

OPEN

Genomic and phenotypic insights point to diverse ecological strategies by facultative anaerobes obtained from subsurface coal seams

Silas H. W. Vick^{1,2*}, Paul Greenfield², Sasha G. Tetu¹, David J. Midgley² & Ian T. Paulsen¹

Microbes in subsurface coal seams are responsible for the conversion of the organic matter in coal to methane, resulting in vast reserves of coal seam gas. This process is important from both environmental and economic perspectives as coal seam gas is rapidly becoming a popular fuel source worldwide and is a less carbon intensive fuel than coal. Despite the importance of this process, little is known about the roles of individual bacterial taxa in the microbial communities carrying out this process. Of particular interest is the role of members of the genus *Pseudomonas*, a typically aerobic taxa which is ubiquitous in coal seam microbial communities worldwide and which has been shown to be abundant at early time points in studies of ecological succession on coal. The current study performed aerobic isolations of coal seam microbial taxa generating ten facultative anaerobic isolates from three coal seam formation waters across eastern Australia. Subsequent genomic sequencing and phenotypic analysis revealed a range of ecological strategies and roles for these facultative anaerobes in biomass recycling, suggesting that this group of organisms is involved in the degradation of accumulated biomass in coal seams, funnelling nutrients back into the microbial communities degrading coal to methane.

Globally, coal represents a key fuel, accounting for almost ~25% of the world's energy consumption (International Energy Agency). The use of coal for power generation, however, is associated with significant environmental and health impacts^{1,2}. As a cleaner alternative, methane derived from coal, known as coal seam gas (CSG) or coal bed methane has become an increasingly important 'bridge' fuel for a global transition to renewables. With the discovery that significant portions of the world's CSG are produced through microbial degradation of the organic matter in coal to methane, has come an increasing interest in understanding the microbial communities involved in this process, with the aim of stimulating CSG production from coal reserves^{3,4}.

Efforts to understand the process of biological coal degradation have focussed primarily on either (1) understanding what nutrients to add to stimulate microbial communities to degrade organic matter in coal to methane, and/or; (2) characterising the microbial community compositions on coal and associated formation waters either *in situ* on or in laboratory grown microcosms (reviewed in Ritter *et al.*, 2015). These studies have typically used 16S rRNA gene based community profiling. This approach has been informative in characterising the members of coal-associated microbial communities, including which taxa appear to be ubiquitous in coal systems and those which appear to be dependent upon local conditions such as coal rank, depth and temperature. This type of research has enabled some generalisations to be made about the composition of these communities: all contain ubiquitous methanogenic archaea, a small number of very abundant bacterial taxa from the Proteobacteria or Actinobacteria; and less frequently Firmicutes, and a long tail of rare organisms from a range of phyla^{3,5,6}.

One key limitation of this approach is that assigned functions of any these uncultured organisms are obtained from related, cultured taxa. While this is useful for some groups such as the methanogenic archaea, where metabolic capabilities are largely constrained (see also^{7,8}), it is considerably less reliable in the majority of bacterial groups where metabolic functions are not always taxonomically conserved. In order to address this, some effort

¹Department of Molecular Sciences, Macquarie University, North Ryde, Australia. ²Commonwealth Scientific and Industrial Research Organisation (CSIRO), Sydney, Australia. *email: silas.vick@nmbu.no

to assign functions to coal seam microbes has been attempted through genome reconstructions from metagenomic sequences^{9–11}. This approach has generated some insights into the metabolisms of members of these environments, notably the discovery of a novel methanogenic taxon¹⁰, however, as many genes in the process are unknown the effectiveness of this method is limited. Finally, classical isolation into either axenic or gnotobiotic culture can be used to elucidate the metabolic roles of individual organisms. Unlike metagenomics-based approaches, isolation has the advantage of being able to experimentally test the metabolic and physiological characteristics of the isolated taxa through growth studies, though a key drawback remains that many microbial species are recalcitrant to isolation.

In 2016, Vick *et al.* identified the presence of several early coal-colonists in a Sydney Basin formation water. One of the highly abundant OTUs (Operational Taxonomic Unit) observed in this 16S rRNA amplicon sequence based study was taxonomically identified as a pseudomonad (OTU_9), and its presence during early colonisation suggested it may play some role in degrading organic matter in coal. Interestingly, *Pseudomonas* species have been previously identified from almost all surveys of coal seams conducted to date^{6,12–25} and pseudomonads isolated from coals in the past have been reported for the production of biosurfactants²⁶ and lignin degradation phenotypes²⁷. Additionally, a previous study has noted the prevalence of seemingly aerobic taxa and metabolisms in deep anoxic hydrocarbon environments²⁸. In order to try to assign functions to these coal-associated pseudomonads, and functionally related facultative aerobic taxa, a culturing effort was undertaken to bring facultatively aerobic organisms, including that corresponding to OTU_9, into axenic culture. This isolation was followed by genomic and phenotypic characterisation to uncover the metabolic and ecological roles played by these organisms in coal seams. As these organisms are facultative aerobes, the present study employed oxic conditions and a complex medium to obtain axenic cultures.

Results

Cellular and colony morphology. Ten bacterial isolates were obtained from three coal enrichment cultures derived from coal seam formation waters from the Sydney, Bowen and Surat coal basins of eastern Australia. The isolates all grew under oxic and anoxic conditions on either NA (Nutrient agar) or TSA (tryptone soy agar). Cellular and colony morphology were characterised for aerobically grown cells on rich complex media (NA and TSA) and are summarised in Fig. 1. Taxonomy of isolates was determined with isolates found to represent members of the Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria and Actinobacteria (Fig. 1).

16S rRNA Phylogenetic analysis. Phylogenetic analysis of coal seam isolates was performed using the full length 16S rRNA gene sequence. Isolates are compared to previously characterised/reported members of each represented genera (Fig. 2). Analysis showed that *Chelatococcus* sp. BOW-1 is closely related to *Chelatococcus caeni*, first isolated from a sludge biofilm wastewater reactor²⁹. Both *Citrobacter* sp. BOW-7 and *Citrobacter* sp. SUR-1 isolates had highly similar 16S rRNA sequences and shared *Citrobacter amalonicus* as their closest relative in the phylogenetic tree. *Pseudomonas* BOW-2 and *Pseudomonas* SYD-2 clustered within different species clades within the genus *Pseudomonas* tree. Bow-2 clustered within the *P. stutzeri* species clade, whereas SYD-2 clustered with a clade containing *P. linyingensis* and *P. sagittaria*. *Actinotalea* sp. SUR-A1, isolated from the Surat coal basin, showed the closest relationship to *A. fermentans*, a cellulolytic organism isolated from a municipal waste dumping site³⁰. *Marinobacter* sp. SUR-4 sits in a clade shared by close taxonomic relatives *M. alkaliphilus* and *M. shengliensis*, isolated from alkaline sub-seafloor mud and oil contaminated soil, respectively^{31,32}. *Vibrio* sp. SUR-5 has a very similar 16S rRNA sequence to *V. cincinnatiensis* and clusters closely to it in the *Vibrio* genus tree. *Tessaracoccus* sp. SUR-6 clusters within a clade containing *T. profundus* and *T. oleiagri*, previously isolated from the deep terrestrial subsurface³³ and oil contaminated soil, respectively³⁴. *Thauera* sp. SYD-3, isolated from formation waters sourced from the Sydney coal basin, does not cluster closely with any named species of *Thauera* but has very similar or identical 16S rRNA gene sequences to several unnamed environmental sequences from diverse industrial wastewater sources (Fig. 2).

Comparing isolate 16S rRNA sequences to those in the CSMB (Coal Seam MicroBiome) dataset⁶ identified the CSMB OTUs within which these isolates would be grouped. From this analysis *Chelatococcus* sp. BOW-1 matched CSMB_850, *Pseudomonas* sp. BOW-2 matched CSMB_20, *Pseudomonas* sp. SYD-2 matched CSMB_98, *Citrobacter* sp. BOW-7 and *Citrobacter* sp. SUR-1 both matched CSMB_148, *Actinotalea* sp. SUR-A1 matched CSMB_664, *Marinobacter* sp. SUR-4 matched CSMB_51, *Vibrio* sp. SUR-5 matched CSMB_562 and *Thauera* sp. SYD-3 matched CSMB_192.

Biolog carbon utilisation phenotyping. Biolog phenotype microarray plates PM1 and PM2 were used to characterise the catabolic capabilities of the coal seam bacterial isolates under aerobic metabolisms. Two strains, *Tessaracoccus* SUR-6 and *Pseudomonas* SYD-2, showed no observable growth or respiration in Biolog media under the conditions required for the assays, and were therefore not able to be assessed for carbon utilisation profiles. Other isolates grew on a range of different carbon substrates with the exception of *Actinotalea* sp. SUR-A1, which only showed growth on the carbolic acids: acetic acid and acetoacetic acid (Fig. 3).

