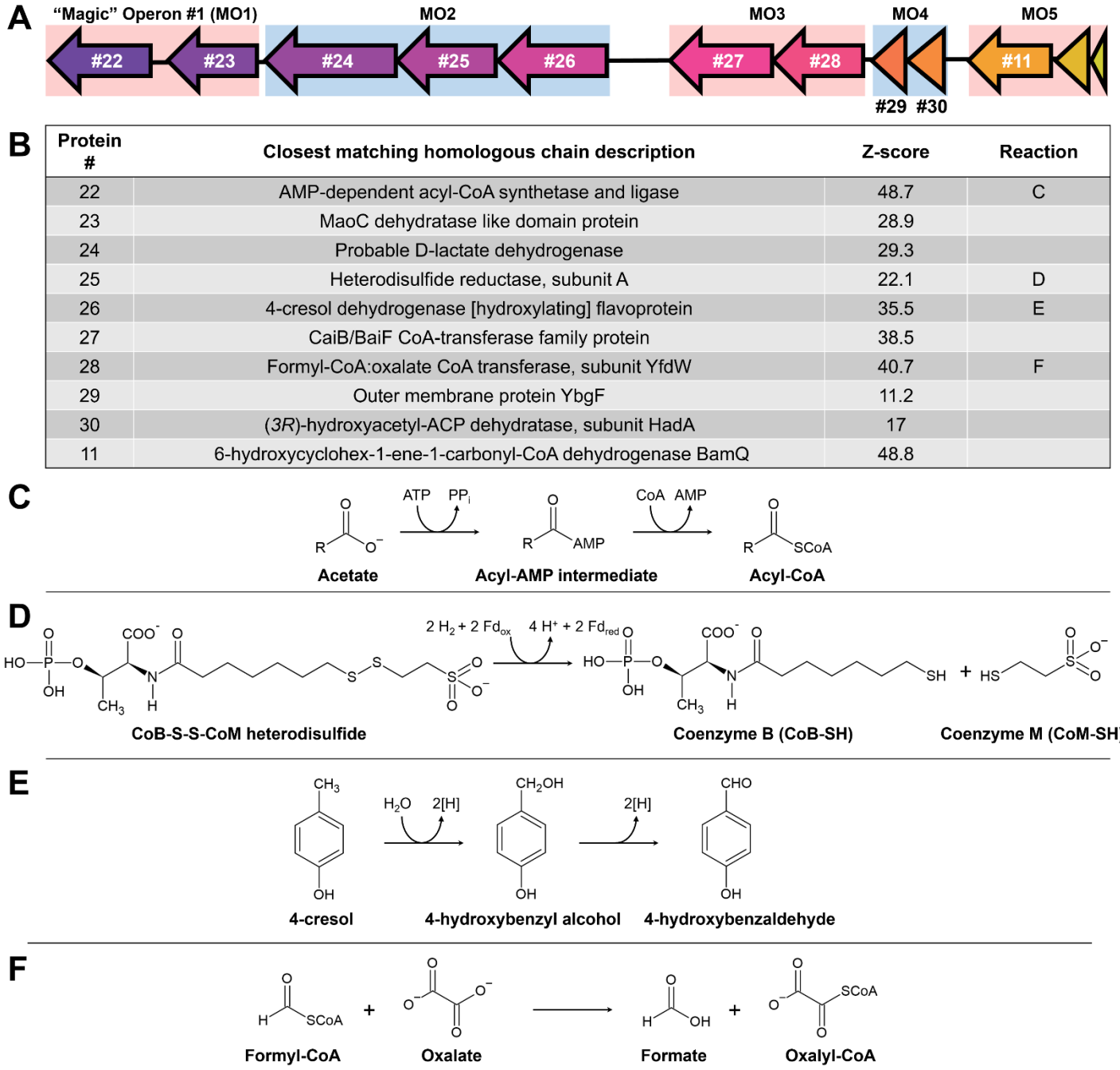


Figure 3. Genomic neighborhoods of all publicly available MAGs in NCBI and IMG producing significant alignments to the "Magic" and "Nanopod" gene clusters in ORM2a and ORM2b_MAG4. Homologous genes are color-coded by function; sequence alignment statistics for relevant proteins are shown in Table 3; additional alignment statistics are provided in Table S11. 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.



