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], Se Gong[c], David J. Midgley[c], Ian T. Paulsen[b]

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

13 ***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.

37 **Importance**

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

INTRODUCTION

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

463

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

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

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

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

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

548

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

557

558 **Acknowledgements**

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

563

564 **Competing interests statement**

565 The authors declare no competing interests.

566

567 **Data availability**

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

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.

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

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

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

600 9. Wang B, Yu Z, Zhang Y, Zhang H. 2019. Microbial communities from the Huaibei Coalfield
601 alter the physicochemical properties of coal in methanogenic bioconversion. Int J Coal Geol
602 202:85–94.

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

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

608 12. McKay LJ, Smith HJ, Barnhart EP, Schweitzer HD, Malmstrom RR, Goudeau D, Fields MW.
609 2021. Activity-based, genome-resolved metagenomics uncovers key populations and
610 pathways involved in subsurface conversions of coal to methane. ISME J 1–12.

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

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

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

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

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

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

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

631 20. Kühner S, Wöhlbrand L, Fritz I, Wruck W, Hultschig C, Hufnagel P, Kube M, Reinhardt R,
632 Rabus R. 2005. Substrate-Dependent Regulation of Anaerobic Degradation Pathways for
633 Toluene and Ethylbenzene in a Denitrifying Bacterium, Strain EbN1. J Bacteriol 187:1493–
634 1503.

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

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

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

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

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

650 26. Gray ND, Sherry A, Grant RJ, Rowan AK, Hubert CRJ, Callbeck CM, Aitken CM, Jones DM,
651 Adams JJ, Larter SR, Head IM. 2011. The quantitative significance of Syntrophaceae and
652 syntrophic partnerships in methanogenic degradation of crude oil alkanes. Environ Microbiol
653 13:2957–2975.

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

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

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

661 30. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–
662 2069.

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

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

668 33. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for
669 Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 428:726–
670 731.

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

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

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

681 37. Greenfield P, Tran-Dinh N, Midgley D. 2019. Kelpie: generating full-length 'amplicons' from
682 whole-metagenome datasets. PeerJ 6:e6174.

683 38. Apprill A, McNally S, Parsons R, Weber L. 2015. Minor revision to V4 region SSU rRNA 806R
684 gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol
685 75:129–137.

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

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

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

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

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

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

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

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

710 47. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz
711 F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A,
712 Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM,
713 Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ,
714 Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen
715 JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I,
716 Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Murat Eren A, Schriml L,
717 Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified

718 genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea.
719 8. Nat Biotechnol 35:725–731.