Ecophysiological genome analyses. In order to better understand the lifestyles and metabolic strategies of the bacterial taxa isolated in the current study from coal seam formation waters, the genomes were examined for genes and genetic elements relevant to ecological lifestyle. Isolates were all isolated under oxic conditions indicating they are capable of using oxygen as a terminal electron acceptor and the requisite metabolic pathways/enzymes for this are observed in each genome. The genomes of both *Citrobacter* isolates (BOW-7 and SUR-1) and *Actinotalea* SUR-A1 encoded genes for the respiratory nitrate reductase (NarG) indicating that they are capable of using nitrate as a terminal electron acceptor also. *Actinotalea* SUR-A1 was also predicted to be able to use sulphate as a terminal electron acceptor. IMG/ER annotation also revealed that *Pseudomonas* BOW-2 was the sole isolate to carry the genes necessary for nitrogen fixation.

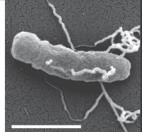
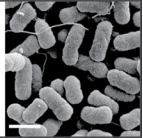

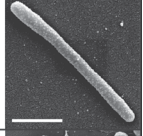

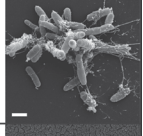
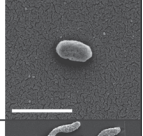



Isolate name	CSMB match	Colony morphology	Optimal medium	Cellular morphology
Chelatococcus sp. BOW-1	CSMB_850	Large, thick, raised, opaque white colonies with a glossy surface, smooth margin and mucous like appearance with round to amorphous shape.	NA	
Citrobacter sp. BOW-7	CSMB_148	Off white, round, raised colonies with a slightly glossy to matt surface with a smooth edge surrounded by a thin region of rough spreading morphology.	NA	
Pseudomonas sp. BOW-2	CSMB_20	Very diffuse, thin, light tan translucent colonies with a rough/bumpy surface and a rough and indistinct margin	NA	
Actinotalea sp. SUR-A1	CSMB_664	Small, translucent yellow, raised, circular colonies with a glossy surface and smooth margin	TSA	
Citrobacter sp. SUR-1	CSMB_148	Raised, white to off white, translucent colonies with a smooth margin and glossy surface.	NA	
Marinobacter sp. SUR-4	CSMB_51	Variable morphology. Generally raised, white to off white, opaque colonies with glossy surface and smooth margin. Some colonies show a moderate spreading morphology with a rough surface and diffuse margins. Overall more flattened.	NA	
Tessaracoccus sp. SUR-6	CSMB_3786	Small, translucent yellow, convex, circular colonies with a glossy surface and smooth margin. and overall more flattened.	NA	
Vibrio sp. SUR-5	CSMB_562	Opaque off white, round, raised colonies with a glossy surface and smooth margin.	NA	
Pseudomonas sp. SYD-2	CSMB_98	Off white to muddy yellow colonies spread very thinly with flat colony centres and a moderately strong spreading morphology with indistinct margins.	TSA	
Thauera sp. SYD-3	CSMB_192	Off white to yellow colonies spread very thinly with a denser raised centre. margins are rough and so thin as to be indistinct.	NA	

Figure 1. Colony and cellular morphology of isolates. Optimal growth media are nutrient agar (NA) and tryptone soy agar (TSA). Scale bars represent 1 μm .

Aromatic hydrocarbon degradation. Isolate genomes were also analysed with KEGG (Kyoto Encyclopedia of Genes and Genomes) to determine the presence of aromatic hydrocarbon degradation genes and pathways. All isolates except *Actinotalea* SUR-A1, *Vibrio* SUR-5 and *Tessaracoccus* SUR-6 showed a high number of genes in the benzoate catabolism category, a central intermediate in several aromatic degradation pathways. This included three complete pathways and two pathways missing only a single gene component. Additionally, *Marinobacter* SUR-4 and *Thauera* SYD-3 contained complete catabolism pathways for benzene and toluene, while *Pseudomonas* SYD-2 contained complete pathways for benzene only. Interestingly, all benzene catabolism pathways observed appeared to proceed through the intermediate catechol, indicative of aerobic catabolism rather than benzoyl-CoA which is indicative of anaerobic catabolism (Fig. 4).

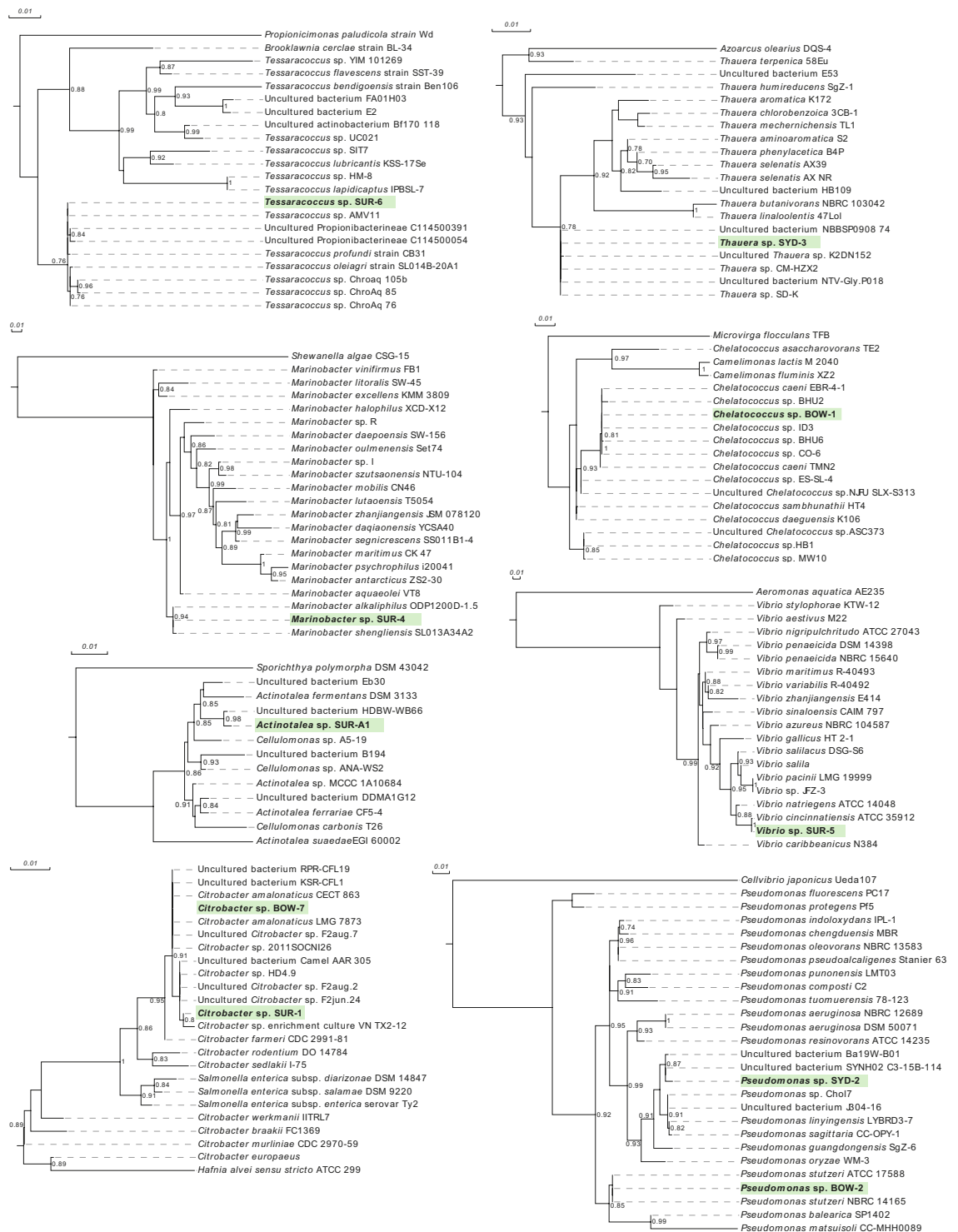


Figure 2. Phylogenetic tree analysis of coal seam isolates based on full length 16S rRNA gene sequences. Isolates from the current study are shown with bold font and highlighted in green. Numbers at branch junctions are a confidence index supporting branches as described in⁶⁷.

