1 **TITLE PAGE**

2 ***Title:*** Krumholzibacteriota and Deltaproteobacteria contain rare genetic potential to liberate

3 carbon from monoaromatic compounds in subsurface coal seams

4 ***Running title:*** Genetic potential for carbon liberation in coal seams

5 ***Authors:*** Bronwyn C. Campbell[a,b]#, Paul Greenfield[b,c], Elliott P. Barnhart[d], Se Gong[c], David J.

6 Midgley[c], Ian T. Paulsen[b] and Simon C. George[b]

7 [a]Environment Business Unit, Commonwealth Scientific and Industrial Research Organisation

8 (CSIRO), Floreat, WA, 6014, Australia

9 [b]School of Natural Sciences, Macquarie University, North Ryde, NSW, 2109, Australia

10 [c]Energy Business Unit, Commonwealth Scientific and Industrial Research Organisation (CSIRO),

11 Lindfield, NSW, 2070, Australia

12 [d]U.S. Geological Survey, Wyoming-Montana Water Science Center, Helena, MT, 59601, USA

13 ***#Corresponding author:*** Bronwyn.Campbell@csiro.au

14 ***Competing interests statement:*** The authors declare no competing interests

ABSTRACT

Biogenic methane in subsurface coal seam environments is produced by diverse consortia of microbes. Although this methane is useful for global energy security, it remains unclear which microbes can liberate carbon from the coal. Most of this carbon is relatively resistant to biodegradation, as it is contained within aromatic rings. Thus, to explore for coal-degrading taxa in the subsurface, this study used coal seam metagenomes to reconstruct important metagenome-assembled genomes (MAGs) using a key genomic marker for the anaerobic degradation of monoaromatic compounds as a guide: the benzoyl-CoA reductase gene (*bcrABCD*). Three taxa were identified with this genetic potential. The first was a novel taxon from the Krumholzibacteriota phylum, which this study is the first to describe. This Krumholzibacteriota sp. contained a full set of genes for benzoyl-CoA dearomatisation, in addition to other genes for anaerobic catabolism of monoaromatics. Analysis of Krumholzibacteriota MAGs from other environments revealed that this genetic potential may be common within this phylum, and thus they may be important organisms for the liberation of recalcitrant carbon on a global scale. Further, two taxa from the Deltaproteobacteria class were also implicated in monoaromatic degradation; two geographically unrelated *Syntrophorhabdus aromaticivorans* MAGs, and a Syntrophaceae sp. MAG. Each of these three taxa are potential rate-limiting organisms for subsurface coal-to-methane biodegradation. Their description here provides an understanding of their function within the coal seam microbiome, and will help inform future efforts in coal bed methane stimulation, anoxic bioremediation of organic pollutants, and assessments of anoxic carbon cycling and emissions.

38 **Importance**

39 Subsurface coal seams are highly anoxic and oligotrophic environments, where the main source of
40 carbon is “locked away” within aromatic rings. Despite these challenges, biogenic methane
41 accumulates within many of these coal seams, which implies that the coal seam microbiome can
42 “unlock” this carbon source *in situ*. For over two decades, researchers have been working to
43 understand which organisms are responsible for these processes. This study provides the first
44 descriptions of these organisms. Here, we report metagenomic insights into the liberation of
45 carbon from aromatic molecules typically found within coal, the degradation pathways involved,
46 and descriptions of the Krumholzibacteriota sp., *Syntrophorhabdus aromaticivorans*, and
47 Syntrophaceae sp. that contain this genetic potential. Additionally, this is the first time that the
48 Krumholzibacteriota phylum has been implicated in anaerobic dearomatisation of aromatic
49 hydrocarbons. This potential is identified here in numerous taxa within the phylum from other
50 subsurface environments, implicating Krumholzibacteriota in global-scale carbon-cycling
51 processes.

52 **INTRODUCTION**

53

54 The global energy transition from hydrocarbons to renewables requires low-emissions fuels for
55 facilitating energy security during this shift. For this transition fuel need, methane is one such
56 lower-emission alternative to coal (1). Methane gas provides a dispatchable source of energy that,
57 unlike coal, produces neither particulates nor harmful nitrous and sulfur oxides during combustion
58 (2, 3). Thus, significant research interest exists in enhancing rates of methane gas production in
59 subsurface coal by using the coal seam microbiome (4–6).

60

61 Within the coal itself, carbon is primarily (~60 to 100 %) contained within aromatic rings, which
62 increase in abundance with thermal maturity (7). Thus, taxa capable of aromatic ring degradation
63 may be the dominant contributors of carbon to the coal seam microbiome, especially in more
64 thermally mature coals. Over the last decade, a good understanding has been formed of the range
65 of microbes that occur in subsurface coal seams, which are typically characterised by a few
66 relatively abundant taxa and a modest tail of rarer taxa (4, 8, 9). Coal seam microbiome studies
67 have also revealed that the dominant organisms, such as taxa within *Desulfuromonas* and
68 *Desulfovibrio*, are not involved in aromatic degradation but rather degrade simpler intermediates
69 (8, 10). Aromatic-degrading taxa therefore probably occur within the rarer taxa.

70

71 Identification of aromatic-degrading taxa is important both for enhancing applied outcomes such
72 as industrial gas production, and for understanding carbon mobilisation from the fossilised
73 geosphere. Indeed, the research effort to identify these taxa has been ongoing since the early
74 papers in the subsurface coal microbiology field (11). Some taxa have previously been identified
75 with the genetic tools for aerobic aromatic ring degradation within coal seams (12–14), however,
76 aerobic pathway contributions are likely restricted to very shallow regions of meteoric water

77 infiltration, since subsurface coal seams are overall highly anoxic environments (14). Within other
78 anoxic hydrocarbon-degrading environments, such as oil reservoirs (15), the catabolism of
79 aromatic substrates progresses at least in part via intermediate aromatic compounds. Although a
80 wide range of monoaromatic biodegradation pathways are known, these pathways are dependent
81 upon a relatively small range of central intermediate aromatic compounds. One intermediate
82 aromatic compound in particular, known as benzoyl-CoA, is central to the anaerobic degradation
83 of a particularly wide range of monoaromatics (16). Further catabolism from this central
84 intermediate requires the benzoyl-CoA reductase enzyme, which is responsible for the first
85 dearomatisation steps of the ring (17). This enzyme for ring cleavage is a crucial step for accessing
86 the carbon contained within the ring structures, as the thermodynamic stability of the aromatic
87 ring structure renders it highly resistant to degradation (15). The ability to encode this enzyme and
88 those for the proceeding metabolites in the benzoyl-CoA pathway (depicted in Figure 1) would
89 provide organisms with access to much of the carbon locked up in the organic substrates present
90 in coal.
91
92 To date, numerous studies have attempted to identify aromatic-degrading taxa in coal seams using
93 a range of strategies. Some of these have included the enrichment of microbes capable of
94 degrading aromatic compounds, using either a variety of mono- and polyaromatic compounds or
95 organic matter of differing maturities as sole sources of carbon (18, 19). These studies identified
96 putative aromatic-degraders among the enriched taxa, but also obscure primary aromatic-
97 degraders amid the numerous taxa subsequently enriched on their downstream degradative
98 products.
99
100 Metagenome-assembled genomes (MAGs) have also been used to identify aromatic-degrading
101 taxa by using known genes required for hydrocarbon degradation, performed alongside methods

102 such as bio-orthogonal non-canonical amino acid tagging (BONCAT) to differentiate active from
103 inactive cells (12, 13). Within the active taxa, this strategy has identified a range of genes used for
104 rearranging the substituents of the aromatic ring in peripheral pathways above the benzoyl-CoA
105 intermediate, for example genes for catabolism of ethylbenzene to acetophenone in Chlorobi and
106 *Geobacter* taxa (12). Although these studies identify an abundance of genes and taxa actively
107 involved in the coal-to-methane degradation pathways, the specific anaerobic monoaromatic-
108 degrading genes identified are not involved in catalysing dearomatisation reactions or capable of
109 liberating carbon from their targeted substrates (15, 20, 21). Consequently, the identification of
110 taxa containing the genes for dearomatisation of these monoaromatic substrates, such as the
111 benzoyl-CoA reductase gene, remains an unanswered and critical step for understanding the *in*
112 *situ* liberation of carbon from coal.
113
114 Recently, Syntrophorhabdaceae and Syntrophaceae were implicated as potentially important coal-
115 degrading families using the linear discriminant analysis effect size statistical method on a group of
116 algal-amended coal seam microbiomes from the Powder River Basin, USA (22).
117 Syntrophorhabdaceae is a monotypic family, of which *Syntrophorhabdus aromaticivorans* is
118 presently the sole described species (23). *S. aromaticivorans* has previously been identified within
119 an enriched Surat Basin (Australia) coal seam microbial community, which had likely responded to
120 the increased surface area of the provided coal that had undergone solvent extraction (24).
121 Outside the coal seam environment, a *S. aromaticivorans* isolated from an anaerobic digester was
122 demonstrated to anaerobically catabolise multiple monoaromatic substrates to acetate, to utilise
123 a model organic electron acceptor, and was proposed to be capable of interspecies electron
124 transfer in partnership with a hydrogenotrophic methanogen (23). *S. aromaticivorans* is thus a
125 promising candidate for monoaromatic degradation capabilities within the coal seam
126 environment. In contrast to this, Syntrophaceae spp. are more commonly associated with aliphatic

127 degradation than aromatic degradation, however, some taxa within the family can utilise a limited
128 range of more labile monoaromatic substrates (25, 26). Although Syntrophaceae spp. and *S.*
129 *aromaticivorans* have been implicated as taxa that may be important for *in situ* coal degradation,
130 there has been no demonstration that strains from this environment have the genetic tools
131 required to access aromatic compounds or culture-based demonstrations of their activity against
132 aromatic constituents of coal.
133
134 One alternative strategy to those outlined above is to 'mine' metagenomic data for specific
135 aromatic-degrading genes, and then use binning techniques to reassemble genomes of potential
136 aromatic-degraders. Accordingly, this study aimed to identify aromatic-degrading taxa by mining
137 metagenomic data from Australia and North America through first identifying assembled DNA
138 sequences that contained benzoyl-CoA reductase gene subunits, and then using these as
139 references to assemble metagenome-assembled genomes (MAGs) for the identification of these
140 taxa and their capabilities.
141
142 **MATERIALS AND METHODS**
143
144 **Sources of metagenomic DNA**
145 For further details of metagenomic sequence data sourcing, see Campbell et al. 2022 (27). Briefly,
146 whole-genome shotgun sequences selected for use were required to: (1) be from a subsurface
147 coal seam, (2) state the coal seam or associated geological basin, (3) have been sequenced using
148 Illumina, and (4) be available as unassembled sequence reads. Four metagenomes from Australia
149 and nine from the USA were found to be suitable (Supplementary Table S1).
150
151 **Reassembly, annotation and detection of benzoyl-CoA reductase gene subunits**

