environmental microbiology reports

Environmental Microbiology Reports (2020) **12**(5), 525–533 doi:10.1111/1758-2229.12868

Brief Report

Subsurface carbon monoxide oxidation capacity revealed through genome-resolved metagenomics of a carboxydotroph

Andre Mu,[1,2] Brian C. Thomas,[3] Jillian F. Banfield[3,4] and John W. Moreau[5†*]

[1]*Department of Microbiology and Immunology, The Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Australia.*

[2]*Doherty Applied Microbial Genomics, Department of Microbiology and Immunology, The Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Australia.*

[3]*Department of Earth and Planetary Science, University of California, Berkeley, CA, USA.*

[4]*Innovative Genomics Institute, University of California, Berkeley, CA, USA.*

[5]*School of Earth Sciences, University of Melbourne, Melbourne, Australia.*

Summary

Microbial communities play important roles in the biogeochemical cycling of carbon in the Earth's deep subsurface. Previously, we demonstrated changes to the microbial community structure of a deep aquifer (1.4 km) receiving 150 tons of injected supercritical CO_2 ($scCO_2$) in a geosequestration experiment. The observed changes support a key role in the aquifer microbiome for the thermophilic CO-utilizing anaerobe *Carboxydocella*, which decreased in relative abundance post-$scCO_2$ injection. Here, we present results from more detailed metagenomic profiling of this experiment, with genome resolution of the native carboxydotrophic *Carboxydocella*. We demonstrate a switch in CO-oxidation potential by *Carboxydocella* through analysis of its carbon monoxide dehydrogenase (CODH) gene before and after the geosequestration experiment. We discuss the potential impacts of $scCO_2$ on subsurface flow of carbon and electrons from oxidation of the metabolic intermediate carbon monoxide (CO).

Received 30 March, 2020; accepted 2 July, 2020. *For correspondence. E-mail john.moreau@glasgow.ac.uk. Tel: +44 (0)141 330 5461. †Present address: School of Geographical and Earth Sciences, University of Glasgow, Glasgow, UK.

Introduction

Carbon monoxide (CO) provides a carbon and/or energy source for a wide range of anaerobic microorganisms (Oelgeschläger and Rother, 2008) as its oxidation can be coupled to multiple terminal electron accepting processes (TEAPs), e.g., iron(III)-reduction (Slobodkin *et al*., 1999), sulfate-reduction (Parshina *et al*., 2005, 2010) hydrogenogenesis, acetogenesis or methanogenesis (Diender *et al*., 2015). The electrons produced from CO oxidation can be coupled to the reduction of H_2O and metals (Techtmann *et al*., 2009). The substrate versatility of CO results from the low redox potential (E°; −524 to −558 mV) of the half-reaction for carbon dioxide (CO_2) reduction to CO (Ragsdale, 2004; Techtmann *et al*., 2009). This potential is lower than that of the H^+/H_2 redox couple, meaning that CO can replace H_2 as an electron donor for microorganisms carrying genes encoding for carbon monoxide dehydrogenase (CODH). The CODH gene cluster in the genus, *Carboxydocella*, consists of 13 genes encoding for: redox- and CO-sensitive transcriptional regulator (*cooA*), [Ni, Fe]-CODH (*cooS*), [Ni, Fe]-CODH accessory nickel-insertion protein (*cooC*), CO-induced hydrogenase membrane anchor (*cooM*), energy-converging hydrogenase (*cooKLXUH*), hydrogenase maturation protein (*hypA*), 4Fe-4S di-cluster domain containing electron transfer protein (*cooF*), energy-converging hydrogenase catalytic subunit (*cooH*), and the upstream enigmatic *cooSC* operon of unknown function (Toshchakov *et al*., 2018). The upstream *cooS* gene is notably missing in the closely related *Carboxydothermus hydrogenoformans* (Toshchakov *et al*., 2018). The catalytic subunit of CODH is composed of the CooS protein, while the CooF electron-transfer protein is known to associate with redox reactions such as the production of H_2 or reduced metals (Soboh *et al*., 2002). Bi-functional CODH enzyme complexes contain domains that catalyse the oxidation of CO (CODH) and the formation of acetyl-CoA (acetyl-CoA synthase [ACS]; Ragsdale

and Kumar, 1996). Biochemically, the electrons produced by CODH flow on to the ACS module for downstream production of acetyl-CoA. Bi-functional CODH/ACS complexes are therefore fundamental to acetoclastic methanogens and acetogenesis in prokaryotes using the Wood–Ljungdahl pathway to form acetyl-CoA from either CO_2 or CO (Ragsdale, 2004). The CODH/ACS complex is also responsible for the majority of autotrophic utilization of CO (Techtmann *et al*., 2009). An increase in CO_2 partial pressure may therefore inhibit microbial carboxydotrophy, resulting in increasing and decreasing levels of CO and H_2 respectively. Such changes potentially hold consequences for both carbon and electron flow through various microbial metabolisms (Ragsdale, 2004; Techtmann *et al*., 2009).