Carbohydrate-active enzymes. Predicted protein sequences from all isolate genomes were annotated for carbohydrate-active enzyme genes using dbCAN to assess the capabilities of isolates for utilising various carbohydrates³⁵. Genes which matched carbohydrate-active enzymes were then analysed with signalP to determine if the enzymes were likely excreted or transferred to the outer cell surface³⁶. A number of carbohydrate-active enzyme genes were identified in all of the isolate genomes (Supplementary Material 1). PCA (Principal Component Analysis) of these carbohydrate-active enzyme gene profiles identified two main clades containing similar profiles. These consisted of *Pseudomonas* BOW-2, *Pseudomonas* SYD-2, *Thauera* SYD-3 and *Marinobacter* SUR-4 in

Compound group	Total number of compounds in group	<i>Chelatococcus</i> sp. BOW-1	<i>Citrobacter</i> sp. BOW-7	<i>Pseudomonas</i> sp. BOW-2	<i>Actinotalea</i> sp. SUR-A1	<i>Citrobacter</i> sp. SUR-1	<i>Marinobacter</i> sp. SUR-4	<i>Vibrio</i> sp. SUR-5	<i>Thauera</i> sp. SYD-3
Alcohols	5	1	0	0	0	1	0	0	0
Amides	3	0	1	1	0	1	0	0	1
Amines	5	0	0	0	0	0	0	0	0
Amino acids	25	16	11	12	0	11	5	7	7
Carboxylic acids	59	39	29	27	2	29	14	12	22
Polymers	11	0	2	3	0	1	3	5	0
Dipeptides	4	3	4	1	0	4	0	0	0
Disaccharides	12	2	6	1	0	5	1	5	1
Glycosides	3	2	3	0	0	2	0	2	0
Methyl-saccharides	7	1	2	0	0	2	0	1	0
Miscellaneous	4	1	0	1	0	0	0	0	0
Monosaccharides	23	16	15	2	0	18	0	5	1
Nucleosides	5	0	5	1	0	5	0	3	0
Saccharide phosphates	4	1	4	0	0	4	0	0	0
Sugar alcohols	13	5	3	3	0	4	0	3	0
Surfactants	3	0	0	3	0	0	3	0	0
Tri and tetra saccharides	4	0	1	1	0	1	1	1	0

Figure 3. Heatmap showing the number of carbon substrates supporting growth in the various compound categories as measured by reduction of a tetrazolium dye on Biolog phenotype microarray plates 1 and 2. Isolates *Tessaracoccus* SUR-6 and *Pseudomonas* SYD-2 are not included as they appeared unable to grow under the conditions supplied by the Biolog assay.

Catabolic pathway	<i>Chelatococcus</i> sp. BOW-1	<i>Citrobacter</i> sp. BOW-7	<i>Pseudomonas</i> sp. BOW-2	<i>Actinotalea</i> sp. SUR-A1	<i>Citrobacter</i> sp. SUR-1	<i>Marinobacter</i> sp. SUR-4	<i>Vibrio</i> sp. SUR-5	<i>Tessaracoccus</i> sp. SUR-6	<i>Thauera</i> sp. SYD-3	<i>Pseudomonas</i> sp. SYD-2
Aromatic hydrocarbon degradation										
Benzoate	29	24	29	1	25	38	8	6	38	37
Benzene	0	0	0	0	0	2	0	0	2	2
Ethylbenzene	3	2	2	1	2	2	3	1	4	1
Toluene	3	4	3	0	4	7	0	1	13	7
Xylene	8	5	4	0	6	21	0	1	16	17
Naphthalene	3	4	5	2	5	2	4	1	3	1
Polycyclic aromatic hydrocarbons*	4	3	2	1	3	2	1	0	1	0
Xenobiotics biodegradation and metabolism	93	60	69	18	64	65	31	22	66	63

Figure 4. Number of genes in catabolic pathways for aromatic hydrocarbon degradation as determined by KEGG pathways analysis. Counts in enclosed boxes indicate a complete catabolic path and underlined counts indicate pathways containing all but one enzyme in the pathway. *With the exception of naphthalene.

the first group, characterised by having relatively fewer glycoside hydrolase (GH) and glycosyl transferase (GT) genes, and *Chelatococcus* BOW-1, *Tessaracoccus* SUR-6, *Citrobacter* BOW-7 and *Citrobacter* SUR-1 in the second, which had higher numbers of GH and GT genes. Outside of these two clusters *Vibrio* SUR-5 and *Actinotalea* SUR-A1 were outliers, having markedly different carbohydrate-active enzyme gene profiles (Fig. 5). SignalP analysis of the carbohydrate-active enzyme genes in the isolate genomes revealed that *Actinotalea* SUR-A1 encodes significant numbers of carbohydrate binding module (CBM), GH and S-layer homology domain (SLH) genes with accompanying signal peptide sequences, suggesting that this isolate can form a cellulosome. *Vibrio* SUR-5 also appeared to have a high number of GH and CBM genes, however it lacked the SLH genes characteristic of cellulosomes and a lower number of its genes contained signal peptides (Supplementary Material 1).

Membrane transporters. Genomes were annotated using TransportDB 2.0 to identify genes involved in membrane transport as these genes provide information on metabolism, through import transporters, and resistances to certain toxic compounds through export transporters³⁷. Isolate genomes contained between 263 and 850 genes involved in membrane transport from 90 transporter families (Supplementary Material 1).

PCA of the transporter profiles in the ten isolates showed that most isolates had a similar membrane transporter gene profile while *Chelatococcus* BOW-1 and both *Citrobacter* isolates (BOW-7 and SUR-1) were outliers having quite different profiles (Fig. 5). The difference in the *Chelatococcus* BOW-1 membrane transporter profile appeared to be driven by a number factors including: a larger number of membrane transport genes overall, a higher number of TRAP-T (Tripartite ATP-independent Periplasmic Transporters) family of transporters, involved in dicarboxylate uptake, and a larger number of ABC transporters involved in the uptake of amino acids, sulphate and polyamines. Differences in the *Citrobacter* transporter profiles appear to be driven by increased numbers of sugar transporters in the SSPTS (Sugar Specific Phosphotransferase System), GPH