152 All metagenomes were downloaded as unassembled reads. These reads were error-corrected
153 using Blue v2 (28) prior to assembly using SPAdes v3.13.2 with the meta flag (29). Resultant
154 contigs were annotated using prokkaMeta v1.14.5 (30), and these annotations were then used to
155 identify the various subunits of benzoyl-CoA reductase within the contigs.
156
157 **Selection of contigs with benzoyl-CoA reductase gene subunits**
158 Contigs with benzoyl-CoA reductase gene subunits were interrogated using the Prokka
159 annotations for other genes in the benzoyl-CoA pathway. Contigs selected for further analysis
160 contained at least one benzoyl-CoA reductase gene subunit (*bcrABCD*), were of lengths greater
161 than seven kilo–base-pairs, and had coverages greater than five.
162
163 **Generation of trimer signatures, correlations, bin quality control, and accessioning**
164 In order to identify other contigs within the metagenome from the same taxon, a trimer approach
165 was used (31). In brief, a custom Python script was used to count the proportion of each of the 64
166 possible trimers in all contigs containing at least 1000 base pairs. The trimer signature for those
167 contigs containing benzoyl-CoA reductase gene subunits was then used as a reference, and the
168 trimer signatures for all contigs in the metagenome were examined against these reference
169 contigs using a Pearson correlation coefficient in SciPy 1.7.3 (scipy.org). Those contigs with
170 Pearson's R values greater than 0.95 were collected into bins for quality control. As a subsequent
171 step in bin quality control, the coverages of each contig within the bin were inspected manually
172 and contigs with aberrant coverage were removed. Each bin was then subject to inspection for
173 completeness and contamination using CheckM v1.1.3 (32).
174
175 **Characterisation of the genomic content of the bins**

176 Summary contig statistics (total bin length, total contigs, mean contig length, median contig
177 length, N50, maximum contig length, and GC-content) were determined using a custom Python
178 script. The bins were then submitted to online tools to further characterise the genomic content of
179 each bin. For use with KEGG Mapper (genome.jp/kegg/mapper), the Prokka-annotated amino acid
180 files were submitted to the BlastKOALA v2.2 online tool and run against the "species_prokaryotes"
181 database (33). The Prokka-annotated amino acid files were also used to determine which
182 transport proteins were present in each bin, by submission to the TransportDB 2.0 TransAAP
183 (membranetransport.org) for transporter annotation (34). In order to characterise the
184 carbohydrate active enzymes, dbCAN HMMdb v10.0 (bcb.unl.edu/dbCAN2) was run on the
185 unannotated contigs within each bin (35). Similarly, the CRISPR sequences and *cas* genes were
186 identified in the contigs within each bin using CRISPRCasFinder online (crisprcas.i2bc.paris-
187 saclay.fr; (36)). Resultant data from these tools were summarised and used to characterise the
188 genomes in each bin and estimate their ecological roles within the coal seam environment.
189
190 **Identification of the bins**
191 As all bins lacked 16S ribosomal DNA (rDNA) regions, the elongation factor G genes were used to
192 identify the closest known relatives of each bin. For this, BLASTX searches of the non-redundant
193 protein database were used (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Subsequently, putative
194 identities from elongation factor G were compared with BLASTN searches of 16S rDNA sequence
195 data obtained from each metagenome using the Earth Microbiome Project V4 primers with Kelpie
196 (37–39). This manual comparison between the elongation factor G taxonomy and Kelpie-derived
197 16S rDNA data was used to identify putative taxa at the level of family or lower, with a minimum
198 per-taxon abundance of 10 applied for the 16S rDNA data. In addition, all 16S rDNA sequences
199 generated using Kelpie were then compared to the Coal Seam Microbiome (CSMB) reference set,

200 using USEARCH v11.0.667 at 97 % identity, to determine any associated CSMB reference taxa (8,
201 40).
202
203 Where close elongation factor G sequence % identities were not found with BLASTX, the
204 phylogeny of the bin was estimated against other Prokka-annotated representative genomes using
205 multilocus sequence analysis (MLSA). The common housekeeping genes selected for this were
206 restricted to those present within both the bin and the majority of the representative genomes.
207 The constructed fasta files of the selected housekeeping genes were aligned and clustered in
208 Mega v11.0.13 (41) with the ClustalW function, and edited as Newick tree format files in FigTree
209 (http://tree.bio.ed.ac.uk/software/figtree/).
210
211 **Figures**
212 Final editing of each figure was performed using Inkscape v1.1.2 or Adobe Illustrator v26.0.2.
213
214 **Data availability**
215 Final contig bins, Prokka nucleotide annotations, and Kelpie-generated 16S rDNA sequences were
216 uploaded to the CSIRO Data Access Portal (https://data.csiro.au/collection/csiro:54006). An
217 overview of the methods can be found in Supplementary Figure S1.
218
219 **RESULTS**
220
221 **Distribution of benzoyl-CoA reductase genes across the examined metagenomes**
222 Subunits of the benzoyl-CoA reductase gene (*bcrABCD)* were detected in the Prokka annotations
223 of all examined metagenomes except for Central Appalachian 89 (Supplementary Table S2). The
224 greatest number of benzoyl-CoA reductase subunits were detected in the Powder River 40

225 metagenome, with 43 subunits observed. In contrast, aside from the Central Appalachian 89
226 metagenome, the Bowen 3 metagenome contained the fewest number of benzoyl-CoA reductase
227 subunits (4 subunits; Supplementary Table S2). Overall, subunits *bcrB* and *bcrC* were identified far
228 more often than subunits *bcrA* and *bcrD*; no *bcrA* or *bcrD* subunits were detected in the Australian
229 metagenomes, and in the Powder River Basin metagenomes *bcrBC* were an order of magnitude
230 more abundant than *bcrAD* (215 total compared to 23 total, respectively).
231
232 Twenty-one bins were obtained that contained benzoyl-CoA reductase gene subunits, and
233 comparison of these indicated three groupings of similar genomes (Supplementary Table S3).
234 From these groupings, four bins were characterised in detail (bins 1.1, 2.2, 2.6 and 3.1; Figure 1;
235 Table 1). The remaining 17 bins were determined to be chimeric, as they were too large to
236 represent a single genome, or to have unacceptably low CheckM contamination scores
237 (Supplementary Table S3).
238
239 **Phylogeny of the four contig bins that were chosen for further analysis**
240 16S rDNA sequences were not recovered in any of the 21 contig bins compiled in this study,
241 however, phylogenetic analyses using BLASTX to identify elongation factor G gene sequences with
242 high % identities to the four contig bins and probable 16S rDNA OTUs revealed that the four bins
243 selected for characterisation represented a Syntrophaceae species (bin 1.1), two
244 *Syntrophorhabdus aromaticivorans* species (bins 2.2 and 2.6), and a novel member of the
245 Fibrobacteres-Chlorobi-Bacteroidetes (FCB) superphylum (bin 3.1). Although the closest relative of
246 bin 1.1 by its elongation factor G gene was a Syntrophaceae sp., possible Syntrophaceae matches
247 for this bin in the 16S rDNA results from the metagenome were limited to two distinct taxa
248 (OTU_64 and OTU_114; detected 12 and 10 times, respectively). Both of these have >90 %
249 identities to type sequence *Smithella propionica* LYP (96.84 % and 94.86 %, respectively; GenBank

250 accession NR_024989.1) as well as to each of the three formally described species within the
251 *Syntrophus* genus (93.68–95.65 % and 92.89–94.09 %, respectively). Accordingly, bin 1.1 was
252 referred to here as a Syntrophaceae sp., being closely related to the *Smithella* and *Syntrophus*
253 genera. It is noteworthy that, based on phylogenomic analysis, there has been recent discussion
254 around updating the phylogenetic ranks and assignments within the Deltaproteobacteria.
255 Accordingly, *S. aromaticivorans* and Syntrophaceae sp. would be reclassified into a phylum named
256 Thermodesulfobacteriota or Desulfobacterota, and, to incorporate high percent identities to both
257 *S. propionica* and the *Syntrophus* spp., the nearest common rank would then be the Syntrophales
258 order (42, 43) for bin 1.1. The nomenclature used in the present study does not reflect these
259 proposed updates, and instead aligns with currently entries for Deltaproteobacteria and
260 Syntrophaceae in the LPSN (https://lpsn.dsmz.de/).
261
262 Although 16S rDNA sequences were able to be matched by close relatives using only elongation
263 factor G genes for the Deltaproteobacteria taxa (Syntrophaceae sp. bin 1.1, and *Syntrophorhabdus*
264 *aromaticivorans* bins 2.2 and 2.6), the novelty of bin 3.1 impeded identification through close
265 relatives.
266
267 Multilocus sequence analysis (MLSA) was performed to further characterise bin 3.1, after it was
268 placed loosely within the FCB superphylum by low % identity NCBI BLAST results using its
269 elongation factor G gene. MLSA was performed first on 37 representative genomes from the FCB
270 superphylum (44), and then was also performed on 47 genomes from the Krumholzibacteriota,
271 Delphibacteria, and Latescibacteria candidate phyla; the portion of the FCB superphylum that was
272 most closely related to bin 3.1 (Supplementary Figure S2; Supplementary Table S4). These phyla
273 are not yet well resolved, and the Genome Taxonomy Database classifies the Delphibacteria and
274 some Latescibacteria within the Krumholzibacteriota phylum (45). From these MAGs, six