Using phenotypic and physiological assays, Sokolova *et al*. (2002) demonstrated that a 40% decrease in CO in cultures of the anaerobic Firmicute, *Carboxydocella*, led to a 30% increase in the gas phase concentrations of H_2 and CO_2 (Sokolova *et al*., 2002). Furthermore, pure cultures of *Desulfotomaculum kuznetsovii* and *D. thermobenzoicum* subsp. *thermosynthrophicum* were shown to grow on CO as the sole electron donor, and at concentrations as high as 50%–70%, in the presence of hydrogen/CO_2. However, co-culture of *Desulfotomaculum kuznetsovii* and *D. thermobenzoicum* subsp. *thermosynthrophicum* with a carboxydotroph (i.e. *Carboxydothermus hydrogenoformans*) supported sulfate-reducing bacteria (SRB) growth in 100% CO, with CO oxidation coupled to SO_4^{2-} reduction and acetogenesis (pCO = 120 kPa) (Parshina *et al*., 2005).

Injection of massive volumes of supercritical CO_2 ($scCO_2$) into deep aquifers forms a principal current strategy for geological carbon storage (Benson and Surles, 2006; Gibbins and Chalmers, 2008). These $scCO_2$ 'plumes' will enrich groundwaters locally with respect to dissolved CO_2, with implications for both microbial metabolic activity and water-rock chemical reactions (Phillips *et al*., 2012). Understanding the *in situ* impacts of increased CO_2 on microbial CODH activity will yield insights into the long-term and large-scale potential responses of the subsurface microbial biosphere to geological CO_2 sequestration.

Here, we report results from a field-scale geological $scCO_2$ injection project in the Paaratte Formation of the Otway Basin (1.4 km below ground, Southeastern Australia) that revealed a steep decline in the relative abundance of a dominant native carboxydotroph representing >96% of the aquifer microbial community before injection (Mu *et al*., 2014). Previous studies of subsurface microbial responses to quickly elevated CO_2 levels have inferred their findings on the basis of 16S rRNA gene data (Bordenave *et al*., 2013; Lavalleur and Colwell, 2013; Mu *et al*., 2014). Here, we used time-series relative abundance *in situ* metagenomic sampling to resolve a near-complete genome from the carboxydotrophic genus *Carboxydocella* and established the microbial host of the CODH gene cluster.

Results

Metagenomic analysis of the Paaratte Formation

A total of 13 samples collected from the Paaratte Formation over the course of a field-scale demonstration of supercritical carbon dioxide ($scCO_2$) geosequestration were processed for shotgun metagenomics (Table 1) with multiple displacement amplification (MDA). Over 500 tons of ground water were produced during the pre-$scCO_2$ injection phase, followed by the injection of 150 tons of $scCO_2$. Ten samples were collected during the pre-CO_2 injection phase, two samples during the post-CO_2 injection phase, and one sample from the core mud. The mean number of reads that passed quality control during the pre-$scCO_2$ injection phase was 600 766, while for the post-$scCO_2$ injection phase was 281 383. Mean GC content throughout the geosequestration experiment was 53%. The number of reads with identified functional categories was 76 662. The metagenome of the pre-CO_2 injection phase sample PF13 contained 1 034 457 reads post-QC. Anaerobic CODH (represented by *cooS*) is absent from the two post-CO_2 injection phase samples (Fig. 1A), while aerobic CODH (represented by *coxL*) on the other hand was present at low levels during the pre-CO_2 injection phase, but proliferated by ≥2.7-fold in sample PF145, which is during the late phases of post-CO_2 injection (Fig. 1B).

Assembly of metagenomic sequence data

Biodiversity analyses from our previous study (Mu *et al*., 2014) informed the targeted selection of four samples from the pre-CO_2 injection phase (samples PF10, 11, 13, and 15) for genome-resolved metagenomic analyses. Metagenomic assembly of these samples generated a total of 10 492 contigs with an average of 57.83% GC content and 31 465 predicted coding sequences (Table 2).

Metagenomic reconstruction of aquifer-native carboxydotroph

Metagenomic bins were created for a carboxydotroph *Carboxydocella* using ggKbase from samples PF10, PF11, PF13, and PF15, based on the following bin phylogenetic profile; Bacteria > *Firmicutes* > *Clostridia* > *Clostridiales*; and GC contents (%) of 43.56–51.99 for PF10, 40.7–53.98 for PF11, 43.13–55.0 for PF13, and 39.43–51.5 for PF15. The metagenomic bin for *Carboxydocella* from PF13 had the most bacterial

Table 1. Summary of 13 samples analysed in this study.