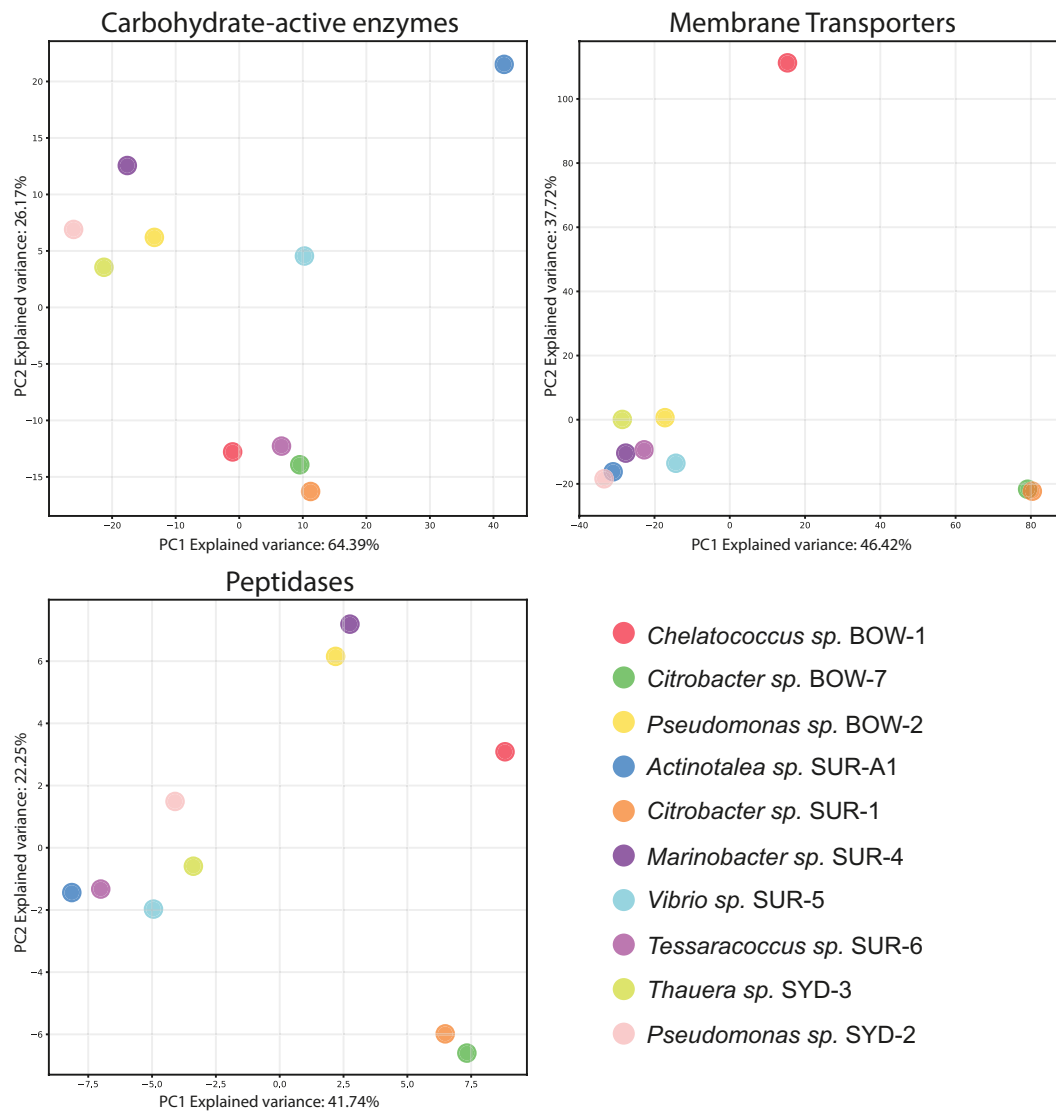


Figure 5. Principle components analysis (PCA) of gene abundances from genome annotations of carbohydrate-active enzyme, membrane transporter and peptidase genes. Abundances were measured as gene counts for the different classes of carbohydrate-active enzyme and peptidase genes (Online Resource 1) and gene counts for the different membrane transporter gene classes (or predicted substrates for ABC transporters) (Online Resource 1).

(Glycoside-pentoside-hexuronide) and ABC transporter families, and amino acid transporters in the HAAAP (Hydroxy/Aromatic Amino Acid Permease) and APC (amino acid-polyamine-organocation) transporter families (Supplementary Material 1).

Peptidases. Analysis using the MEROPS database identified a suite of putative peptidase genes in all isolate genomes³⁸. PCA of the peptidase gene family profiles for the isolates identified a group of isolates with similar peptidase profiles. This group, which included *Pseudomonas* SYD-2, *Thauera* SYD-3, *Vibrio* SUR-5, *Tessaracoccus* SUR-6 and *Actinotalea* SUR-A1, was characterised by a lower total number of putative peptidase genes. Other genomes with higher peptidase numbers separated out, included *Marinobacter* SUR-4 and *Pseudomonas* BOW-2 which group together, *Citrobacter* BOW-7 and SUR-1 also grouping and *Chelatococcus* BOW-1, with the highest number of peptidase genes, sitting as an outlier (Fig. 5) (Supplementary Material 1).

Secondary metabolites. The antiSMASH tool was used to annotate secondary metabolite gene clusters from the isolate genomes, of interest due to the roles secondary metabolites play in bacterial competition and biosurfactant production for substrate solubilisation³⁹. All isolates showed the presence of secondary metabolite gene clusters from a range of classes including non-ribosomal peptide synthetases (nrps), homoserine lactones (hserlactone) commonly used for quorum sensing, bacteriocins involved in competitive bacteriocidal interactions, ectoine used as a compatible solute and siderophores used for iron scavenging (Fig. 6).

	<i>Chelatococcus</i> sp. BOW-1	<i>Citrobacter</i> sp. BOW-7	<i>Pseudomonas</i> sp. BOW-2	<i>Actinotalea</i> sp. SUR-A1	<i>Citrobacter</i> sp. SUR-1	<i>Marinobacter</i> sp. SUR-4	<i>Vibrio</i> sp. SUR-5	<i>Tessaracoccus</i> sp. SUR-6	<i>Thauera</i> sp. SYD-3	<i>Pseudomonas</i> sp. SYD-2
Genetic element										
Insertion sequence elements										
Predicted number of IS elements	6	16	15	22	28	10	12	26	15	9
CRISPR sequences										
CRISPR Loci	2	0	1	3	0	0	4	0	1	1
Number of spacers	70	0	40	182	0	0	54	0	26	40
Secondary metabolite synthesis genes										
nrps	2	1	0	0	1	1	1	0	0	0
hserlactone	0	1	1	0	1	0	1	0	1	0
transatpks	0	0	0	0	0	0	0	0	0	0
bacteriocin	1	0	1	0	0	0	0	0	1	0
other	2	0	0	0	0	0	0	0	0	0
terpene	1	0	1	1	0	0	0	2	1	0
ectoine	0	0	1	0	0	2	1	0	0	1
t3pks	1	0	0	1	0	0	0	0	0	0
siderophore	0	0	1	0	0	1	0	0	0	0
thiopeptide	0	1	0	0	1	0	0	0	0	0
arylpolyene	0	1	0	0	1	0	0	0	0	0

Figure 6. Heatmap showing the abundance of genes or genetic elements observed in the genomes of coal isolates. nrps (non-ribosomal peptide synthetase), hserlactone (homoserine lactone), transatpks (trans-AT polyketide synthase), t3pks (type 3 polyketide synthase).

Insertion sequences and CRISPRs. In addition to genes produced by bacterial species for catabolism and competition the presence and abundance of insertion sequences and CRISPRs (clustered regularly interspaced short palindromic repeats) were investigated to assess horizontal gene transfer and viral predation on the bacterial isolates. Analysis using the ISSaga tool identified numerous insertion sequence elements in all isolate genomes with *Citrobacter* SUR-1 and *Tessaracoccus* SUR-6 having notably high numbers while *Chelatococcus* BOW-1 had notably less than other genomes (Fig. 6)⁴⁰. Genome analysis with CRISPRFinder (<https://crispr.i2bc.paris-saclay.fr/Server/>) identified CRISPR sequences in six of the isolate genomes while four didn't appear to have any CRISPR elements (Fig. 6)⁴¹. It is noteworthy that the size of CRISPR spacer arrays was quite variable between isolates while the number of CRISPR loci was less variable (Fig. 6).

Discussion

Pseudomonas represents a ubiquitous bacterial genus observed in almost all surveyed coal seam microbial communities. Despite this, the ecological and metabolic roles played by these pseudomonads and related facultatively aerobic taxa in coal seams remains poorly understood. The current study aimed to isolate representatives of these pseudomonads from multiple coal seam environments and to characterise their capabilities both genomically and phenotypically, in order to better understand their roles in this environment. Aerobic isolation from three coal seam formation waters on peptide rich media successfully generated two pseudomonad isolates from different species clades. This isolation effort also resulted in the generation of eight additional bacterial isolates from the actinobacterial genera *Tessaracoccus* and *Actinotalea* and the proteobacterial genera *Thauera*, *Marinobacter*, *Chelatococcus*, *Citrobacter* and *Vibrio*. As a phylogenetically diverse range of bacterial isolates were obtained it was decided to pursue genomic and phenotypic characterisation of all bacterial isolates to investigate putative ecological functions for this set of facultatively anaerobic, heterotrophic bacteria. The cellular morphology of all isolates was examined through scanning electron microscopy, carbon metabolisms characterised through Biolog phenotype screening and genomes sequenced and annotated to examine ecophysiological relevant genes and genetic elements.

The roles of microbes in coal seam microbial communities are typically thought of in terms of their carbon catabolism, where the organic components of coal are sequentially degraded to methane in a linear fashion by a succession of microbial species. The ecological functions in a microbial community, however, are more complex and include a range of community functions including primary and secondary degradation, predation, scavenging and biomass recycling, amongst others. In fields of macro-ecology these ecological roles are often interrogated in terms of life strategy categories such as the competitive, stress tolerant and ruderal (CSR) classification used to classify the ecological strategies of plants during succession⁴². These lifestyle classifications and the catabolisms and phenotypes associated with them are, in microbial systems, dependent upon the type of environment and how the energy and nutrient resources are introduced into the environment⁴³. The coal seam constitutes an endogenous heterotrophic environment in that it is a closed system without regular microbial, nutrient or energy inputs from external sources⁴³. In this type of environment initial colonising ruderal species are heterotrophic organisms which modify the physicochemistry of the environment through their metabolic processes until the environment is modified to a point where slower growing, stress tolerating organisms or those with specialised metabolic capabilities are able to compete and proliferate⁴³.