275 housekeeping genes were selected for analysis: adenylosuccinate synthase, ATP synthase subunit
276 beta, elongation factor G, elongation factor P, malate dehydrogenase, and protein RecA genes,
277 which were identified in bin 3.1 in addition to a maximum number of the representative genomes.
278 Importantly, each of these genes occurred within two or more genomes from Delphibacteria and
279 Krumholzibacteriota; the two phyla with which bin 3.1 had the highest % identities in the BLASTX
280 results for the elongation factor G gene. The results of the MLSA indicate that bin 3.1 is likely a
281 novel member of the Krumholzibacteriota phylum (Figure 2), of which the closest known relative
282 appears to be from a marine sediment enrichment culture from the Bothnian Sea (Supplementary
283 Table S4) (46).
284
285 For the purpose of aiding cross-study analysis of Krumholzibacteriota sp. bin 3.1, the four MAGs in
286 the MLSA results that contained 16S rDNA sequences (Figure 2) were used to determine both the
287 most probable 16S rDNA OTU in the present study, and the most probable CSMB reference set
288 sequence (8). These four MAGs consisted of a Delphibacteria, two Krumholzibacteriota, and a
289 Latescibacteria MAG (NCBI Assemblies GCA_002747875, GCA_022711765, GCA_024276305, and
290 GCA_012103555, respectively). Each of these were compared against the metagenomic 16S rDNA
291 OTUs and the CSMB reference set using USEARCH (90 % minimum threshold). Although no 16S
292 rDNA OTU or CSMB identities were found for Krumholzibacteriota GCA_024276305 or
293 Latescibacteria GCA_012103555, the USEARCH results indicated that Krumholzibacteriota sp. bin
294 3.1 may correspond to OTU_57 (94.1 and 92.9 % identities; GCA_022711765 and
295 GCA_002747875), which also is in agreement with the distribution of similar bin 3 group MAGs
296 across the metagenomes in the present study (Supplementary Table S3; assuming loss of 16S rDNA
297 sequences from rarer taxa in the more shallowly sequenced metagenomes). The CSMB identities
298 indicated that Krumholzibacteriota sp. bin 3.1 may correspond to CSMB_1092 (95.5 and 93.9 %

299 identities; GCA_022711765 and GCA_002747875), which has been previously reported from coal
300 seams in the Bowen, Surat, and Sydney basins, Australia (Supplementary Table S5).
301
302 **General description of the Syntrophaceae sp., *S. aromaticivorans*, and Krumholzibacteriota sp.**
303 **genome bins**
304 The four bins chosen for further characterisation ranged in size from 1.96 to 4.18 Mbp, with *S.*
305 *aromaticivorans* bin 2.6 having the smallest bin size, and the Krumholzibacteriota sp. having the
306 largest bin size (Table 1). GC-content was broadly similar for the three Deltaproteobacteria MAGs
307 (bins 1.1, 2.2 and 2.6), with an average of 46.7 %. Krumholzibacteriota sp. bin 3.1, however, had a
308 higher GC-content of 71.1 % (Table 1) consistent with the other closely related
309 Krumholzibacteriota MAGs from freshwater lakes and marine sediment (69.2 to 70.8 %; Figure 2;
310 Supplementary Table S4) (46). Across the four bins, CheckM results indicated that contamination
311 was very low (<2 %). Although only taxonomically resolved to the phylum level, genome
312 completeness was greatest for Krumholzibacteriota sp. bin 3.1 (95 %), whilst genome
313 completeness was lowest for the Powder River Basin-sourced *S. aromaticivorans* bin 2.6 (55 %;
314 Table 1; Supplementary Table S3). Overall, these bins are medium- to high-quality draft MAGs,
315 based on the combination of completeness and contamination scores with the varying presence of
316 key marker genes within the assemblies (47). Although completeness and contamination scores
317 for the Krumholzibacteriota sp. bin 3.1 were very good, it is ineligible for the classification of high-
318 quality draft due to the lack of recovered rDNA sequences.
319
320 **Dearomatisation genes and other notable ecophysiological genomic characteristics**
321 The most commonly detected benzoyl-CoA reductase gene subunits were those of *bcrABCD*, from
322 the class-I enzyme system, associated with facultative anaerobes such as *Magnetospirillum* and
323 *Thauera* spp. (15, 48). The Prokka annotations contained the highest number of these, however,

324 the BlastKOALA annotations included an additional enzyme system (Table 2; Supplementary Table
325 S2). This additional enzyme system, found in *S. aromaticivorans* bin 2.2, was the ATP-independent
326 class-II benzoyl-CoA reductase gene subunit, *bamB*, associated with obligate anaerobes (Figure 1)
327 (48).
328
329 In addition to their use for MLSA, the Prokka annotations of the representative MAGs from the
330 FCB superphylum, including those from the Krumholzibacteriota, Latescibacteria, and
331 Delphibacteria phyla, were searched for benzoyl-CoA pathway genes. From the 37 MAGs used for
332 MLSA from the FCB superphylum, no genes of either the *bcr* or *bam* enzyme systems were
333 detected (Supplementary Data Table S4a). Other genes from the benzoyl-CoA pathway were also
334 rare, with the highest number (two) being found in the MAG of *Longimicrobium terrae* CECT 8660,
335 from the Gemmatimonadetes phylum, for the production of benzoyl-CoA from 4-hydroxybenzoyl-
336 CoA and benzoate (*hcrABC* and *badA*; respectively). In contrast, the *bcr* gene and other benzoyl-
337 CoA reductase genes were far more common among the Krumholzibacteriota, Latescibacteria, and
338 Delphibacteria MAGs used for MLSA: 17 out of 40 contained one or more subunits from *bcrABCD*,
339 with 12 of these containing all four subunits (Supplementary Data Table S4b) (46, 49, 50). Further,
340 25 of the 40 MAGs contained one or more genes (or gene subunits) from the benzoyl-CoA
341 pathway. Of these, four taxa contained an identical distribution of pathway genes to
342 Krumholzibacteriota sp. bin 3.1, and were sourced from a hydrothermal vent chimney, and
343 freshwater lakes in Puerto Rico and Finland (GCA_024275485, GCA_903859215, GCA_903912525,
344 and GCA_903834005).
345
346 Aside from the benzoyl-CoA pathway genes, the most notable BlastKOALA results regarded
347 flagellar assembly and chemotaxis, anabolic antimicrobial genes, and nutrient and electron
348 acceptor scavenging. Substantial numbers of genes for flagellar assembly and chemotaxis were

349 detected within Krumholzibacteriota sp. bin 3.1, as well as genes for biosynthesis of antimicrobial
350 substances. Genes for nutrient and electron acceptor scavenging of nitrogen and sulfur
351 compounds were most commonly detected within the *S. aromaticivorans* bins 2.2 and 2.6.
352
353 **Carbohydrate-active enzymes**
354 Each of the four bins contained a small number of carbohydrate-active enzymes (Table 2;
355 Supplementary Table S2). The highest number of carbohydrate-active enzymes was detected in
356 Krumholzibacteriota sp. bin 3.1 (143 total), of which the majority were glycosyltransferases and
357 glycoside hydrolases. The lowest number of carbohydrate-active enzymes was found in *S.*
358 *aromaticivorans* bin 2.6 (44 total), more than half of which were glycosyltransferases. Overall,
359 glycosyltransferases were the most commonly detected carbohydrate-active enzymes in each bin.
360 No polysaccharide lyases were detected in any of the bins.
361
362 **Membrane transporter proteins**
363 Numerous transporter protein genes were detected in each bin (Table 2; Supplementary Table S2).
364 *Syntrophorhabdus aromaticivorans* bins 2.2 and 2.6, along with Krumholzibacteriota sp. bin 3.1,
365 contained higher numbers of transporter proteins (totalling 269, 254 and 266 respectively). In
366 contrast, Syntrophaceae sp. bin 1.1 contained a lower number of transporter proteins (totalling
367 171). Across all four bins, the ATP-binding cassette transporters were the most commonly
368 detected, however, it is also notable that the tripartite ATP-independent periplasmic transporters
369 were unusually abundant within the *S. aromaticivorans* bins 2.2 and 2.6 (32 and 36, respectively).
370
371 **CRISPRs and *cas* genes**
372 Three of the four bins included multiple CRISPR loci, and two of these (bins 1.1 and 2.2) contained
373 several *cas* genes (Table 2; Supplementary Table S2). Syntrophaceae sp. bin 1.1 had four CRISPR

374 loci containing 18 spacers, *S. aromaticivorans* bin 2.2 had three CRISPR loci and 30 spacers, and
375 Krumholzibacteriota sp. bin 3.1 had two CRISPR loci and seven spacers. Notably, while
376 Syntrophaceae sp. bin 1.1 and *S. aromaticivorans* bin 2.2 contained the aforementioned *cas* genes
377 (7 and 9, respectively), no *cas* genes were detected within Krumholzibacteriota sp. bin 3.1 and
378 neither CRISPR loci nor *cas* genes were detected in *S. aromaticivorans* bin 2.6.
379
380 **DISCUSSION**
381
382 For microbiologists working to understand the catabolism of coal in the subsurface, the
383 identification of taxa engaged in degrading aromatic compounds has proved elusive. Indeed,
384 microbial communities from the coal seam environment tend to be numerically dominated by one
385 or two methanogens (12, 51), a handful of bacterial taxa presumably engaging in syntrophic
386 partnerships with these methanogens, and a long tail of rarer taxa (52). It has generally been the
387 view that bacterial syntrophs could not degrade complex aromatic or aliphatic compounds (4, 53,
388 54), which leaves the largely unexplored tail of the microbial community in coal as a likely place to
389 find those microbes capable of coal degradation. Further, it is ecologically reasonable to
390 hypothesise that this unexplored tail contains microbes with a more diverse array of nutritional
391 strategies, since presumably there are abundant and less-highly competed nutritional niches
392 within the diverse heterogeneity of the organic matter in coal. Data presented here demonstrate
393 for the first time the identity of multiple lineages with crucial monoaromatic-degrading genes from
394 subsurface coal seams on two continents. Four taxa were observed in this study: one
395 Syntrophaceae sp. (bin 1.1), two *Syntrophorhabdus aromaticivorans* (bins 2.2 and 2.6) and a novel
396 taxon from the Krumholzibacteriota phylum in the FCB superphylum (bin 3.1). This
397 Krumholzibacteriota sp. is likely engaged in monoaromatic degradation in the Powder River Basin,
398 and had not previously been implicated as having a role in these processes in the coal seam

399 environment. While it has been previously stated that *S. aromaticivorans* may be important for
400 coal degradation (22), this is the first demonstration that coal seam strains of this taxon carry
401 genes for monoaromatic degradation. For the Syntrophaceae sp., taxa within this family are well-
402 known hydrocarbon-degraders in subsurface environments, such as aliphatic molecule
403 degradation in oil reservoirs (26), however, they had not previously been implicated in aromatic
404 degradation in the coal seam environment.
405
406 **Genomic characterisation of the Krumholzibacteriota sp. bin3.1 MAG**
407 Krumholzibacteriota taxa are identified here for the first time as aromatic-degraders in the coal
408 seam environment, as well as more broadly. From the annotations of its genome, the
409 Krumholzibacteriota sp. is a primary coal degrader with the ability to anaerobically catabolise a
410 wide range of monoaromatic compounds (Figure 3; Supplementary Table S2). Further, the
411 Krumholzibacteriota sp. may also compete in other ecological niches such as biomass recycling,
412 since in addition to genes for monoaromatic degradation, it also has a high number of glycoside
413 hydrolase genes (relative to the three Deltaproteobacteria MAGs; Table 2). It is probable that the
414 organism is performing biomass recycling to supplement its catabolism of carbon from aromatic
415 substrates, rather than specialising in biomass recycling, since the abundance of glycoside
416 hydrolase genes is lower in the Krumholzibacteriota sp. relative to other biomass degraders
417 described from the coal seam environment (52). In addition to being able to obtain nitrogen and
418 phosphorus through biomass recycling, the Krumholzibacteriota sp. may be capable of extracting
419 nutrients from coal via the degradation of nitrogen and phosphate-containing aromatic
420 compounds, since it has a near-complete set of genes for the catabolism of benzamide and
421 benzoyl phosphate to acetate (Figure 3). Moreover, it is also the only MAG described here to
422 contain most of the genes necessary for flagellar assembly and chemotaxis, which it may use to
423 seek out desirable regions of organic matter within the coal. Finally, the Krumholzibacteriota sp.