	Baseline water production												CO_2 injection		Post-injection water production	
	50 tons day^{-1} for 10.2 days												37.5 tons day^{-1} for 4 days		45 tons day^{-1} for 3.8 days	
	Total of 510 tons of formation water												Total of 150 tons of CO_2 injected		Total of 170 tons of gas saturated formation water	
	Sample ID	Coremud	PF8	PF10	PF11	PF13	PF14	PF15	PF16	PF19	PF21	PF22	CO_2 injection	CO_2 injection	PF115	PF145
Geochemistry data	Injection phase	Pre-CO_2 drilling mud	Pre-CO_2	Pre-CO_2	Pre-CO_2	Pre-CO_2	Pre-CO_2	Pre-CO_2	Pre-CO_2	Pre-CO_2	Pre-CO_2	Pre-CO_2	Beginning	End	Post-CO_2	Post-CO_2
	Timeline[a]	Pre 17 June	17 June	18 June	18 June	19 June	20 June	20 June	21 June	21 June	23 June	23 June	10 July	14 July	August	August
	pH	-	7.93	7.97	7.9	8.47	8.09	7.94	8.13	8.22	8.06	8.04			5.59	5.57
	Total dissolved solids (mg l^{-1})	-	953	887	949	863	892	892	990	863	864	850			1063	1140
Genomic data	Sequence data available	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
	Number of reads post-QC	380 967	864 693	821 323	488 243	1,034 457	629 790	813 519	730 189	375 497	456 554	13 309			37 912	524 855
	Mean GC (%)	57	57	55	53	49	48	54	55	46	51	56			58	58
	Number of reads with identified functional categories	51 949	170 774	142 767	83 260	57 699	81 597	110 492	128 812	55 940	75 695	1 282			4 068	32 276
	Range sequence length (bp)	193 ± 92	164 ± 78	178 ± 86	178 ± 86	183 ± 94	192 ± 101	189 ± 92	212 ± 101	220 ± 102	199 ± 97	163 ± 89			240 ± 102	208 ± 99

[a]The temporal relationship, as indicated by the day and month, of each sample occurs over the year 2011.
-: Data are unavailable due to the dedicated sampling time points required by multi-disciplinary collaborative experiments.
PF–: where PF represents 'Paaratte Formation', and '–' represents the number of U-tube samples since time origin.

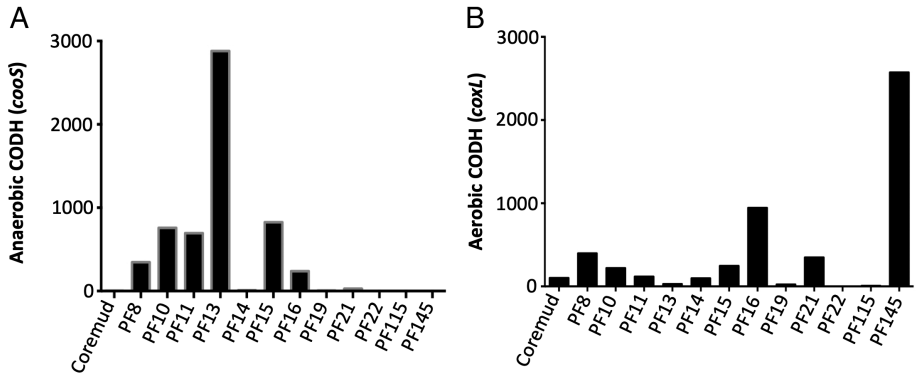

Fig 1. Abundances of anaerobic and aerobic CODH genes throughout the geosequestration project.A. An abundance profile of sequence data from the Paaratte Formation mapped to CooS. Coos is the catalytic subunit associated with the anaerobic carbon monoxide dehydrogenase enzyme.B. An abundance profile of sequence data from the Paaratte Formation mapped to CoxL. CoxL is the large subunit associated with the aerobic carbon monoxide dehydrogenase enzyme.

Table 2. Genomic features of metagenomic bins using pre-CO_2 injection phase samples.