Examination of the genotypic and phenotypic analysis of isolate from the current study suggests the presence of three broad ecological life strategies. The first resembles a ruderal life strategy and is shared by *Citrobacter* SUR-1, *Citrobacter* BOW-7 and *Pseudomonas* BOW-2, characterised by rapid growth on common, labile biomolecules with few genes associated with specialised metabolisms. The second life strategy appeared to be an opportunistrophic lifestyle as described by Singer *et al.*⁴⁴. This was characterised by genes associated with diverse

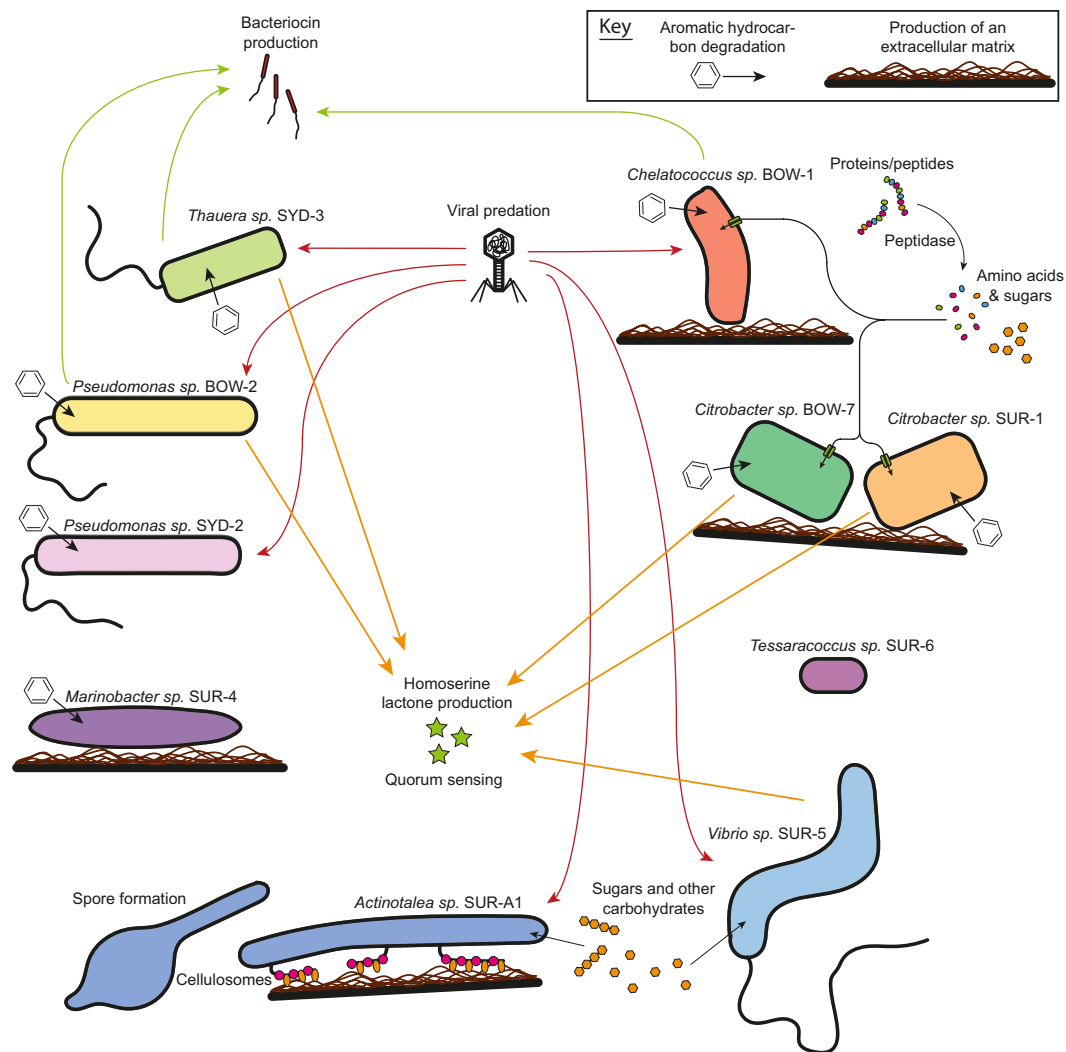


Figure 7. Conceptual model of the ecologically relevant capabilities of bacterial isolates determined through genomic and phenotypic analysis.

metabolisms suitable for survival across a wide range of different environments and was observed in *Thauera* SYD-3, *Marinobacter* SUR-4 and *Pseudomonas* SYD-2. The third life strategy group represents the specialists. This group is broader in function and includes different specialist strategies but shares the common trait of having narrower metabolisms with a greater genomic investment in a targeted catabolic process. Isolates which could be considered to have specialist life strategies include *Chelatococcus* BOW-1 and *Actinotalea* SUR-A1. In addition to isolates that had characteristics aligning with known ecological strategies, two isolates (*Tassaracoccus* SUR-6 and *Vibrio* SUR-5) had life strategies that did not appear to conform to known models based on the genotypic and phenotypic characterisation performed here (Fig. 7). From our analyses of the isolates and their genomes described here, it is not believed that any of these isolates partakes in a lifestyle or strategies centred around direct biodegradation of coal compounds *in situ*.

The first life strategy group constituted isolates displaying a ruderal life strategy. Comparison of the isolates from the current study to OTUs observed in a previous study of succession in coal seams, using the CSMB OTU reference set⁵, shows that *Citrobacter* SUR-1, *Citrobacter* BOW-7 and *Pseudomonas* BOW-2 matched to OTUs which, in a previous study, showed initial rapid proliferation, quickly dominating the communities, before declining in abundance at later time-points⁴⁵. This pattern of growth is characteristic of ruderals which rapidly proliferate on the most labile compounds available in an environment before being replaced by microbes with more specialised metabolisms once this labile resource is exhausted. In a previous study, these OTUs matching *Citrobacter* SUR-1, *Citrobacter* BOW-7 and *Pseudomonas* BOW-2 have also showed a very high level of variation in relative abundance across replicate microcosms compared to other taxa in the community. This high variation in abundance is due to competition by different ruderal taxa in other replicates as success as a ruderal is often decided by small stochastic affects during inoculation⁴⁶. The phenotypic and genotypic features presented in the current study support this hypothesis of a ruderal lifestyle and suggest catabolic niches supporting them. These putatively ruderal isolates all showed growth on a wide variety of amino acids, peptides and carboxylic acids

(Fig. 3). The genomes of these organisms also contained large numbers of peptidase genes and membrane transport genes for amino acids, sugars and carboxylic acid transporters (Supplementary Material 1). This collection of genes and phenotypes suggests that growth on cellular debris, particularly proteinaceous material, may drive the initial ruderal response in coal seam microbial communities. This cellular debris likely originates from cell death due to stress associated with the inoculation event but in native coal environments cellular debris could be liberated by chemical or physical perturbations to the coal seams or the ingress of meteoric waters.

The second life strategy type observed amongst our isolates was an opportunistrophic lifestyle, *sensu* Singer *et al.*⁴⁴. Isolates conforming to this grouping (*Thauera* SYD-3, *Marinobacter* SUR-4 and *Pseudomonas* SYD-2) have been observed in a previous study of succession in coal seams⁴⁵. In this previous study these opportunistrophic organisms remained at low abundances during the succession process⁴⁵. These were also the only isolates to have considerable numbers of genes involved in aromatic hydrocarbon degradation (Fig. 4). Interestingly, in all of these organisms the pathways proceed through the metabolic intermediate catechol, an intermediate indicative of aerobic catabolism⁴⁷. As all the eastern Australian coal seams sampled in the current study are highly anoxic and chemically reduced it is unlikely that these isolates are using these pathways for aromatics degradation in the coal seam under the conditions from which they were isolated. It should be noted, however, that these aromatic molecules are common in these coal seams and these may become available in the more oxic regions of aquifers into which coal seam formation waters may migrate. This observation of genes involved in aerobic aromatics catabolism, along with the facultative anaerobic growth patterns of these isolates, suggests that many of these organisms have diverse metabolic potentials allowing for persistence and growth across a large range of environments, scavenging diverse carbon compounds when environmental conditions allow it. A lifestyle in line with that described for the prototypical opportunistroph *Marinobacter aquaeolei*^{44,48}. It is noteworthy that two pseudomonads with distinctly different metabolic and ecological lifestyles were observed in this study. This likely reflects the wide metabolic and lifestyle diversity in the genus which spans parasitism to mutualism^{49–51} and free-living lifestyles⁵².