424 contained a high number of multidrug efflux pumps relative to the other organisms examined
425 here, which may be relevant for removal of toxic metabolites produced during catabolism of
426 aromatic substrates. Overall, these findings implicate this Krumholzibacteriota sp. as potentially
427 important for both the dearomatisation of monoaromatic compounds, and the production of
428 simpler methanogenic substrates in the coal seam environment. Further studies verifying the
429 function of this microbe in the coal seam environment are needed, ideally in the form of axenic
430 culturing to test its response to different variables such as complex carbon substrates. Axenic
431 culturing would also provide details on the cell morphology of this Krumholzibacteriota sp., and
432 may aid in understanding its ability to access different regions within the coal seam environment.
433 Such efforts to obtain this taxon in pure culture may be tailored using its described characteristics
434 from the genome provided here.
435
436 Taxonomically, the classification of this Krumholzibacteriota sp. was more challenging than for the
437 Deltaproteobacteria MAGs, both because the assembly contained no annotated sequences of 16S
438 rDNA, and because the elongation factor G gene did not return any high % identities to other well
439 characterised genomes in the NCBI collection. The results of the MLSA indicate that the closest
440 relatives of this Krumholzibacteriota sp. were other taxa from within the same phylum, found in
441 marine sediments from the Bothnian Sea in Scandinavia (46), a stratified freshwater reservoir of
442 the Cañas River in Puerto Rico, and Lake Lovojärvi in Finland (Figure 2; Supplementary Table S4).
443 Although 16S rDNA sequences for the Krumholzibacteriota sp. (OTU_57; CSMB_1092) were able to
444 be inferred using four representative MAGs of the same phylum, confirmation of the 16S rDNA
445 sequences specific to this Krumholzibacteriota sp. is needed for global context in the subsurface
446 coal seam environment, as well as other environments where the Krumholzibacteriota spp. may
447 play a key ecological role.
448

449 **Krumholzibacteriota is a globally-relevant phylum for liberation of recalcitrant carbon**

450 Indeed, dearomatisation and downstream catabolism of monoaromatic compounds in anoxic
451 environments may be relatively common within the Krumholzibacteriota phylum (Figure 2). Of the
452 40 other Krumholzibacteriota MAGs used for MLSA, 13 contained a complete or near-complete set
453 of genes for catabolism of benzoyl-CoA to 3-hydroxypimeloyl-CoA (Supplementary Table S4). In
454 addition to this genetic potential for anaerobic dearomatisation, aerobic dearomatisation has
455 recently been suggested regarding a Krumholzibacteriota MAG that contained one subunit of the
456 gene encoding for benzoate/toluate 1,2-dioxygenase reductase (*benC-xylZ*), and was sourced from
457 an oil-contaminated environment in the Persian Gulf (55). Again, axenic studies of these carbon-
458 liberating Krumholzibacteriota would be beneficial in clarifying the phylogenetic and metabolic
459 capabilities of taxa within this phylum. Specifically, this clarification is essential for a clearer
460 understanding of the potential rate-limiting ecological roles of Krumholzibacteriota spp. in the
461 subsurface coal seam environment. Axenic studies can also aid in understanding the roles of these
462 organisms in other carbon-cycling environments where they have been detected, such as marine
463 sediments, deep marine cold seep fluids, hydrothermal vents, and surface freshwaters.

464

465 ***Syntrophorhabdus aromaticivorans* in the coal seam environment**

466 Unlike the novel Krumholzibacteriota sp., *S. aromaticivorans* has been identified within the coal
467 seam environment on numerous previous occasions. Indeed, the CSMB reference set presently
468 lists six distinct taxa from the *Syntrophorhabdus* genus in coal-bearing basins of Australia
469 (Supplementary Table S5), and they have also been detected in the Powder River Basin in the USA
470 (22). A recent study of microbial communities from the Powder River Basin implicated this taxon
471 as an important potential primary degrader of organic matter in coal, based on both its increases
472 in abundance after algal-amendment, and the known degradative abilities of the single strain of
473 this taxon in culture (23). It should be noted, however, that only a single isolate of this species has

474 been described, and this isolate was obtained from wastewater, not from the subsurface. It was
475 thus unclear whether other strains of *S. aromaticivorans* were capable of aromatic degradation.
476
477 Data from the current study demonstrate, for the first time, that two other genomes of *S.*
478 *aromaticivorans* have at least partial pathways for the degradation of monoaromatic compounds
479 (Figure 1). This is likely to be a common feature in the genome of this species, given that these two
480 new strains come from different coal seams on different continents. Taken together, these data
481 indicate that *S. aromaticivorans* can likely directly access carbon contained within coal via
482 aromatic molecules, such as 4-hydroxybenzoate and benzoate. In addition to its ability to
483 catabolise monoaromatic compounds, both the North American and Australian *S. aromaticivorans*
484 strains described here appear to have an unusually large array of tripartite ATP-independent
485 periplasmic (TRAP) transporters, which may indicate that dicarboxylates (such as fumarate) are an
486 important source of carbon for *S. aromaticivorans* (Figure 3). Biomass recycling, on the other
487 hand, is unlikely to be a source of carbon for this taxon, as neither strain contained substantial
488 numbers of genes for glycoside hydrolase and other carbohydrate enzymes. Interestingly, the
489 Australian *S. aromaticivorans* genome contains abundant CRISPR spacers, suggesting substantial
490 pressure on this taxon from viral predation in the Bowen Basin. In contrast, the lack of CRISPR
491 sequences detected in the North American genome indicates little to no viral predation stress on
492 this taxon in the Powder River Basin. Viral predation has been previously implicated as an
493 important process in the Surat Basin (14), and numerous other taxa from this environment have
494 been previously demonstrated to harbour significantly CRISPR spacers arrayed against a host of
495 viruses (10, 56). The absence of these sequences in the Powder River Basin may indicate this
496 species, at least, is not under the same level of viral predation as in other subsurface systems.
497 Alternatively, it may be that these sequences were simply not recovered due to the insufficient
498 completeness of this bin (Table 1). Lastly, the ability of *S. aromaticivorans* to use organic electron

499 acceptors (9,10-anthraquinone-2,6-disulfonate) when isolated from other environments is
500 intriguing (23), and it may be that aromatic compounds from the coal seam environment could
501 also act as alternative electron acceptors, though experimental work with the organism in culture
502 would be important to confirm this speculation. Regardless, certainly for Australian strains, there
503 is clear evidence that this species has the genetic potential to catabolise aromatic compounds
504 from the coal seam, and, if these taxa are also subject to viral predation, it may be a mechanism
505 by which this carbon is made available to a wider range of taxa post-lysis of the *S. aromaticivorans*
506 cells.
507
508 **Syntrophaceae taxa may utilise aromatic carbon in coal seams**
509 The other Deltaproteobacteria taxon identified was a bin associated with the family
510 Syntrophaceae, having genes for monoaromatic degradation in the Powder River Basin. Data from
511 the CSMB reference set reveal that the Syntrophaceae family is commonly present within coal
512 seams in Australia, in all of the North American coal seams in the present study, and also in the
513 Ishikari Basin, Japan (Supplementary Table S5; (8)). The Syntrophaceae family, and *Smithella*
514 *propionica* specifically, is well known for its use of aliphatic hydrocarbons (26, 57). Indeed,
515 *Smithella* spp. are well-known alkane-degraders from anoxic environments such as oil reservoirs,
516 where they can grow in syntrophy with hydrogenotrophic methanogens (26). Despite this
517 prevalence in hydrocarbon-rich environments, Syntrophaceae sp. bin 1.1 represents the first time
518 that this family has been implicated in primary degradation of aromatic substrates in the coal
519 seam environment. Of additional note regarding metabolism and ecological niches of this taxon, is
520 the absence of substantial numbers of genes for carbohydrate degradation. This absence, in
521 addition to its low number of ABC transporters, suggests that it does not participate in biomass
522 recycling, nor does it experience substantial competitive stress. This genome does, however,
523 appear to experience considerable pressure from viral predation, as it contains a high number of

524 CRISPR spacers relative to the other Powder River Basin taxa examined here (Figure 3; Table 2).
525 The Syntrophaceae sp. co-occurs with the relatively viral predation-free *S. aromaticivorans* in the
526 Powder River Basin, and it thus seems likely that viral predation may be an important process in
527 the Powder River Basin, as it is in other subsurface habitats.
528
529 **Ecological functions and implications for understanding the coal seam microbiome**
530 In terms of their life strategies in the coal seam, all three taxa likely have stress-tolerant ecological
531 profiles (in the sense of Grime (58)). All three appear to possess specialised genetic tools for the
532 degradation of a plentiful, but difficult to access, carbon source within subsurface coal seams.
533 Interestingly, the Krumholzibacteriota taxon, as well as hosting these genes for monoaromatic
534 degradation, also hosts an array of other genes for accessing nutrients in moribund cells or plant
535 material (Figure 3; Table 2). As described earlier, coal seams are oligotrophic environments, and
536 access to other macronutrients, such as nitrogen, in coal organic matter may be important for
537 competition in this environment. Further, while it appears that the Krumholzibacteriota organism
538 described here may have a relatively high number of genes involved in carbohydrate metabolism,
539 it has relatively few compared to truly ruderal taxa that occur in coal seams (14).
540
541 Importantly for scientific efforts to enhance or control methane yields from coal, the four putative
542 aromatic-degrading taxa described here may hold key rate-limiting roles in the biodegradation of
543 coal to methane in the subsurface. If the Krumholzibacteriota sp., the two *S. aromaticivorans*, and
544 the Syntrophaceae sp. are all indeed capable of direct access to the carbon within coal, further
545 study of their metabolic strategies may provide important tools for altering biodegradation of coal
546 in the subsurface. One key goal of future research should be to obtain isolates of each of these
547 taxa, which would not only be beneficial in confirming their function, but also provide a suite of
548 tools for further manipulative experiments.