Samples	Completeness	Size	%GC	Average coverage	Number of contigs	Number of genes
PF10	RP: 47 bSCG: 47	10.29 Mbp	60.4	16.1	4326	13 351
PF11	RP: 41 bSCG: 42	5.37 Mbp	58.01	13.94	2582	7348
PF13	RP: 47 bSCG: 50	389.27 Kbp	53.69	101.69	183	509
PF15	RP: 36 bSCG: 33	7.72 Mbp	59.2	36.58	3401	10 257
Clostridiales bins						
PF10	RP: 44 bSCG: 48	2.33 Mbp	49.48	45.56	195	2521
PF11	RP: 42 bSCG: 46	2.13 Mbp	49.49	37.05	216	2357
PF13	RP: 47 bSCG: 50	2.43 Mbp	49.07	226.19	170	2659
PF15	RP: 35 bSCG: 38	1.95 Mbp	50.05	59.54	427	2318

bSCG = bacterial single-copy genes; RP = Ribosomal proteins.

single-copy genes (bSCGs) (50 of 51 expected bSCGs). This sample contained highest relative abundances of *Carboxydocella*, as determined by amplicon 16S rRNA gene profiling. Samples PF10, PF11, and PF15 returned *Carboxydocella* genomes with 48, 46 and 38 of each of the 51 expected bSCGs, respectively (Table 1). Notably, the gene encoding for the catalytic subunit of CODH enzyme (*cooS*) was present in the genomes from all four metagenomic datasets (Fig. 2B); also present were genes associated with *Flagella-related proteins*.

The manually curated genome-resolved metagenome of the aquifer-native carboxydotroph is 2 544 661 bp in length, consisting of six rRNA genes, 64 tRNAs, 2584 coding sequences (CDS), and has a GC content of 49.2% GC. These genomic characteristics are similar to those of the two reference *Carboxydocella* strains. The genome is classified as near-complete (94.6% complete) when assessed using *C. thermautotrophica* strain 019 as the reference genome (Table 3). BLASTP and phylogeny-based analyses revealed high similarity between aquifer-native carboxydotroph and the two reference strains (*C. thermautotrophica* strain 019, accession number: CP028491; and strain 41, accession number: CP028514). The genome has the complete gene cluster determining hydrogenic carboxydotrophy: *cooA* – transcriptional regulator; *cooS* – Ni,Fe-CODH gene; *cooC* – Ni,Fe-CODH accessory nickel-insertion protein gene; *cooMKLXUH* – energy-converting hydrogenase gene; *hypA* – hydrogenase maturation protein gene; *cooF* - ferredoxin-like protein gene; and *cooH* – ECH catalytic subunit gene. The gene cluster also includes an upstream enigmatic *cooSC* operon, in which the function in *Carboxydocella* is unknown (Fig. 2D). The reconstructed aquifer-native *Carboxydocella* genome is available under the European Nucleotide Archive (ENA) study accession number: PRJEB37484.