Amongst the isolates to be aligned with the specialist life strategy group is *Chelatococcus* BOW-1. *Chelatococcus* BOW-1 displays very high number of diverse membrane transport genes, particularly ABC transporters involved in the uptake of amino acids, sugars, and polyamines. This factor, coupled with its growth on a wide range of carbon sources and high peptidase gene number but non-ruderal succession observed in previous studies and paucity of genes involved in aromatic hydrocarbon degradation suggests that this organism likely specialises in scavenging low quantities of diverse cellular material for nutrients and energy. This lifestyle would allow this organism to survive in diverse environments so long as other organisms were present to provide it with biomass to recycle. Similarly, *Actinotalea* SUR-A1 showed a distinctive ecophysiological profile setting it apart from the other isolates examined. *Actinotalea* SUR-A1, in contrast to *Chelatococcus* BOW-1 showed a reduced metabolic diversity compared to other isolates having a low number of transporters and growing on very few carbon compounds. *Actinotalea* SUR-A1 did, however, display a large number of carbohydrate active enzymes including all components involved in the formation of a cellulosome⁵³. This suggests *Actinotalea* SUR-A1 may be responsible for complex polysaccharide degradation in the coal seam, presumably from biofilm or extra-cellular material as has been previously reported to be produced by coal seam microbial communities^{24,45}. The genome of *Actinotalea* SUR-A1 also displayed a number of traits involved in cellular defence including a large number of CRISPR elements and a large number of daunorubicin and other multidrug efflux pumps often associated with antibiotic production in Actinobacteria⁵⁴. Together with the observed production of spores by this isolate (Fig. 1), these results indicate that *Actinotalea* SUR-A1 follows a stress tolerator lifestyle, relying on a specialised metabolism based around carbohydrate catabolism rather than a broader opportunistrophic lifestyle, although sharing metabolism of common biomolecules as well as the ability for environmental persistence through sporulation.

A small number of isolates from the current study (*Tessaracoccus* SUR-6 and *Vibrio* SUR-5) did not appear to conform to any clear life strategy models based on the genotypic and phenotypic characterisations performed in the current study. It may be that these organisms are also involved in biomass recycling albeit on a smaller range of cellular biomolecules than the other scavenger and opportunistroph organisms. Alternatively, it may be that these organisms are involved in novel or unexamined catabolic niche processes in coal or other subsurface waters.

From a broader perspective, some notable differences were observed in the numbers of insertion sequence and CRISPR elements between the different isolate genomes. One interesting observation was the large difference in the number of insertion sequence elements observed in the closely related *Citrobacter* isolates BOW-7 and SUR-1 (Fig. 6). This increase in insertion sequence number in *Citrobacter* sp. SUR-1 may indicate recent adaptations as insertion sequences have been implicated in genomic shuffling and reorganisation⁵⁵. CRISPR elements have previously been reported as unexpectedly common in a metagenomic exploration of a deep fractured shale environment and have been suggested to be a common feature of the terrestrial subsurface⁵⁶. The current study supports these findings with observations of CRISPR sequences in six out of ten isolate genomes examined. If this high abundance of viral particles in subsurface environments is indeed the case, then viral predation may be a contributing factor in controlling biomass turnover, cell density and species distributions in subsurface coal environments, as it has been shown to be in other microbial environments⁵⁷. This could have implications for subsequent methane production by methanogenic coal seam communities and warrants further attention as a potential limiting factor for methane production rates from coal.

Much of the literature in the field of coal seam microbiology is focused on improving coal seam gas production rates through nutrient and microbial amendments³. This has meant that studies have primarily focused on the microbial compositions of formation waters and coal solids, considering less the continuous nature of aquifers and whether those organisms that live within them require flexibility as they may find themselves in different physical and chemical environments along the course of the aquifer. Further, it would be valuable to be able to distinguish those taxa that are subsurface generalists from those that are coal seam specialists. A concept that has recently been explored by Barnhart and colleagues⁵⁸. Based on their metabolisms as facultative anaerobes and their genomic potential for recycling of cellular biomass, it is likely that the isolates from the current study belong

to a group of subsurface generalised organisms rather than coal seam specialists. It should be noted, however, that the current observations are made from genome sequences and expression of these genes under *in situ* conditions remains undetermined. Examination of expression of these genes under anaerobic conditions mimicking the coal seam environment may shed further light on the lifestyles of these microbes.

In summary, bacteria in coal seams are thought to carry out the majority of steps involved in coal degradation, with methanogenic archaea responsible for only the very final stages of conversion of acetate, CO₂ or other simple methylated compounds to methane³. These initial stages of degradation, carried out by the bacteria taxa in the community, are thought to be the rate limiting step in the conversion of coal to methane and so identifying and characterising the bacterial taxa responsible for these processes is important for understanding the microbial coal to methane conversion process⁴. The current study reports on the successful isolation of two pseudomonads and eight other bacterial isolates, from eastern Australian coal seam formation waters. Phenotypic and genomic characterisation of these bacterial facultative anaerobes from the coal seam identified metabolisms supporting ruderal, opportunitrophic and specialist lifestyles centred around recycling carbohydrates and proteinaceous material in the coal seam. This study represents an initial step in assigning ecological and metabolic functions to the bacterial taxa observed in coal seams and utilised a protein rich isolation media and aerobic isolation conditions to target members of the genus *Pseudomonas*, which have been identified as ubiquitous members of coal seam microbial communities.

Future work should focus on isolation strategies to target other metabolisms involved in the conversion of organic matter in the coal to methane, these include anaerobic hydrocarbon degraders to identify taxa responsible for breakdown of hydrocarbons in coal or sulphate reducing bacteria which constitute a large proportion of coal seam microbial communities in eastern Australian coal seam formation waters despite the absence of elemental sulphur and sulphate from these environments and for whom a definitive catabolic role is not yet known.

Methods

Sources of isolates. Formation water for enrichment cultures was sampled anoxically from the wellhead of CSG wells in the Bowen Basin (QLD, Australia), Sydney Basin (NSW, Australia) and Surat Basin (QLD, Australia). 250 mg/L of Na₂S, 200 mg/L of cysteine HCl and resazurin to a final concentration of 0.0001% were added to the formation water immediately upon collection. The formation water was transported to the laboratory and incubated anoxically on crushed coal from matching formations at 30 °C (Bowen and Sydney) and 40 °C (Surat) for several months prior to use as an inoculum. Chemical analysis of formation water was performed by the ALS Environmental (Sydney Australia).

Actively methanogenic enrichment cultures were generated from these Bowen, Sydney and Surat Basin formation waters. To mimic the coal seam environment as closely as possible sampled waters were used as a culture medium with 25 mg/L NH₄Cl, 400 mg/L K₂HPO₄·2H₂O added to increase growth rates. Gamma sterile (8 kGy) crushed coal (<1 mm) collected from the same coal basins as the corresponding formation waters was used as the sole carbon source. An anoxic headspace consisting of 95% Argon: 5% H₂ was used in the butyl rubber stoppered vials and these enrichment cultures were incubated for several months at temperatures matching those measured in the produced formation waters (40 °C for the Bowen and Sydney Basin and 30 °C for the Surat Basin).

Isolation of bacteria. In order to isolate facultative aerobic taxa, 50 µL of enrichment culture was plated onto multiple nutrient agar (NA) (Lab-lemco powder 1 g/L, yeast extract 2 g/L, peptone 5 g/L, NaCl 5 g/L and agar, 15 g/L) plates which were then incubated under oxic conditions at temperatures matching those of the corresponding enrichment cultures until colonies were apparent. Colonies were then picked based on morphological differences and streaked multiple times on NA and tryptone soy agar (TSA) (Pancreatic digest of casein 15 g/L, enzymatic digest of soya bean 5 g/L, NaCl 5 g/L and agar 15 g/L) plates to produce axenic cultures. Isolates were screened to identify novel isolates by colony PCR amplification of the 16S rRNA gene using Bioline MyTaq following the manufacturer's instructions. Amplicons were purified (QiaQuick PCR purification kit) and sequenced by the Ramaciotti Centre for Genomics in both the forward and reverse directions (University of New South Wales, Sydney, Australia).

Visualisation of cellular morphology. To examine cell morphology isolates were grown in 10 ml of nutrient broth (same as for NA excluding agar) overnight at 40 °C for Bowen and Sydney basin isolates and 30 °C for Surat basin isolates with shaking (140 rpm) or until visible turbidity was observed. 200 µl of broth culture was then spotted onto poly-L-lysine (mol. wt. 150,000–300,000) coated glass coverslips for 5 min. Broth culture was then removed and the coverslips incubated in 3% glutaraldehyde in 0.01 M phosphate buffer (pH 7.4) for 24 hours at room temperature before being washed three times, each for 10 min with 0.01 M phosphate buffer. The coverslips were then incubated for 10 min in each of 20%, 50%, 70%, 80% and 90% ethanol solutions, twice with 100% ethanol before final dehydration with a Leica EM CPD300 critical point dryer. Coverslips were mounted on metal stubs, coated with 20 nm of gold with an Emitech K550 gold sputter coater unit and imaged using a JEOL JSM-7100F field emission scanning electron microscope operating at 5 kV accelerating voltage.