549

550 Much of the previous research into the coal seam microbiome has centred around either
551 descriptive studies of species distributions, or the effect of nutrient mixtures in an effort to
552 enhance gas yields. While both of these approaches are valuable, they provide comparatively little
553 information on the function of individual microbes in these communities. In contrast, this study
554 describes the genomes of four new taxa from the coal seam environment with likely roles in
555 monoaromatic compound degradation. Studies that seek to elucidate processes upstream of
556 monoaromatic degradation, involving the liberation of soluble organic matter from the insoluble
557 coal macromolecule, would further our understanding of this unusual environment.

558

559 **Acknowledgements**

560 The authors thank Dr. Andrew G. McLeish for running the contig bins with CheckM. Dr. Bronwyn
561 Campbell was supported by a Macquarie University Research Training Program scholarship, a
562 CSIRO Energy strategic research initiative, and a PESA Horstmann Federal Postgraduate
563 Scholarship.

564

565 **Competing interests statement**

566 The authors declare no competing interests.

567

568 **Data availability**

569 The datasets sourced for analysis in the current study were downloaded from the NCBI Sequence
570 Read Archive (from BioProject accessions PRJNA330673, PRJNA291107, and PRJNA678021;
571 sequence run accessions can be found in Supplementary Table S1), the JGI Data Portal (Gold
572 project IDs Gp0406113 to Gp0406117, and the CSIRO Data Access Portal
573 (https://doi.org/10.4225/08/5b31ca6373d48).

The MAGs (as contig bins and Prokka annotations), 16S rDNA OTU sequences, and CSMB-OTU 16S rDNA data table generated during the current study are each available in the CSIRO Data Access Portal (https://doi.org/10.25919/d4nf-8f32).

Disclaimer

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Reference List

1. Hardisty PE, Clark TS, Hynes RG. 2012. Life Cycle Greenhouse Gas Emissions from Electricity Generation: A Comparative Analysis of Australian Energy Sources. 4. Energies 5:872–897.

2. Markandya A, Wilkinson P. 2007. Electricity generation and health. The Lancet 370:979–990.

3. Schandl H, Baynes T, Haque N, Barrett D, Geschke A. 2019. Final Report for GISERA Project G2 - Whole of Life Greenhouse Gas Emissions Assessment of a Coal Seam Gas to Liquefied Natural Gas Project in the Surat Basin, Queensland, Australia. CSIRO Aust.

4. Strąpoć D, Mastalerz M, Dawson K, Macalady J, Callaghan AV, Wawrik B, Turich C, Ashby M. 2011. Biogeochemistry of Microbial Coal-Bed Methane. Annu Rev Earth Planet Sci 39:617–656.

5. Ritter D, Vinson D, Barnhart E, Akob DM, Fields MW, Cunningham AB, Orem W, McIntosh JC. 2015. Enhanced microbial coalbed methane generation: A review of research, commercial activity, and remaining challenges. Int J Coal Geol 146:28–41.

597 6. Wang H, Lin H, Rosewarne CP, Li D, Gong S, Hendry P, Greenfield P, Sherwood N, Midgley
598 DJ. 2016. Enhancing biogenic methane generation from a brown coal by combining different
599 microbial communities. Int J Coal Geol 154–155:107–110.

600 7. He X, Liu X, Nie B, Song D. 2017. FTIR and Raman spectroscopy characterization of
601 functional groups in various rank coals. Fuel 206:555–563.

602 8. Vick SHW, Greenfield P, Tran-Dinh N, Tetu SG, Midgley DJ, Paulsen IT. 2018. The Coal
603 Seam Microbiome (CSMB) reference set, a *lingua franca* for the microbial coal-to-methane
604 community. Int J Coal Geol 186:41–50.

605 9. Wang B, Yu Z, Zhang Y, Zhang H. 2019. Microbial communities from the Huaibei Coalfield
606 alter the physicochemical properties of coal in methanogenic bioconversion. Int J Coal Geol
607 202:85–94.

608 10. McLeish AG, Greenfield P, Midgley DJ, Paulsen IT. 2021. Draft Genome Sequence of
609 *Desulfovibrio* sp. Strain CSMB_222, Isolated from Coal Seam Formation Water. Microbiol
610 Resour Announc https://doi.org/10.1128/MRA.00564-21.

611 11. Johnson ER, Klasson KT, Basu R, Volkwein JC, Clausen EC, Gaddy JL. 1994. Microbial
612 conversion of high-rank coals to methane. Appl Biochem Biotechnol 45:329.

613 12. McKay LJ, Smith HJ, Barnhart EP, Schweitzer HD, Malmstrom RR, Goudeau D, Fields MW.
614 2021. Activity-based, genome-resolved metagenomics uncovers key populations and
615 pathways involved in subsurface conversions of coal to methane. ISME J 1–12.

616 13. Schweitzer HD, Smith HJ, Barnhart EP, McKay LJ, Gerlach R, Cunningham AB, Malmstrom
617 RR, Goudeau D, Fields MW. 2022. Subsurface hydrocarbon degradation strategies in low-
618 and high-sulfate coal seam communities identified with activity-based metagenomics. 1. Npj
619 Biofilms Microbiomes 8:1–10.

620 14. Vick SHW, Greenfield P, Tetu SG, Midgley DJ, Paulsen IT. 2019. Genomic and phenotypic
621 insights point to diverse ecological strategies by facultative anaerobes obtained from
622 subsurface coal seams. 1. Sci Rep 9:1–13.

623 15. Carmona M, Zamarro MT, Blázquez B, Durante-Rodríguez G, Juárez JF, Valderrama JA,
624 Barragán MJL, García JL, Díaz E. 2009. Anaerobic catabolism of aromatic compounds: a
625 genetic and genomic view. Microbiol Mol Biol Rev MMBR 73:71–133.

626 16. Foght J. 2008. Anaerobic Biodegradation of Aromatic Hydrocarbons: Pathways and
627 Prospects. Microb Physiol 15:93–120.

628 17. Boll M, Fuchs G. 1995. Benzoyl-Coenzyme A Reductase (Dearomatizing), a Key Enzyme of
629 Anaerobic Aromatic Metabolism. Eur J Biochem 234:921–933.

630 18. McLeish AG, Gong S, Greenfield P, Midgley DJ, Paulsen IT. 2021. Microbial Community
631 Shifts on Organic Rocks of Different Maturities Reveal potential Catabolisers of Organic
632 Matter in Coal. Microb Ecol https://doi.org/10.1007/s00248-021-01857-x.

633 19. Campbell BC, Gong S, Greenfield P, Midgley DJ, Paulsen IT, George SC. 2021. Aromatic
634 compound-degrading taxa in an anoxic coal seam microbiome from the Surat Basin,
635 Australia. FEMS Microbiol Ecol 97:fiab053.

636 20. Kühner S, Wöhlbrand L, Fritz I, Wruck W, Hultschig C, Hufnagel P, Kube M, Reinhardt R,
637 Rabus R. 2005. Substrate-Dependent Regulation of Anaerobic Degradation Pathways for
638 Toluene and Ethylbenzene in a Denitrifying Bacterium, Strain EbN1. J Bacteriol 187:1493–
639 1503.

640 21. Schink B, Philipp B, Müller J. 2000. Anaerobic degradation of phenolic compounds.
641 Naturwissenschaften 87:12–23.

642 22. Smith HJ, Schweitzer HD, Barnhart EP, Orem W, Gerlach R, Fields MW. 2021. Effect of an
643 algal amendment on the microbial conversion of coal to methane at different sulfate
644 concentrations from the Powder River Basin, USA. Int J Coal Geol 248:103860.

645 23. Qiu Y-L, Hanada S, Ohashi A, Harada H, Kamagata Y, Sekiguchi Y. 2008. Syntrophorhabdus
646 aromaticivorans gen. nov., sp. nov., the First Cultured Anaerobe Capable of Degrading
647 Phenol to Acetate in Obligate Syntrophic Associations with a Hydrogenotrophic Methanogen.
648 Appl Environ Microbiol 74:2051–2058.

649 24. Vick SHW, Gong S, Sestak S, Vergara TJ, Pinetown KL, Li Z, Greenfield P, Tetu SG, Midgley
650 DJ, Paulsen IT. 2019. Who eats what? Unravelling microbial conversion of coal to methane.
651 FEMS Microbiol Ecol 95:fiz093.

652 25. Kuever J. 2014. The Family Syntrophaceae, p. 281–288. *In* Rosenberg, E, DeLong, EF, Lory,
653 S, Stackebrandt, E, Thompson, F (eds.), The Prokaryotes: Deltaproteobacteria and
654 Epsilonproteobacteria. Springer, Berlin, Heidelberg.

655 26. Gray ND, Sherry A, Grant RJ, Rowan AK, Hubert CRJ, Callbeck CM, Aitken CM, Jones DM,
656 Adams JJ, Larter SR, Head IM. 2011. The quantitative significance of Syntrophaceae and
657 syntrophic partnerships in methanogenic degradation of crude oil alkanes. Environ Microbiol
658 13:2957–2975.

659 27. Campbell BC, Greenfield P, Gong S, Barnhart EP, Midgley DJ, Paulsen IT, George SC. 2022.
660 Methanogenic archaea in subsurface coal seams are biogeographically distinct: an analysis
661 of metagenomically-derived mcrA sequences. Environ Microbiol 24:4065–4078.

662 28. Greenfield P, Duesing K, Papanicolaou A, Bauer DC. 2014. Blue: correcting sequencing
663 errors using consensus and context. Bioinformatics 30:2723–2732.