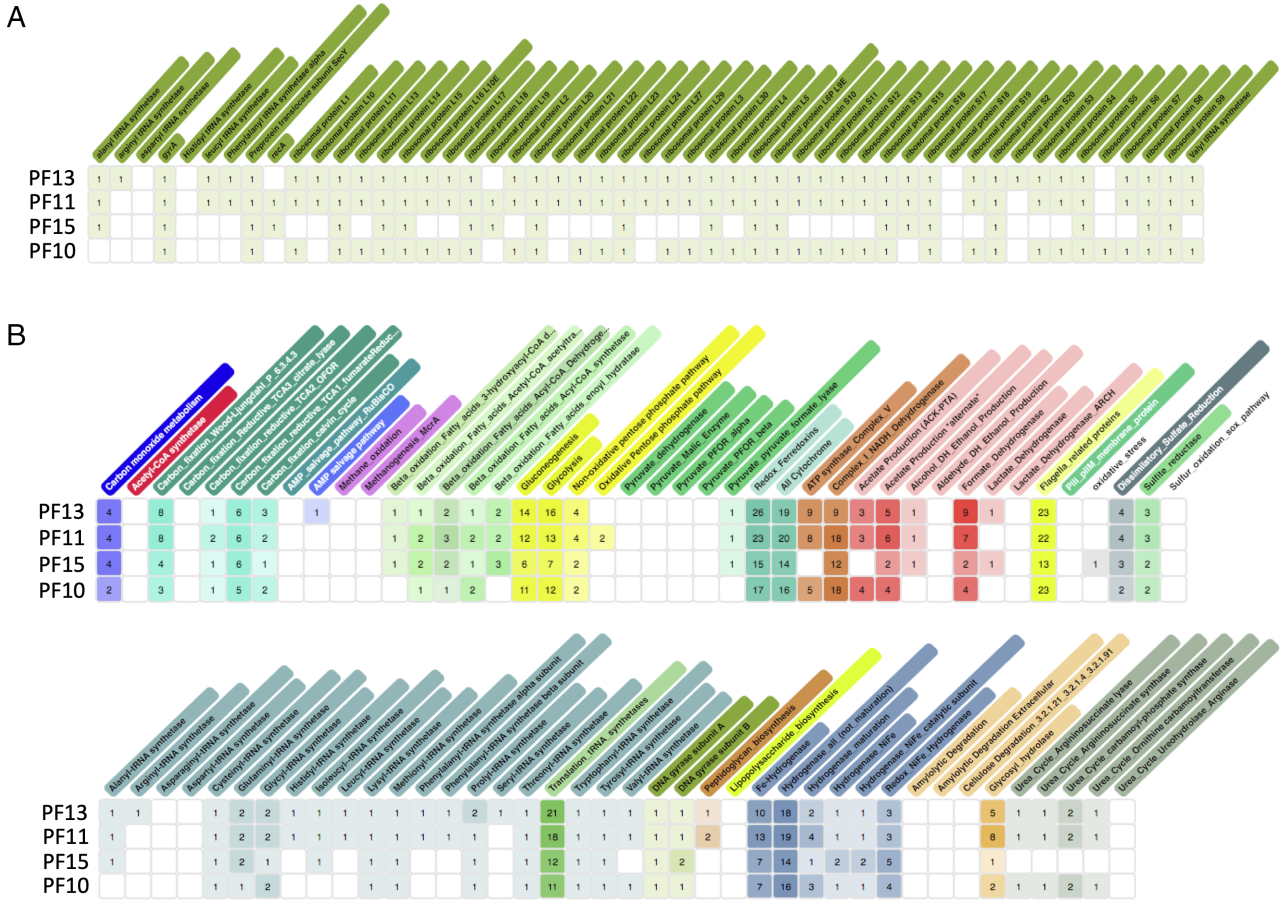

Fig 2. Genomic characterization of genome-resolved aquifer-native carboxydotroph metagenome.A. Bacterial single-copy gene markers for metagenomic bins representing the *Carboxydocella* operational taxonomical unit. Bacterial single-copy genes are used as genetic biomarkers to validate the genome is derived from a single community isolate, and as a proxy for genome completeness. Contigs were clustered together using the ggKbase analysis tool.B. Genomic features of aquifer-native carboxydotrophic metagenomic bins. Each row is representative of a *Carboxydocella* genome from a single time point of the geosequestration project. The presence of a genomic feature is indicated by a coloured cell, and its abundance by the numerical value.

Table 3. Genomic features of the aquifer-native *Carboxydocella* compared to *C. thermautotrophica* reference strains.

	Genome length (bp)	% GC	CDS	rRNA	tRNA	Number of contigs	Ns	Min contig	Max contig	N50
PF-AMU[a]	2 544 661	49.2	2584	6	64	55	15	1369 bp	252 013 bp	89 432 bp
C. thermautotrophica 019	2 676 584	49.1	2689	15	71					
C. thermautotrophica 41	2 690 058	49.1	2697	15	73					

[a]Near-complete genome-resolved metagenome of aquifer native *Carboxydocella*.

Discussion

Analysis of 16S rRNA gene data revealed a proliferation of *Comamonadaceae* and *Sphingomonadaceae*, and a decline of *Carboxydocella*, following the injection of $scCO_2$ to the Paaratte Formation (Mu *et al*., 2014). Temporal shifts in phylogeny also related to changes in metabolic potential and provided insights into likely effects of $scCO_2$ on CO metabolism by *Carboxydocella* (Mu *et al*., 2014). *Carboxydocella* is a Gram positive *Firmicute* that has been phenotypically (including biochemically) described by Sokolova and colleagues (2002) to be motile (lateral flagella), and capable of oxidizing CO as its sole carbon source. However, there are few sequenced reference strains available in public databases for attributing the molecular mechanisms behind the oxidation of CO in terrestrial subsurface environments (Doukov *et al*., 2002).

Metagenomic 'binning' of sequence data based on organism-specific characteristics (e.g., GC content,

phylogeny, coverage, and bacterial single-copy genes), and manual curation of assembly data produced a high-quality genome for an aquifer-native carboxydotroph; this genome represents a near-complete genome at 94.6% complete compared to *C. thermautotrophica* strain 019, and at 49.2 percentage GC (Fig. 3A–C). Despite our necessary use of MDA to prepare genomic samples for sequencing, which has inherent biases to GC% content and taxonomy, the metagenomic bins and reconstructed carboxydotroph genome demonstrated comparable genomic characteristics to reference isolates (Toshchakov *et al*., 2018). Reconstruction of the genome for aquifer-native carboxydotrophs revealed the presence of anaerobic Ni-dependent CODH CooS (*cooS*) belonging to *Carboxydocella*, supporting the capacity of this organism for CO oxidation (Fig. 3D). Consistent with expectations based on genomes of related bacteria, the aquifer-native carboxydotroph genome encodes flagella-related proteins