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

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

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

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

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

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

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

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

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

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

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

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

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

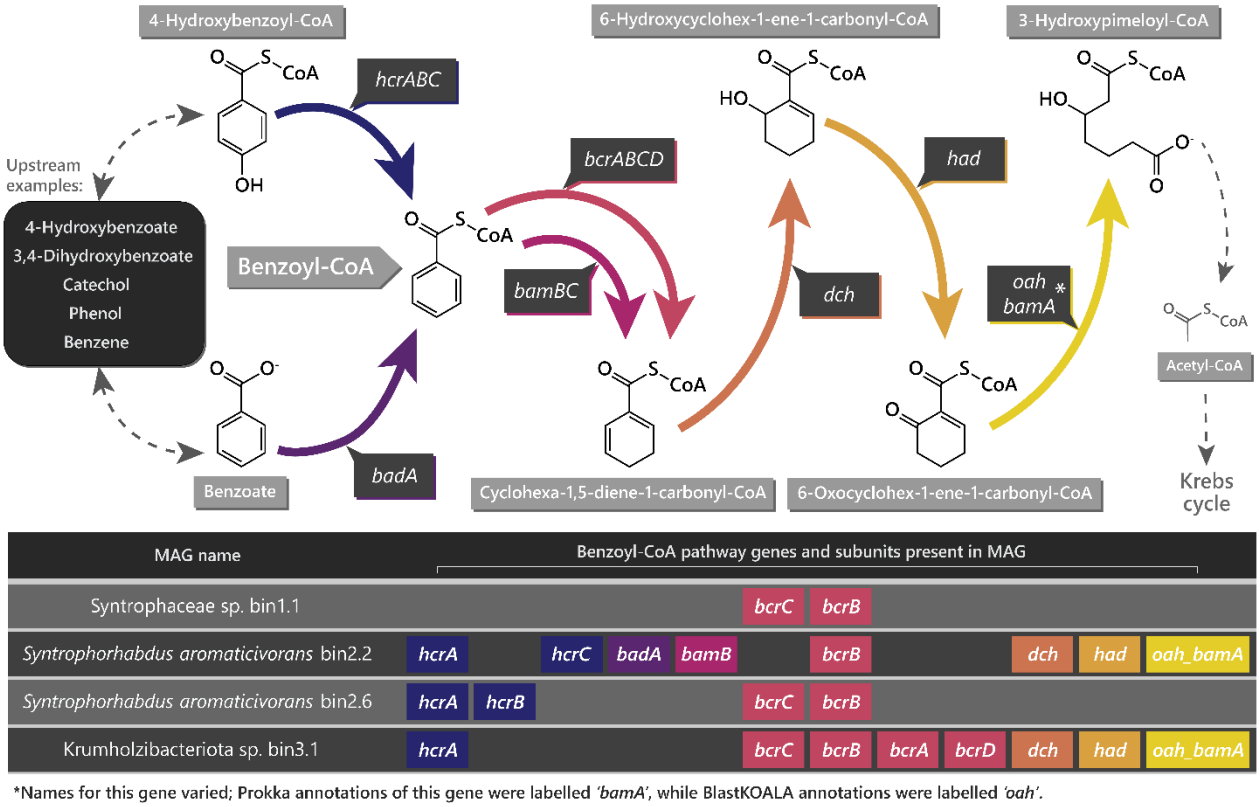
760 61. Robbins SJ, Singleton CM, Chan CX, Messer LF, Geers AU, Ying H, Baker A, Bell SC,
761 Morrow KM, Ragan MA, Miller DJ, Forêt S, Voolstra CR, Tyson GW, Bourne DG. 2019. A
762 genomic view of the reef-building coral Porites lutea and its microbial symbionts. 12. Nat
763 Microbiol 4:2090–2100.

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

767

768 **Figures**

769



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'*.

770

771 **Figure 1** Benzoyl-CoA (Coenzyme A) pathway genes and subunits, from 4-hydroxybenzoyl-CoA and

772 benzoate to 3-hydroxypimeloyl-CoA, and their presence in the selected contig bins (MAGs).

773 Benzoyl-CoA reductase genes for two enzyme systems are displayed: the class-I, ATP-dependant

774 *bcrABCD* (associated with facultative anaerobes), as well as the class-II, ATP-independent *bamB*

775 (associated with obligate anaerobes) (15, 48). See Supplementary Table S2 for number of

776 occurrences of each gene in the MAGs and metagenomes. The displayed benzoyl-CoA pathway

777 was adapted from www.genome.jp/kegg/.