664 29. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes De
665 Novo Assembler. Curr Protoc Bioinforma 70:e102.

666 30. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–
667 2069.

668 31. Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, Banfield JF. 2009.
669 Community-wide analysis of microbial genome sequence signatures. Genome Biol 10:R85.

670 32. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing
671 the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
672 Genome Res 25:1043–1055.

673 33. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for
674 Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 428:726–
675 731.

676 34. Elbourne LDH, Tetu SG, Hassan KA, Paulsen IT. 2017. TransportDB 2.0: a database for
677 exploring membrane transporters in sequenced genomes from all domains of life. Nucleic
678 Acids Res 45:D320–D324.

679 35. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018.
680 dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids
681 Res 46:W95–W101.

682 36. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC,
683 Vergnaud G, Gautheret D, Pourcel C. 2018. CRISPRCasFinder, an update of CRISRFinder,
684 includes a portable version, enhanced performance and integrates search for Cas proteins.
685 Nucleic Acids Res 46:W246–W251.

686 37. Greenfield P, Tran-Dinh N, Midgley D. 2019. Kelpie: generating full-length ‘amplicons’ from
687 whole-metagenome datasets. PeerJ 6:e6174.

688 38. Apprill A, McNally S, Parsons R, Weber L. 2015. Minor revision to V4 region SSU rRNA 806R
689 gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol
690 75:129–137.

691 39. Parada AE, Needham DM, Fuhrman JA. 2016. Every base matters: assessing small subunit
692 rRNA primers for marine microbiomes with mock communities, time series and global field
693 samples. Environ Microbiol 18:1403–1414.

694 40. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST.
695 Bioinformatics 26:2460–2461.

696 41. Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis
697 Version 11. Mol Biol Evol 38:3022–3027.

698 42. Waite DW, Chuvochina M, Pelikan C, Parks DH, Yilmaz P, Wagner M, Loy A, Naganuma T,
699 Nakai R, Whitman WB, Hahn MW, Kuever J, Hugenholtz P. 2020. Proposal to reclassify the
700 proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum
701 Thermodesulfobacteria into four phyla reflecting major functional capabilities. Int J Syst Evol
702 Microbiol 70:5972–6016.

703 43. Oren A, Garrity GM. 2021. Valid publication of the names of forty-two phyla of prokaryotes. Int
704 J Syst Evol Microbiol 71:005056.

705 44. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN,
706 Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R,
707 Thomas BC, Banfield JF. 2016. A new view of the tree of life. 5. Nat Microbiol 1:1–6.

708 45. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 2022. GTDB:
709 an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent,
710 rank normalized and complete genome-based taxonomy. Nucleic Acids Res 50:D785–D794.

711 46. Dalcin Martins P, de Jong A, Lenstra WK, van Helmond NAGM, Slomp CP, Jetten MSM,
712 Welte CU, Rasigraf O. 2021. Enrichment of novel Verrucomicrobia, Bacteroidetes, and
713 Krumholzibacteria in an oxygen-limited methane- and iron-fed bioreactor inoculated with
714 Bothnian Sea sediments. MicrobiologyOpen 10:e1175.

715 47. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz
716 F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A,
717 Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM,
718 Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ,

719 Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen
720 JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I,
721 Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Murat Eren A, Schriml L,
722 Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified
723 genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea.
724 8. Nat Biotechnol 35:725–731.

725 48. von Netzer F, Granitsiotis MS, Szalay AR, Lueders T. 2020. Next-Generation Sequencing of
726 Functional Marker Genes for Anaerobic Degraders of Petroleum Hydrocarbons in
727 Contaminated Environments, p. 257–276. *In* Boll, M (ed.), Anaerobic Utilization of
728 Hydrocarbons, Oils, and Lipids. Springer International Publishing, Cham.

729 49. Murphy CL, Biggerstaff J, Eichhorn A, Ewing E, Shahan R, Soriano D, Stewart S, VanMol K,
730 Walker R, Walters P, Elshahed MS, Youssef NH. 2021. Genomic characterization of three
731 novel Desulfobacterota classes expand the metabolic and phylogenetic diversity of the
732 phylum. Environ Microbiol 23:4326–4343.

733 50. Vega MAP, Scholes RC, Brady AR, Daly RA, Narrowe AB, Bosworth LB, Wrighton KC,
734 Sedlak DL, Sharp JO. 2022. Pharmaceutical Biotransformation is Influenced by
735 Photosynthesis and Microbial Nitrogen Cycling in a Benthic Wetland Biomat. Environ Sci
736 Technol 56:14462–14477.

737 51. Rosewarne CP, Greenfield P, Li D, Tran-Dinh N, Midgley DJ, Hendry P. 2013. Draft Genome
738 Sequence of *Methanobacterium* sp. Maddingley, Reconstructed from Metagenomic
739 Sequencing of a Methanogenic Microbial Consortium Enriched from Coal-Seam Gas
740 Formation Water. Genome Announc 1:e00082-12.

741 52. Vick SHW, Greenfield P, Pinetown KL, Sherwood N, Gong S, Tetu SG, Midgley DJ, Paulsen
742 IT. 2019. Succession Patterns and Physical Niche Partitioning in Microbial Communities from
743 Subsurface Coal Seams. iScience 12:152–167.

53. Rosewarne CP, Greenfield P, Li D, Tran-Dinh N, Bradbury MI, Midgley DJ, Hendry P. 2013. Draft Genome Sequence of *Clostridium* sp. Maddingley, Isolated from Coal-Seam Gas Formation Water. Genome Announc 1:e00081-12.

54. Schink B. 2006. Syntrophic Associations in Methanogenic Degradation, p. 1–19. *In* Overmann, J (ed.), Molecular Basis of Symbiosis. Springer, Berlin, Heidelberg.

55. Rezaei Somee M, Dastgheib SMM, Shavandi M, Ghanbari Maman L, Kavousi K, Amoozegar MA, Mehrshad M. 2021. Distinct microbial community along the chronic oil pollution continuum of the Persian Gulf converge with oil spill accidents. 1. Sci Rep 11:11316.

56. Vick SHW, Greenfield P, Willows RD, Tetu SG, Midgley DJ, Paulsen IT. 2019. Subsurface Stappia: Success Through Defence, Specialisation and Putative Pressure-Dependent Carbon Fixation. Microb Ecol 80:34–46.

57. Zengler K, Richnow HH, Rosselló-Mora R, Michaelis W, Widdel F. 1999. Methane formation from long-chain alkanes by anaerobic microorganisms. 6750. Nature 401:266–269.

58. Grime JP. 1977. Evidence for the existence of three primary strategies in plants and its relevance to ecological evolutionary theory. Am Nat 111:1169–1194.

59. Bik EM, Costello EK, Switzer AD, Callahan BJ, Holmes SP, Wells RS, Carlin KP, Jensen ED, Venn-Watson S, Relman DA. 2016. Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. 1. Nat Commun 7:10516.

60. Storey MA, Andreassend SK, Bracegirdle J, Brown A, Keyzers RA, Ackerley DF, Northcote PT, Owen JG. 2020. Metagenomic Exploration of the Marine Sponge Mycale hentscheli Uncovers Multiple Polyketide-Producing Bacterial Symbionts. mBio 11:e02997-19.

61. Robbins SJ, Singleton CM, Chan CX, Messer LF, Geers AU, Ying H, Baker A, Bell SC, Morrow KM, Ragan MA, Miller DJ, Forêt S, Voolstra CR, Tyson GW, Bourne DG. 2019. A genomic view of the reef-building coral Porites lutea and its microbial symbionts. 12. Nat Microbiol 4:2090–2100.

62. Wang W, Tao J, Yu K, He C, Wang J, Li P, Chen H, Xu B, Shi Q, Zhang C. 2021. Vertical Stratification of Dissolved Organic Matter Linked to Distinct Microbial Communities in Subtropic Estuarine Sediments. Front Microbiol 12:697860.

773 **Figure captions**

774

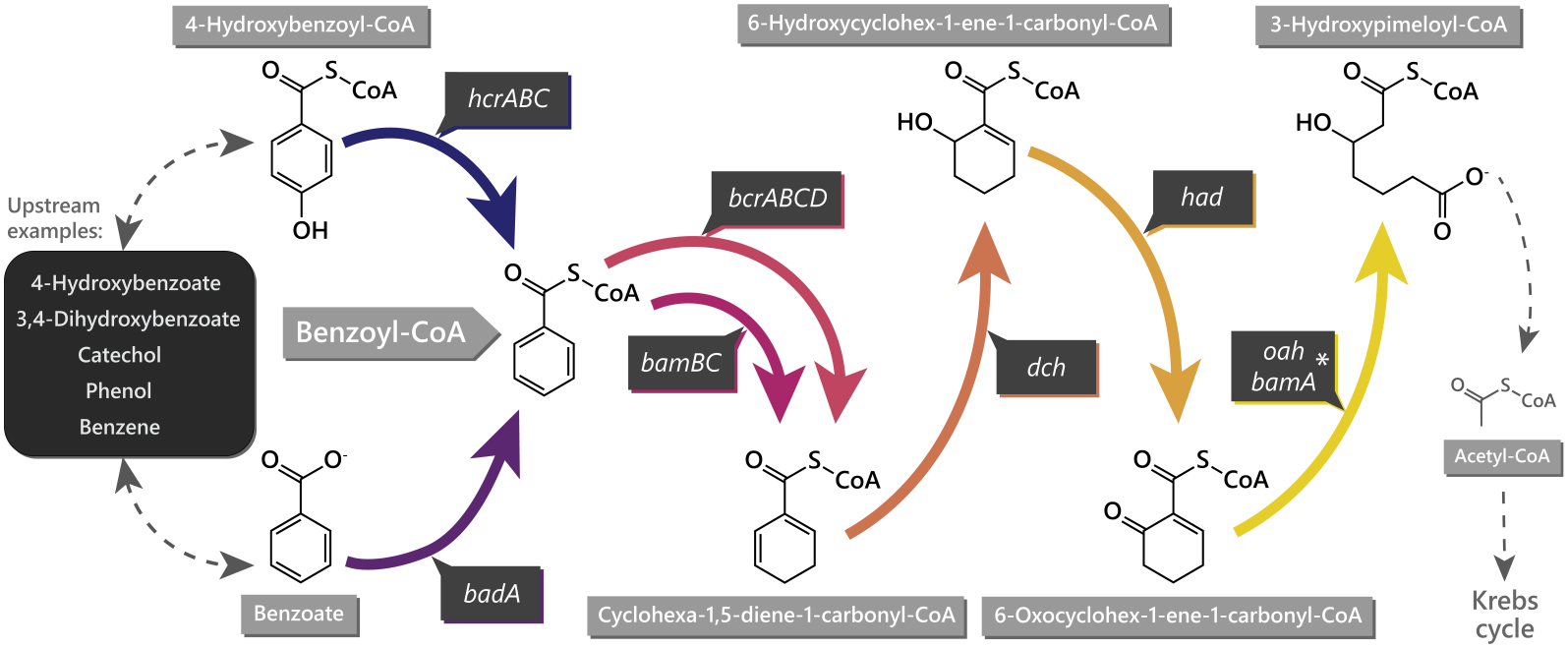
775 **Figure 1** Benzoyl-CoA (Coenzyme A) pathway genes and subunits, from 4-hydroxybenzoyl-CoA and
776 benzoate to 3-hydroxypimeloyl-CoA, and their presence in the selected contig bins (MAGs).
777 Benzoyl-CoA reductase genes for two enzyme systems are displayed: the class-I, ATP-dependant
778 *bcrABCD* (associated with facultative anaerobes), as well as the class-II, ATP-independent *bamB*
779 (associated with obligate anaerobes) (15, 48). See Supplementary Table S2 for number of
780 occurrences of each gene in the MAGs and metagenomes. The displayed benzoyl-CoA pathway
781 was adapted from www.genome.jp/kegg/.