Protein #	Closest matching homologous chain description	Z-score	Reaction
22	AMP-dependent acyl-CoA synthetase and ligase	48.7	C
23	MaoC dehydratase like domain protein	28.9	
24	Probable D-lactate dehydrogenase	29.3	
25	Heterodisulfide reductase, subunit A	22.1	D
26	4-cresol dehydrogenase [hydroxylating] flavoprotein	35.5	E
27	CaiB/BaiF CoA-transferase family protein	38.5	
28	Formyl-CoA:oxalate CoA transferase, subunit YfdW	40.7	F
29	Outer membrane protein YbgF	11.2	
30	(*3R*)-hydroxyacetyl-ACP dehydratase, subunit HadA	17	
11	6-hydroxycyclohex-1-ene-1-carbonyl-CoA dehydrogenase BamQ	48.8	

Figure 4. Structure and functional prediction of proteins coded by the "Magic" gene cluster. Panel A shows the operon organization of this gene cluster. Panel B summarizes the best Protein Data Bank (PDB) search results for each "Magic" gene product. Z-scores are a measure of protein composite confidence, where more closely matching structures produce higher scores; Z-scores ≥30 are considered significant. More alignment statistics are provided in Table S12. Panels C through F show sample reaction pathways catalyzed by reference proteins.

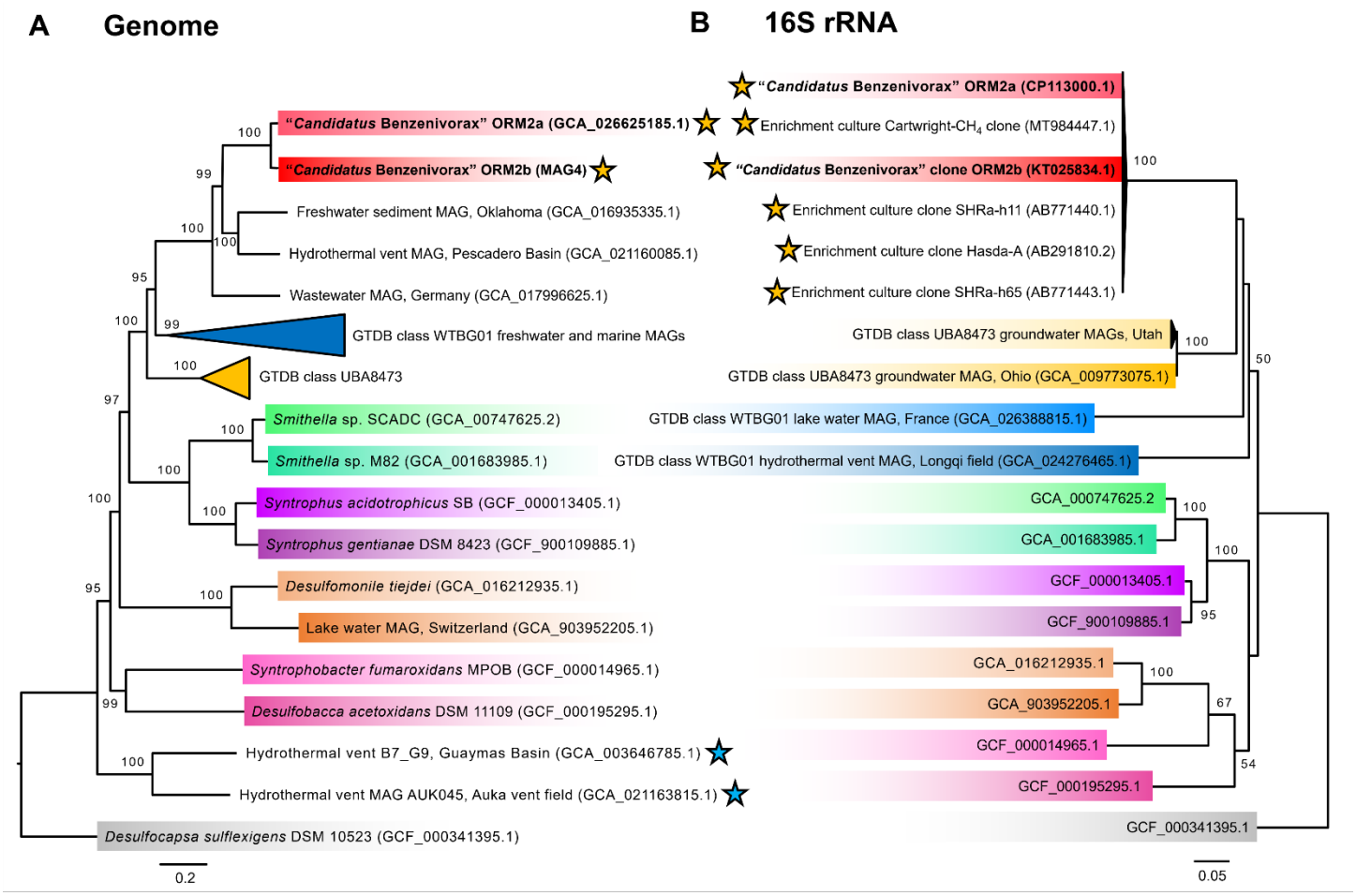


Figure 5. Taxonomic classification of ORM2a and reference *Desulfobacterota* species based on A) genome and B) 16S rRNA gene maximum likelihood phylogenetic trees. Sequences from this study are highlighted in bold. Sequences belonging to known and predicted benzene-degrading anaerobes are marked with 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.