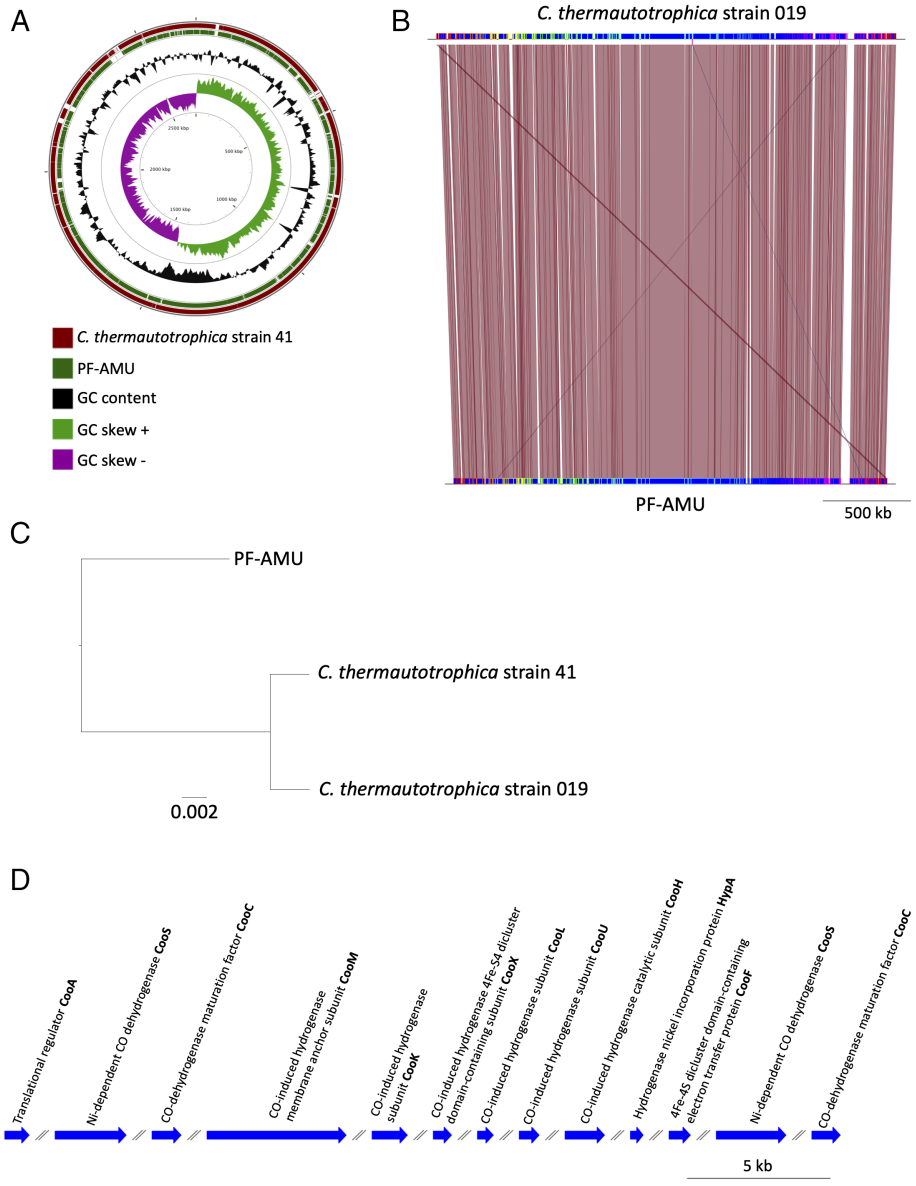

Fig 3. Comparative analysis of the aquifer-native carboxydotroph with representative carboxydotrophic microbial species.A. CGView tools (BLASTP) analysis of *Carboxydocella thermautotrophica* strains 41 and the aquifer-native carboxydotroph (PF-AMU) with *C. thermautotrophica* str 019 as the reference.B. MauveProgressive aligner comparasion of aquifer-native carboxydotroph and *C. thermautotrophica* str 019 genomes as visualized by genoPlotR.C. Core SNP phylogenetic tree of *C. thermautotrophica* strains 41 and 019, and the aquifer-native carboxydotroph. Variant analysis was computed using Prokka for *C. thermautotrophica* strain 41 and the aquifer-native carboxydotroph against *C. thermautotrophica* strain 019 as the reference.D. Schematic of the gene cluster determining hydrogenic carboxydotrophy in the aquifer-native *Carboxydocella* genome. The gene cluster is homologous to that of *C. thermautotrophica* strain 019 and 41.

and lacks lipopolysaccharide biosynthetic genes (Fig. 2B); these key genomic features support the phenotypic characterization of *Carboxydocella* being motile via lateral flagella and having a cell wall structure consistent of a Gram positive microorganism (Sokolova *et al*., 2002).