782

783 **Figure 2** Phylogenetic relationships and environmental sources of selected Krumholzibacteriota,
784 Delphibacteria*, and Latescibacteria* MAGs. Maximum likelihood phylogeny of bin 3.1 and 17
785 representative MAGs using multilocus sequence analysis (MLSA; see Supplementary Figure S2 for
786 the equivalent FCB superphylum tree). Thirty other MAGs from the initial analysis were missing
787 the selected housekeeping genes and thus are not included here (Supplementary Table S4) (46, 49,
788 50, 59–62). Environmental sources are indicated for each MAG; circle size is proportional to the
789 number of times that environment applied within the same phylum.

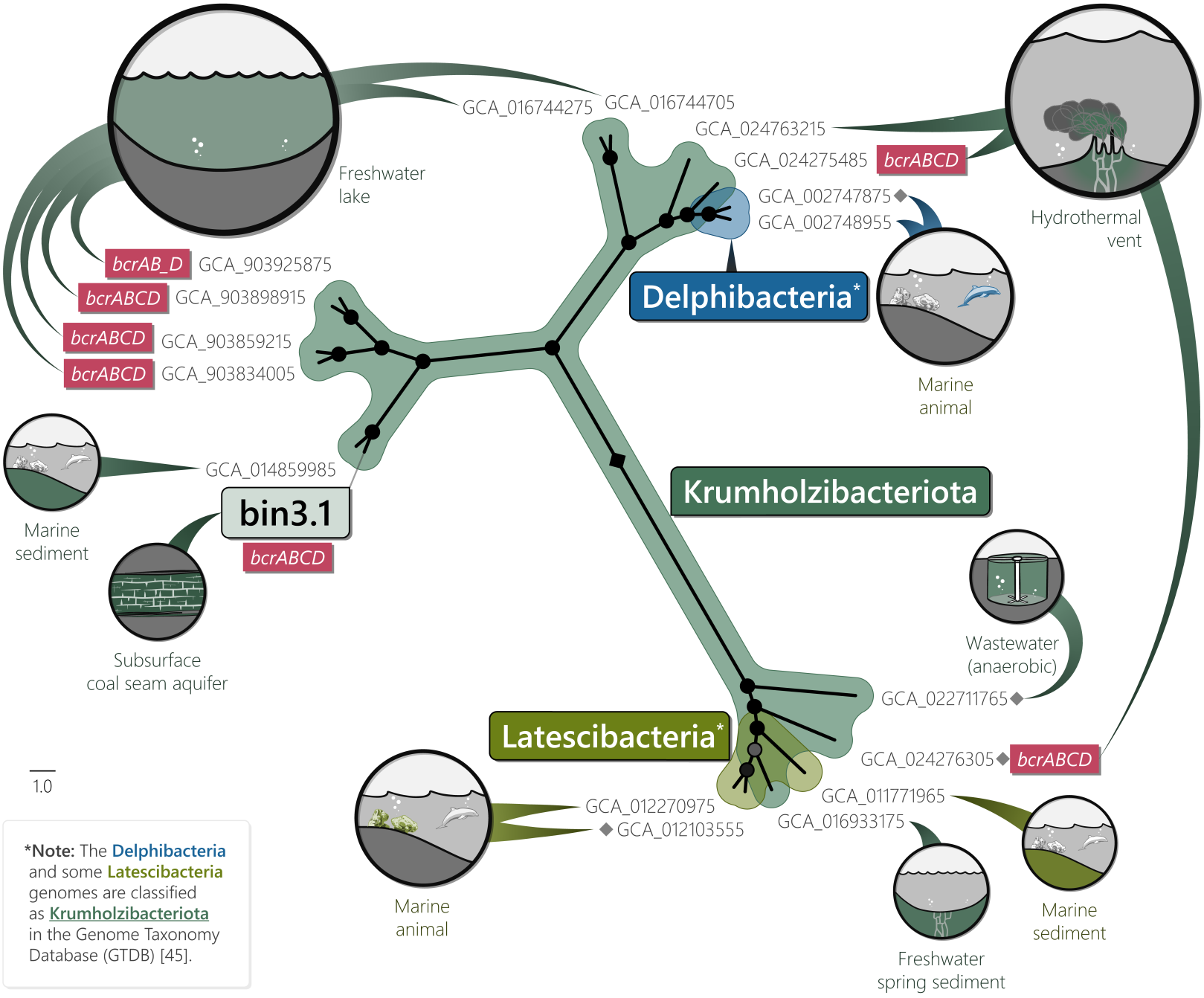
790

791 **Figure 3** Putative ecological functions within the coal seam microbiome. Conceptual model
792 displaying (A) the Syntrophaceae sp., Krumholzibacteriota sp. and *Syntrophorhabdus*
793 *aromaticivorans* within the Nance, Flowers-Goodale and Terret coal seams of the Powder River
794 Basin, USA, and (B) the *Syntrophorhabdus aromaticivorans* within the Bandanna Formation of the
795 Bowen Basin, Australia. The most abundant methanogen/s (according to *mcrA* and 16S rDNA data;
796 (27)) in the metagenome of each bin has been included in the model, however, these are not
797 necessarily the key partner methanogens for the aromatic-degrading taxa.



MAG name	Benzoyl-CoA pathway genes and subunits present in MAG										
Syntrophaceae sp. bin1.1						bcrC	bcrB				
Syntrophorhabdus aromaticivorans bin2.2	hcrA		hcrC	badA	bamB		bcrB		dch	had	oah_bamA
Syntrophorhabdus aromaticivorans bin2.6	hcrA	hcrB				bcrC	bcrB				
Krumholzibacteriota sp. bin3.1	hcrA					bcrC	bcrB	bcrA	bcrD	dch	had	oah_bamA

*Names for this gene varied; Prokka annotations of this gene were labelled *'bamA'*, while BlastKOALA annotations were labelled *'oah'*.



***Note:** The **Delphibacteria** and some **Latescibacteria** genomes are classified as **Krumholzibacteriota** in the Genome Taxonomy Database (GTDB) [45].

KEY:

GENES USED FOR MLSA

purA	Adenylosuccinate synthase	*efp*	Elongation factor P
atpB	ATP synthase subunit beta	*mdh*	Malate dehydrogenase
fusA	Elongation factor G	*recA*	Protein RecA