DNA extraction and whole genome sequencing. Isolates were grown as for visualisation. Cells were pelleted by centrifugation and genomic DNA was isolated using either the Isolate II genomic DNA kit (Bioline), PowerSoil® DNA Isolation Kit (MO BIO Laboratories) or FastDNA® KIT (MP Biomedicals). Library preparation and whole genome sequencing were performed at the Ramaciotti Centre for Genomics (University of New South Wales, Sydney, Australia) using the Nextera XT sample preparation kit (Illumina) and an Illumina Miseq platform to generate 250 bp paired-end reads. Sequence reads were error corrected using Blue (<http://bioinformatics.csiro.au/blue>) before *de novo* assembly with Velvet 1.2.10^{59,60}. Assemblies were submitted to the Genomes Online Database (GOLD) (<https://gold.jgi.doe.gov/>) and annotated using the IMG-ER pipeline⁶¹. Genome assemblies

are available under the project accession ID's: Gp0191703 (*Chelatococcus* sp. BOW-1), Gp0191702 (*Pseudomonas* sp. BOW-2), Gp0191704 (*Citrobacter* sp. BOW-7), Gp0191706 (*Actinotalea* sp. SUR-A1), Gp0151178 (*Citrobacter* sp. SUR-1), Gp0191707 (*Marinobacter* sp. SUR-4), Gp0191708 (*Vibrio* sp. SUR-5), Gp0191709 (*Tessaracoccus* sp. SUR-6), Gp0191711 (*Pseudomonas* sp. SYD-2) and Gp0151176 (*Thauera* sp. SYD-3).

16S rRNA phylogenetic analysis. 16S rRNA gene sequences were taken from IMG annotations of genomes or through the use of the bioinformatic tool Kelpie (<https://github.com/PaulGreenfieldOz/WorkingDogs/tree/master/Kelpie>). BLAST searches were conducted against the NCBI RNA reference sequence database to find closest relatives and percentage identity. 16S rRNA gene sequences for representative sequenced relatives of each of the isolates were obtained from GenBank. Phylogenetic trees were constructed from 16S rRNA gene sequences using the default settings at the Phylogeny.fr webserver (<http://www.phylogeny.fr>)⁶². In brief, sequences were aligned with Muscle⁶³, the alignment checked and corrected using Gblocks⁶⁴, the tree structure calculated using PhyML⁶⁵, prior to visualisation using TreeDyn⁶⁶. Branch support for the 16S rRNA phylogenetic tree was carried out using the aLRT method to bootstrap analysis⁶⁷.

Biogeography of isolates. In order to examine the biogeography of isolates, the V4 region of 16S rRNA gene sequences of isolates were matched to the Coal Seam MicroBiome (CSMB) OTU dataset, which represents a large set of environmental survey OTUs from studied coal seams worldwide⁶. Sequences were matched using the `-usearch_global` command from the USEARCH software with an identity threshold cut-off of 97.5%⁶⁸. CSMB OTU matches were examined to determine the set of locations where the OTUs best matching each cultured strain had been observed previously (<https://doi.org/10.4225/08/5ab47c2bc8cfb>)⁶.

Ecophysiological genome analysis. Genomes were annotated for predicted coding sequences using the PROKKA software⁶⁹ and predicted protein sequences functionally annotated using: CRISPRFinder⁴¹, ISSaga⁴⁰, TransportDB 2.0³⁷, dbCAN³⁵, antiSMASH³⁹, MEROPS³⁸ and signalP 4.1⁷⁰.

Carbon utilization phenotyping. Carbon utilisation capabilities of isolates were assayed with the Biolog microplate system (Biolog, California). PM1 and PM2 (Phenotype Microplate) carbon source microplates were used according to the manufacturer's instructions with either the gram negative or positive protocol depending on the gram stain of the isolate. Respiration was measured by colourimetric changes caused by the reduction of a tetrazolium redox indicator dye mix. The dye mixes used for the individual isolates were chosen according to manufacturer recommendations based on taxonomic affiliation and rates of growth of the isolates. Dye mix A was used for BOW-2, BOW-7, SUR-1 and SYD-3; dye mix H was used for BOW-1, SUR-A1, SUR-4, SUR-6 and SYD-3 and dye mix D was used for SUR-5. Inoculated Biolog microplates were incubated in the Omnilog automated incubator-reader for 36 hours at temperatures matching those found *in situ* for the coal seams from which isolates were obtained: 40 °C for isolates BOW-1, BOW-2 and BOW-7, 30 °C for isolates SUR-A1, SUR-4, SUR-5, SUR-6 and Sur-1 and at 37 °C for isolates SYD-2 and SYD -3 with absorbance at 590 nm measured at 15 minute intervals for each well.

Received: 16 May 2019; Accepted: 4 September 2019;

Published online: 07 November 2019

References

- Rabl, A. & Spadaro, J. V. Environmental impacts and costs of energy. *Ann. N. Y. Acad. Sci.* **1076**, 516–526 (2006).
- Markandya, A. & Wilkinson, P. Electricity generation and health. *The Lancet* **370**, 979–990 (2007).
- Ritter, D. *et al.* Enhanced microbial coalbed methane generation: a review of research, commercial activity, and remaining challenges. *International Journal of Coal Geology* **146**, 28–41 (2015).
- Strapoč, D. *et al.* Biogeochemistry of microbial coal-bed methane. *Annual Review of Earth and Planetary Sciences* **39**, 617–656 (2011).
- Colosimo, F. *et al.* Biogenic methane in shale gas and coal bed methane: A review of current knowledge and gaps. *International Journal of Coal Geology* **165**, 106–120 (2016).
- Vick, S. H. *et al.* The Coal Seam Microbiome (CSMB) reference set, a lingua franca for the microbial coal-to-methane community. *International Journal of Coal Geology* **186**, 41–50 (2018).
- Liu, Y. & Whitman, W. B. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann. N. Y. Acad. Sci.* **1125**, 171–189 (2008).
- Mayumi, D. *et al.* Methane production from coal by a single methanogen. *Science* **354**, 222–225 (2016).
- Robbins, S. J., Evans, P. N., Parks, D. H., Golding, S. D. & Tyson, G. W. Genome-centric analysis of microbial populations enriched by hydraulic fracture fluid additives in a coal bed methane production well. *Frontiers in microbiology* **7**, 731 (2016).
- Evans, P. N. *et al.* Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434–438 (2015).
- Rosewarne, C. P. *et al.* Draft genome sequence of *Methanobacterium* sp. Maddingley, reconstructed from metagenomic sequencing of a methanogenic microbial consortium enriched from coal-seam gas formation water. *Genome announcements* **1**, e00082–00012 (2013).
- Fry, J. C. *et al.* Prokaryotic populations and activities in an interbedded coal deposit, including a previously deeply buried section (1.6–2.3 km) above ~150 Ma basement rock. *Geomicrobiology Journal* **26**, 163–178 (2009).
- Guo, H. *et al.* Pyrosequencing reveals the dominance of methylotrophic methanogenesis in a coal bed methane reservoir associated with Eastern Ordos Basin in China. *International Journal of Coal Geology* **93**, 56–61 (2012).
- Jones, E. J., Voytek, M. A., Corum, M. D. & Orem, W. H. Stimulation of methane generation from nonproductive coal by addition of nutrients or a microbial consortium. *Appl. Environ. Microbiol.* **76**, 7013–7022 (2010).
- Lawson, C. E. *et al.* Patterns of endemism and habitat selection in coalbed microbial communities. *Appl. Environ. Microbiol.* **81**, 7924–7937 (2015).
- Li, D., Hendry, P. & Faiz, M. A survey of the microbial populations in some Australian coalbed methane reservoirs. *International Journal of Coal Geology* **76**, 14–24 (2008).