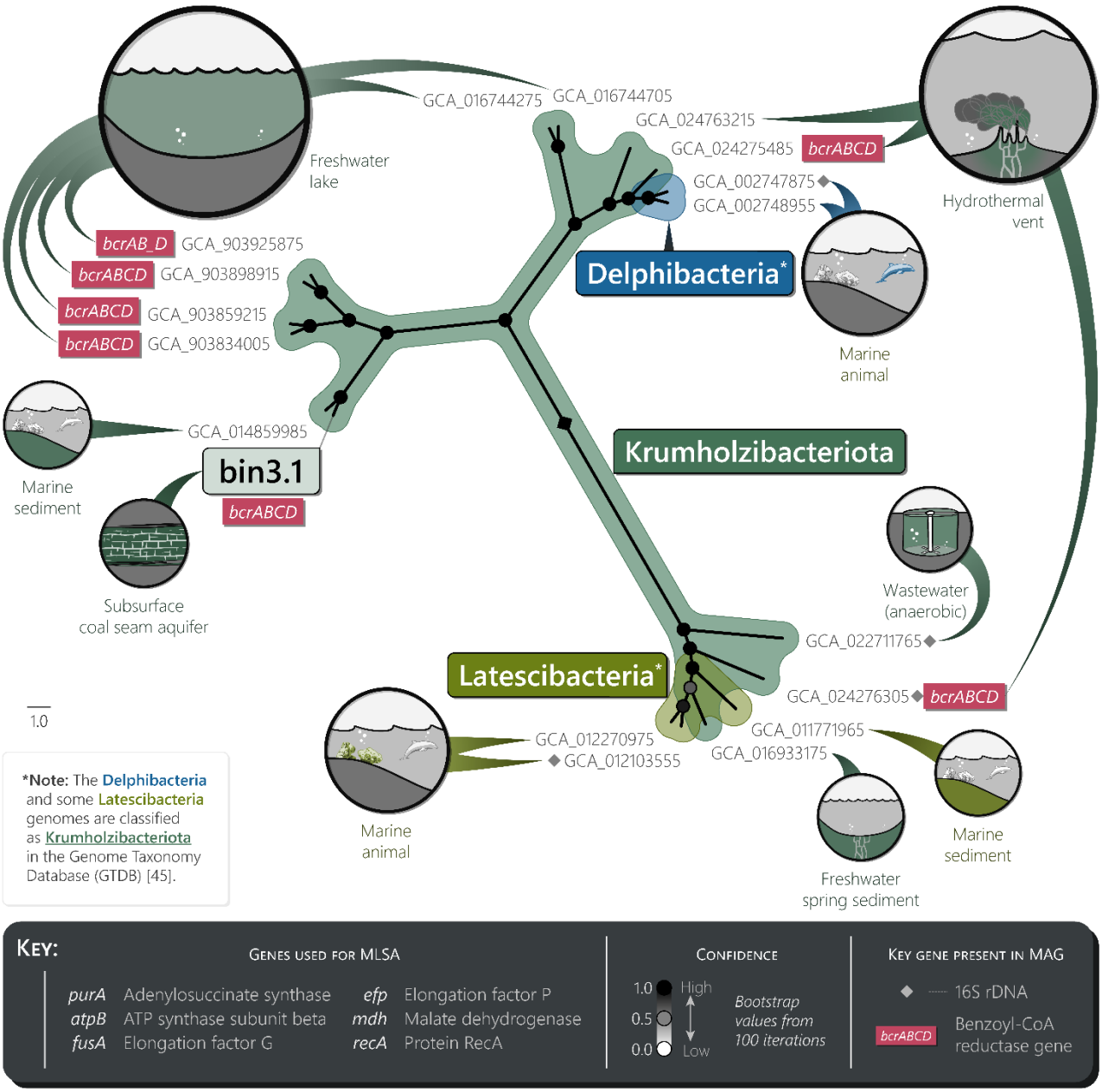


Figure 2 Phylogenetic relationships and environmental sources of selected Krumholzibacteriota, Delphibacteria*, and Latescibacteria* MAGs. Maximum likelihood phylogeny of bin 3.1 and 17 representative MAGs using multilocus sequence analysis (MLSA; see Supplementary Figure S2 for the equivalent FCB superphylum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