CONFIDENCE

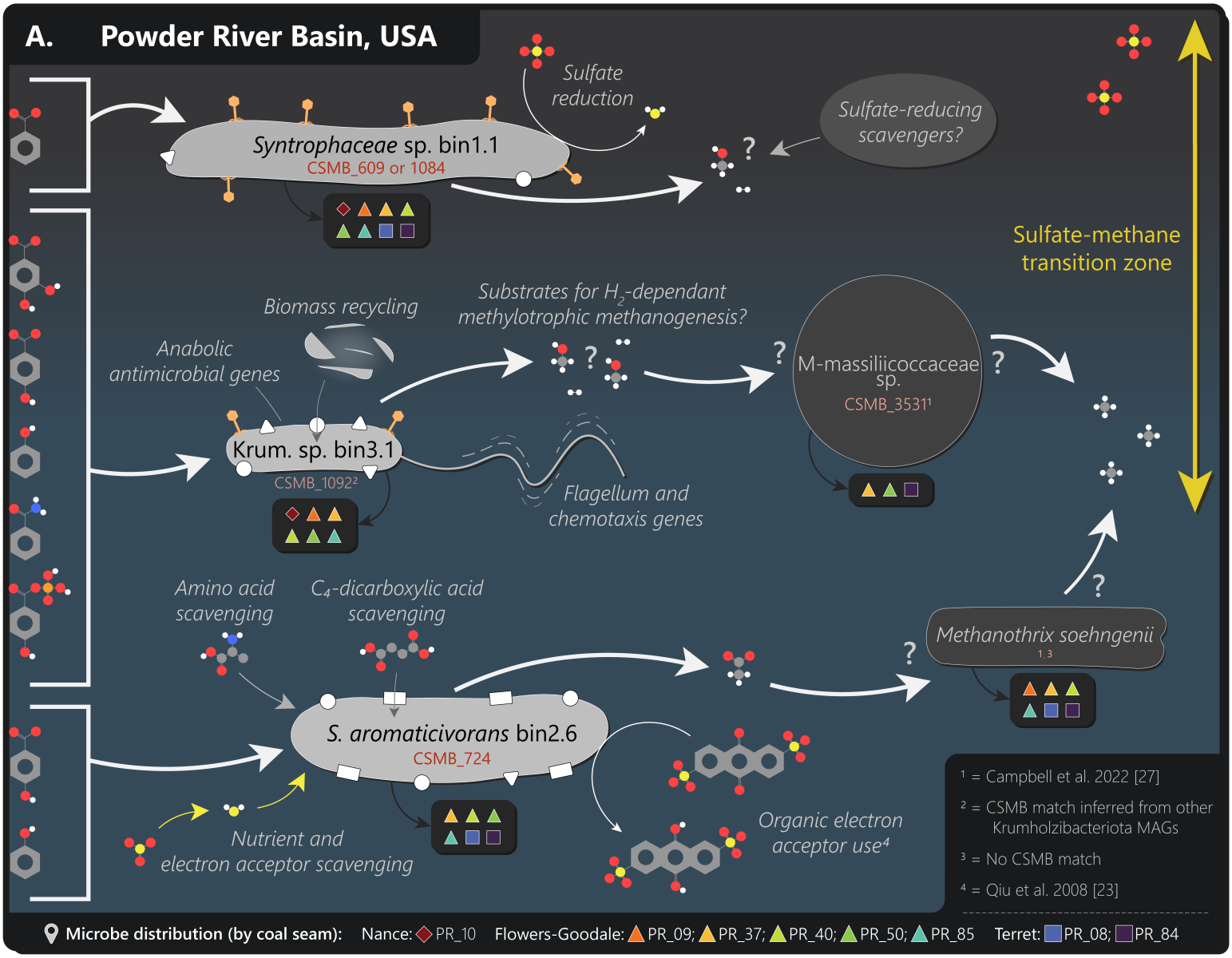
1.0 High
0.5
0.0 Low

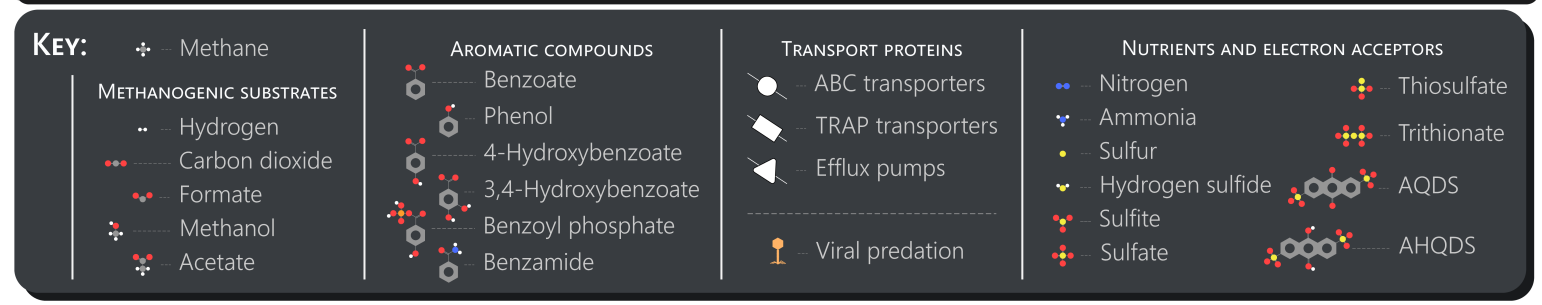
Bootstrap values from 100 iterations

KEY GENE PRESENT IN MAG

◆ 16S rDNA

bcrABCD Benzoyl-CoA reductase gene


A. Powder River Basin, USA
Syntrophaceae sp. bin1.1
CSMB_609 or 1084
Sulfate reduction
Sulfate-reducing scavengers?
Sulfate-methane transition zone
Biomass recycling
Substrates for H2-dependant methylotrophic methanogenesis?
Anabolic antimicrobial genes
M-massiliicoccaceae sp.
CSMB_3531¹
Krum. sp. bin3.1
CSMB_1092²
Flagellum and chemotaxis genes
Amino acid scavenging
C4-dicarboxylic acid scavenging
Methanothrix soehngenii
1, 3
S. aromaticivorans bin2.6
CSMB_724
Nutrient and electron acceptor scavenging
Organic electron acceptor use⁴
¹ = Campbell et al. 2022 [27]
² = CSMB match inferred from other Krumholzibacteriota MAGs
³ = No CSMB match
⁴ = Qiu et al. 2008 [23]
Microbe distribution (by coal seam): Nance: PR_10 Flowers-Goodale: PR_09; PR_37; PR_40; PR_50; PR_85 Terret: PR_08; PR_84


Key:
Methane
Methanogenic substrates
Hydrogen
Carbon dioxide
Formate
Methanol
Acetate
Aromatic compounds
Benzoate
Phenol
4-Hydroxybenzoate
3,4-Hydroxybenzoate
Benzoyl phosphate
Benzamide
Transport proteins
ABC transporters
TRAP transporters
Efflux pumps
Viral predation
Nutrients and electron acceptors
Nitrogen
Ammonia
Sulfur
Hydrogen sulfide
Sulfite
Sulfate
Thiosulfate
Trithionate
AQDS
AHQDS

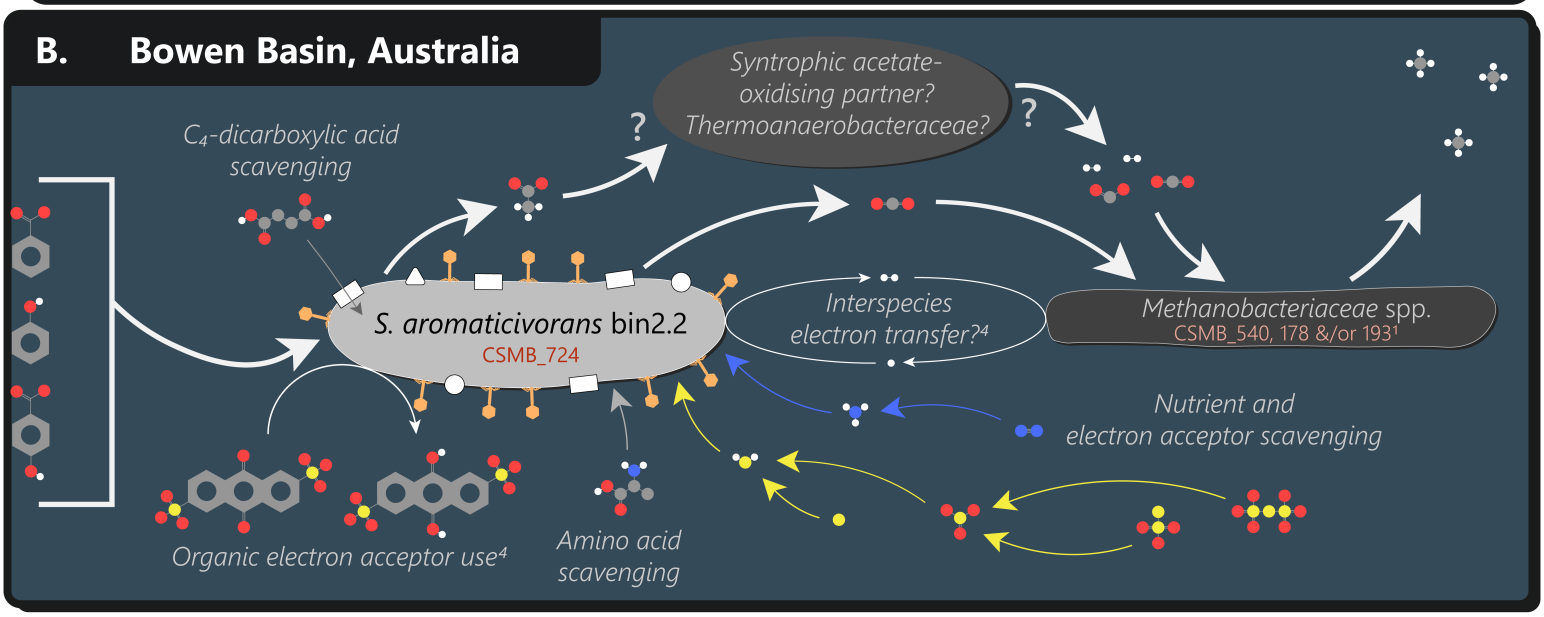

B. Bowen Basin, Australia
Syntrophic acetate-oxidising partner? Thermoanaerobacteraceae?
C4-dicarboxylic acid scavenging
S. aromaticivorans bin2.2
CSMB_724
Interspecies electron transfer?⁴
Methanobacteriaceae spp.
CSMB_540, 178 &/or 193¹
Nutrient and electron acceptor scavenging
Organic electron acceptor use⁴
Amino acid scavenging

Table 1: Bin statistics

Summary statistics category (Mbp)	**Syntrophaceae sp. bin1.1**	***S. aromaticivorans* bin2.2**	***S. aromaticivorans* bin2.6**	**Krumholzibacteriota bin3.1**
Total length	2.15	2.19	1.96	4.18
Total contigs	10	36	19	45
Mean contig length	0.22	0.06	0.1	0.09
Median contig length	0.16	0.05	0.08	0.07
N50	0.39	0.08	0.11	0.14
Maximum contig length	0.5	0.15	0.25	0.56
GC-content (%)	47.6	44.6	47.9	71.1
CheckM results (%)				
Completeness	59.35	77.42	54.62	94.51
Contamination	0.65	1.29	1.68	1.1

Table 2: The abundance of genes or genetic elements identified in the MAGs.

Genetic element (annotation tool) [30, 33-36]	**Syntrophaceae sp. bin 1.1**	***S. aromaticivorans* bin 2.2**	***S. aromaticivorans* bin 2.6**	**Krumholzibacteriota sp. bin 3.1**
Total *bcrABCD* subunits (Prokka)	2	1	3	4
Total *bcrABCD* subunits (BlastKOALA)	0	0	0	4
Annotation summary (Prokka)				
CDS/ORFs	2002	2121	1908	3342
rRNAs	0	0	0	0
tRNAs	33	27	23	52
tmRNAs	0	1	0	1
Carbohydrate-active enzyme class (dbCAN)				
Glycoside hydrolases	6	15	6	45
Glycosyltransferases	50	29	25	72
Polysaccharide lyases	0	0	0	0
Carbohydrate esterases	1	5	3	10
Auxillary activities	6	7	7	2
Carbohydrate-binding modules	5	5	3	14
CRISPR sequences (CRISPRCasFinder)				
CRISPR Loci	4	3	0	2
Number of spacers	18	30	0	7
Combined total length	136	104	0	51
Cas genes total	7	9	0	0
Membrane transporter family (TransportDB)				
ATP-binding cassette transporters	66	108	101	81
F-Type ATPase	13	16	22	16
Tripartite ATP-independent periplasmic transporters	0	32	36	1
Solute/sodium symporters	8	13	13	24
Major facilitator superfamily	6	15	12	13
Resistance-nodulation-division family transporters	9	3	10	20
other	69	82	60	111