Metagenomic data revealed that representative aerobic CODH, *coxL*, genes proliferated during the late phase of post-scCO_2 injection by ≥2.7-fold (Fig. 1B). The observed increase in aerobic CODH may have resulted from a decline in the activity of *Carboxydocella*, which cannot grow in the presence of large volumes of CO_2 (Sokolova *et al*., 2002). However, proliferation of *coxL*, and a decline in *cooS*, suggests a metabolic switch in CO-oxidation potential within the subsurface. CO-oxidation is likely maintained by microbes other than anaerobic carboxydotrophs, such as those using molybdenum containing CO oxidoreductase enzymes (Rajagopalan, 1984; King and Weber, 2007). This observed switch in CO-oxidation potential, rather than a complete loss of function within the aquifer community, implies that sulfate-reducing bacteria and methanogens would have sufficient H_2 to sustain physiological function (Mu and Moreau, 2015). Furthermore, few aerobic carboxydotrophs have been documented in thermophilic environments (King and Weber, 2007); therefore, our findings extend the range of geological environments for aerobic CO-oxidisers.

This study offers insights for future *in vitro* studies of pure culture and mixed consortia responses to scCO_2. Our results elucidate the roles of anaerobic and aerobic CODH in microbial community functionality in a system under scCO_2 stress and expand the ecology of CO-utilizing microbes to include the terrestrial deep subsurface. Furthermore, the genome-resolved metagenome of a Paaratte Formation-derived *Carboxydocella* organism provides a greater molecular understanding of CO-oxidation within the deep subsurface.

Experimental procedures

Groundwater sampling and extraction of whole community genomic DNA

In situ groundwater samples were collected using the U-tube sampling system (Freifeld *et al*., 2005; Freifeld, 2009) as part of the CO2CRC Otway Stage 2B field experiment (Paterson *et al*., 2013). Details of groundwater sampling and extraction of whole community genomic DNA (gDNA) are given in the study by Mu *et al*. (2014). The injection and sampling well was screened at approximately 1400 m true vertical depth subsea (TVDSS) in a sandstone unit of the Paaratte Formation of the Otway Basin at latitude: 38° 31′ 44″ and longitude: 142° 48′ 43″. A series of valve controls at surface were used to pass high-pressure nitrogen gas through the U-tube system and direct formation water to the surface for collection under approximately 2000 psi. Each sample collected via the U-tube system was designated the nomenclature 'PF---'; where 'PF' represents *Paaratte Formation*, and '---' represents the number of U-tube samples since time of origin, which was taken as 17 June 2011. Control samples included nucleic acid extracted from core mud, which is representative of drilling fluid used during the emplacement of injection and sampling wells, and samples from water tanks used to store formation water at the surface. Samples processed for metagenomic analysis are highlighted in Table 1.

Preparation and processing of genomic DNA

Whole genome amplification. To increase the amount of high-quality gDNA to meet input requirements for Illumina's Nextera XT sample preparation kit, it was necessary to amplify extracted gDNA using MDA (Warnecke and Hess, 2009; Kawai *et al*., 2014) with Qiagen's REPLI-g Ultra-Fast Mini Kit following manufacturer's protocol. Genomic DNA from samples PF8, PF10, PF11, PF13, PF14, PF15, PF16, PF19, PF21, PF22, PF115, PF145, and core mud were amplified using MDA.

Genomic DNA preparation for metagenomic sequencing. Sample preparation and metagenomic sequencing were performed using the Illumina MiSeq located in the Department of Microbiology and Immunology at the Peter Doherty Institute for Infection and Immunity, University of Melbourne, Australia.

Multiple displacement amplified gDNA was quantitated using the Qubit dsDNA BR (Molecular probes®) assay system following manufacturer's protocol. Samples of sufficient quality were processed using Illumina's Nextera XT sample preparation kit to generate Clean Amplified Nextera Tagment Amplicons following the manufacturer's protocol. Clean Amplified Nextera Tagment Amplicon-DNA concentrations were checked using Qubit dsDNA High Sensitivity Kit, while DNA fragment sizes were validated and quantified using the Agilent 2100 Bioanalyzer and Agilent high-sensitivity DNA kit. Samples were diluted using Qiagen's EB (elution buffer) in place of Tris-Cl 10 mM 0.1% Tween 20.

Illumina MiSeq sequencing. Samples were processed for sequencing using the Illumina MiSeq reagent kit v2 (500 cycle) on the MiSeq machine following a modified manufacturer's protocol. The following modifications were included: a 1% (v/v) spike-in ratio of PhiX, denatured DNA was diluted to a final concentration of 12 pM with pre-chilled HT1 buffer, and Tris-Cl 10 mM 0.1% Tween 20 was substituted with Qiagen's EB solution to dilute sequencing libraries and PhiX throughout the protocol. Raw sequencing data are available through the European Nucleotide Archive (ENA) study accession number: PRJEB37484.