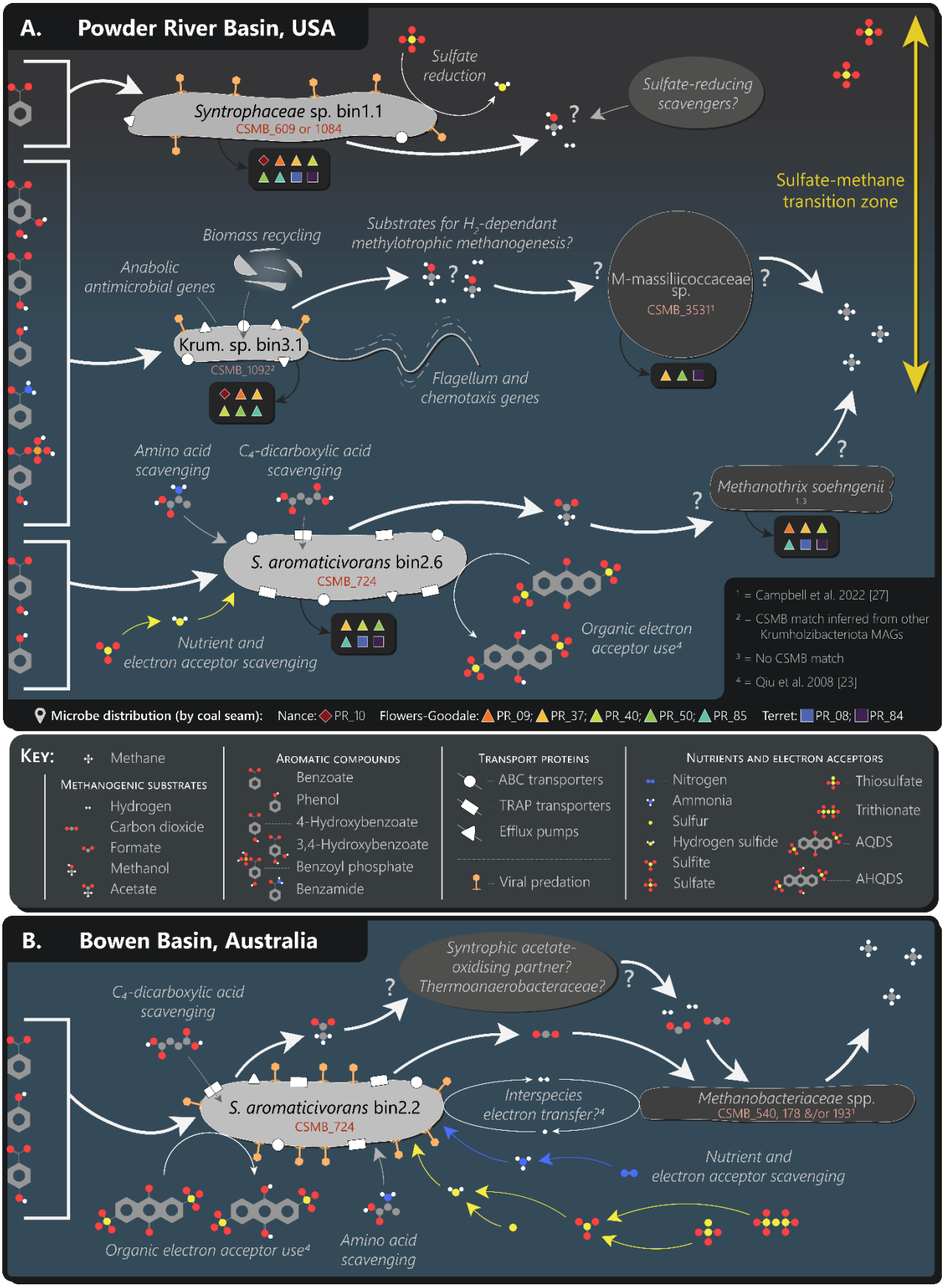


Figure 3 Putative ecological functions within the coal seam microbiome. Conceptual model displaying (A) the Syntrophaceae sp., Krumholzibacteriota sp. and *Syntrophorhabdus*

789 *aromaticivorans* within the Nance, Flowers-Goodale and Terret coal seams of the Powder River

790 Basin, USA, and (B) the *Syntrophorhabdus aromaticivorans* within the Bandanna Formation of the

791 Bowen Basin, Australia. The most abundant methanogen/s (according to *mcrA* and 16S rDNA data;

792 (27)) in the metagenome of each bin has been included in the model, however, these are not

793 necessarily the key partner methanogens for the aromatic-degrading taxa.

794 **Tables**

795

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 (%) [32]				
Completeness	59.35	77.42	54.62	94.51
Contamination	0.65	1.29	1.68	1.1

796

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
Auxiliary 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
TRAP transporters	0	32	36	1
Solute/sodium symporters	8	13	13	24
Major facilitator superfamily	6	15	12	13
RND family transporters	9	3	10	20
other	69	82	60	111

797