17. Penner, T. J., Foght, J. M. & Budwill, K. Microbial diversity of western Canadian subsurface coal beds and methanogenic coal enrichment cultures. *International Journal of Coal Geology* **82**, 81–93 (2010).
18. Shimizu, S. *et al.* Molecular characterization of microbial communities in deep coal seam groundwater of northern Japan. *Geobiology* **5**, 423–433 (2007).
19. Singh, D. N., Kumar, A., Sarbhai, M. P. & Tripathi, A. K. Cultivation-independent analysis of archaeal and bacterial communities of the formation water in an Indian coal bed to enhance biotransformation of coal into methane. *Appl. Microbiol. Biotechnol.* **93**, 1337–1350 (2011).
20. Tang, Y.-Q. *et al.* Diverse microbial community from the coalbeds of the Ordos Basin, China. *International Journal of Coal Geology* **90–91**, 21–33 (2012).
21. Wei, M., Yu, Z. & Zhang, H. Microbial Diversity and Abundance in a Representative Small-Production Coal Mine of Central China. *Energy Fuels* **27**, 3821–3829 (2013).
22. Wei, M., Yu, Z., Jiang, Z. & Zhang, H. Microbial diversity and biogenic methane potential of a thermogenic-gas coal mine. *International Journal of Coal Geology* **134**, 96–107 (2014).
23. Gallagher, L. K. *et al.* The effect of coal oxidation on methane production and microbial community structure in Powder River Basin coal. *International Journal of Coal Geology* **115**, 71–78 (2013).
24. Vick, S. H. *et al.* Revealing colonisation and biofilm formation of an adherent coal seam associated microbial community on a coal surface. *International Journal of Coal Geology* **160**, 42–50 (2016).
25. Beckmann, S. *et al.* Long-term succession in a coal seam microbiome during *in situ* biostimulation of coalbed-methane generation. *The ISME journal* **13**, 632 (2019).
26. Singh, D. N. & Tripathi, A. K. Coal induced production of a rhamnolipid biosurfactant by *Pseudomonas stutzeri*, isolated from the formation water of Jharia coalbed. *Bioresour. Technol.* **128**, 215–221 (2013).
27. Wang, L. *et al.* Diverse Bacteria with Lignin Degrading Potentials Isolated from Two Ranks of Coal. *Frontiers in Microbiology* **7**, 1428 (2016).
28. An, D. *et al.* Metagenomics of hydrocarbon resource environments indicates aerobic taxa and genes to be unexpectedly common. *Environ. Sci. Technol.* **47**, 10708–10717 (2013).
29. Jin, L. *et al.* *Chelatococcus caeni* sp. nov., isolated from a biofilm reactor sludge sample. *International journal of systematic and evolutionary microbiology* **65**, 885–889 (2015).
30. Bagnara, C., Toci, R., Gaudin, C. & Belaich, J. Isolation and characterization of a cellulolytic microorganism, *Cellulomonas fermentans* sp. nov. *International Journal of Systematic Evolutionary Microbiology* **35**, 502–507 (1985).
31. Luo, Y.-J. *et al.* *Marinobacter shengliensis* sp. nov., a moderately halophilic bacterium isolated from oil-contaminated saline soil. *Antonie Van Leeuwenhoek* **107**, 1085–1094 (2015).
32. Takai, K. *et al.* *Marinobacter alkaliphilus* sp. nov., a novel alkaliphilic bacterium isolated from subseafloor alkaline serpentine mud from Ocean Drilling Program Site 1200 at South Chamorro Seamount, Mariana Forearc. *Extremophiles* **9**, 17–27 (2005).
33. Finster, K., Cockell, C., Voytek, M., Gronstal, A. & Kjeldsen, K. U. Description of *Tessaracoccus profundus* sp. nov., a deep-subsurface actinobacterium isolated from a Chesapeake impact crater drill core (940 m depth). *Antonie Van Leeuwenhoek* **96**, 515–526 (2009).
34. Cai, M. *et al.* *Salinarimonas ramus* sp. nov. and *Tessaracoccus oleiagri* sp. nov., isolated from a crude oil-contaminated saline soil. *International journal of systematic and evolutionary microbiology* **61**, 1767–1775 (2011).
35. Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–W451 (2012).
36. Petersen, T. N., Brunak, S., Von Heijne, G. & Nielsen, H. J. Nm SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**, 785 (2011).
37. Elbourne, L. D., Tetu, S. G., Hassan, K. A. & Paulsen, I. T. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res.* **45**, D320–D324 (2017).
38. Rawlings, N. D. *et al.* The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **46**, D624–D632 (2018).
39. Weber, T. *et al.* antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237–W243 (2015).
40. Varani, A. M., Siguier, P., Gourbeyre, E., Charneau, V. & Chandler, M. ISSaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol.* **12**, R30 (2011).
41. Grissa, I., Vergnaud, G. & Pourcel, C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **35**, W52–W57 (2007).
42. Grime, J. P. *Plant strategies, vegetation processes, and ecosystem properties.* (John Wiley & Sons, 2006).
43. Fierer, N., Nemergut, D., Knight, R. & Craine, J. M. Changes through time: integrating microorganisms into the study of succession. *Res. Microbiol.* **161**, 635–642 (2010).
44. Singer, E. *et al.* Genomic potential of *Marinobacter aquaeolei*, a biogeochemical “opportunistic”. *Appl. Environ. Microbiol.* **77**, 2763–2771 (2011).
45. Vick, S. H. *et al.* Succession Patterns and Physical Niche Partitioning in Microbial Communities from Subsurface Coal Seams. *iScience* **12**, 152–167 (2019).
46. Kielak, A. M., Scheublin, T. R., Mendes, L. W., Van Veen, J. A. & Kuramae, E. E. Bacterial community succession in pine-wood decomposition. *Frontiers in microbiology* **7**, 231 (2016).
47. Fuchs, G., Boll, M. & Heider, J. Microbial degradation of aromatic compounds—from one strategy to four. *Nature Reviews Microbiology* **9**, 803–816 (2011).
48. Polz, M. F., Hunt, D. E., Preheim, S. P. & Weinreich, D. M. Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 2009–2021 (2006).
49. Bodey, G. P., Bolivar, R., Fainstein, V. & Jadeja, L. Infections caused by *Pseudomonas aeruginosa*. *Rev. Infect. Dis.* **5**, 279–313 (1983).
50. Hirano, S. S. & Upper, C. D. J. M. M. B. R. Bacteria in the Leaf Ecosystem with Emphasis on *Pseudomonas syringae*—a Pathogen, Ice Nucleus, and Epiphyte. *Microbiol. Mol. Biol. Rev.* **64**, 624–653 (2000).
51. Dowling, D. N. & O’Gara, F. Metabolites of *Pseudomonas* involved in the biocontrol of plant disease. *Trends Biotechnol.* **12**, 133–141 (1994).
52. Stolp, H. & Gadkari, D. In *The prokaryotes* (eds Starr, M. P. *et al.*) 719–741 (Springer, 1981).
53. Schwarz, W. The cellulosome and cellulose degradation by anaerobic bacteria. *Appl. Microbiol. Biotechnol.* **56**, 634–649 (2001).
54. Méndez, C. & Salas, J. A. The role of ABC transporters in antibiotic-producing organisms: drug secretion and resistance mechanisms. *Res. Microbiol.* **152**, 341–350 (2001).
55. Siguier, P., Filée, J. & Chandler, M. Insertion sequences in prokaryotic genomes. *Curr. Opin. Microbiol.* **9**, 526–531 (2006).
56. Daly, R. A. *et al.* Microbial metabolisms in a 2.5-km-deep ecosystem created by hydraulic fracturing in shales. *Nature Microbiology* **1**, 16146 (2016).
57. Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541 (1999).
58. Barnhart, E. P. *et al.* Hydrogeochemistry and coal-associated bacterial populations from a methanogenic coal bed. *International Journal of Coal Geology* **162**, 14–26 (2016).

59. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
60. Greenfield, P., Duesing, K., Papanicolaou, A. & Bauer, D. C. Blue: correcting sequencing errors using consensus and context. *Bioinformatics* **30**, 2723–2732 (2014).
61. Markowitz, V. M. *et al.* IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122 (2012).
62. Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469 (2008).
63. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
64. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
65. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
66. Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B. & Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**, 439 (2006).
67. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539–552 (2006).
68. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
69. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
70. Peterson, D. G., Wessler, S. R. & Paterson, A. H. Efficient capture of unique sequences from eukaryotic genomes. *Trends Genet.* **18**, 547–550 (2002).

Acknowledgements

Microscopy was undertaken at the Macquarie University microscopy department with training and assistance provided by Dr. Sue Lindsay and Dr. Nadia Suarez-Bosche. Mr. Silas Vick was supported by a Macquarie University postgraduate scholarship and the work carried out in the current study was supported by the CSIRO Energy strategic research initiative.

Author contributions

S.H.W.V., S.G.T., D.J.M. and I.T.P. conceived the study. S.H.W.V. performed isolations, phenotypic characterisations, scanning electron microscopy, DNA extractions, 16S rRNA phylogenetic analysis and sequence accessioning. P.G. performed sequence error corrections using Blue and *de novo* assembly with Velvet 1.2.10. S.H.W.V. performed quality control, genome annotation and all subsequent analyses of genomic data. S.H.W.V. drafted the manuscript with assistance from S.G.T., D.J.M. and I.T.P.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-52846-7>.

Correspondence and requests for materials should be addressed to S.H.W.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019