Metagenomic analysis

Pre-processing steps included the removal of artificial sequences generated by sequencing artefacts, and

filtering sequence reads that mapped to the *Homo sapiens* genome from the library intended for further analysis using Bowtie (Langmead and Salzberg, 2012). The preprocessing step aimed to remove contaminating sequences derived from the handling and preparation of gDNA libraries. Sequences were trimmed to contain at most five bases below a Phred score of 15, which was considered to be the lowest quality score included as a high-quality base. Furthermore, the maximum allowed number of ambiguous base pairs per sequence read was five.

Illumina sequence data were processed with Trimmomatic (version 0.33; (Bolger *et al*., 2014) for quality control and to remove adaptor sequences, PhiX contamination, and trace contaminants from Illumina preparation kits. Paired-end reads were merged and assembled using Iterative de Bruijn Graph De Novo Assembler for Uneven sequencing Depth (IDBA-UD; Peng *et al*., 2012) compiled for long reads (i.e. 351 bp). Scaffold headers in data files were amended to include read mapping coverage calculations from Bowtie2 (version 2.2.4; Langmead and Salzberg, 2012) using a ruby script (*add_read_count.rb*; source code available from https://github.com/bcthomas/misc_scripts/tree/master/add_read_count) written by University of California, Berkeley researchers. Prokaryotic Dynamic Programming Gene finding Algorithm (Prodigal v2.6.3; (Hyatt *et al*., 2010) was utilized to predict genes and small RNAs from assembled sequence data greater than 1000 bp in length. Annotation of the predicted proteins were computed by doing similarity searches using *usearch* (v8.1.1861) (Edgar, 2010) and comparing each protein sequence against the following databases: Kyoto Encyclopedia of Genes and Genomes (KEGG; release 71.1 (Kanehisa and Goto, 2000), Consortium, 2009) and Universal Protein Resource (UniProt release 2014_08; Consortium, 2009) including Universal Protein Resource Reference Clusters (UniRef100; Consortium, 2009). Metagenomic data from samples PF10, PF11, PF13, and PF15 were processed for analysis using ggKbase (http://ggkbase.berkeley.edu/) to reconstruct the aquifer-native carboxydotroph. Key metabolic features were analysed against each genome across samples PF10, PF11, PF13, and PF15.

Computational requirements were provided by the Victorian Life Sciences Computation Initiative (Melbourne, Australia), and the University of California, Berkeley (California, USA). Manual curation of the metagenomic bin corresponding to the aquifer-native carboxydotroph was computed using Geneious (v 9.1.8; Biomatters, New Zealand). Prokka was used to annotate the aquifer-native carboxydotroph; annotations were guided by the reference *C. thermautotrophica* strain 019 GenBank file (Accession number: CP028491). Coordinates for the CODH gene cluster were identified and mapped using genoPlotR. CGViewtools and Mauve were also used to visualize genome comparisons between the aquifer-native carboxydotroph, *C. thermautotrophica* strains 019, and 41.

Acknowledgements

The authors acknowledge funding to JWM from the Commonwealth of Australia and industry sponsors through the CO2CRC Program, and the Victorian Life Sciences Computation Initiative (grant number VR0032).

References

Benson, S.M., and Surles, T. (2006) Carbon dioxide capture and storage: an overview with emphasis on capture and storage in deep geological formations. *Proc IEEE* **94**: 1795–1805.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinform Oxf Engl* **30**: 2114–2120.

Bordenave, S., Chatterjee, I., and Voordouw, G. (2013) Microbial community structure and microbial activities related to CO_2 storage capacities of a salt cavern. *Int Biodeter Biodegr* **81**: 82–87.

Consortium, U. (2009) The universal protein resource (UniProt) in 2010. *Nucleic Acids Res* **38**: D142–D148.

Diender, M., Stams, A.J.M., and Sousa, D.Z. (2015) Pathways and bioenergetics of anaerobic carbon monoxide fermentation. *Front Microbiol* **6**: 1275.

Doukov, T.I., Iverson, T.M., Seravalli, J., Ragsdale, S.W., and Drennan, C.L. (2002) A Ni-Fe-Cu center in a bifunctional carbon monoxide dehydrogenase/acetyl-CoA synthase. *Science* **298**: 567–572.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinform Oxf Engl* **26**: 2460–2461.

Freifeld, B. (2009) The U-tube: a new paradigm for borehole fluid sampling. *Sci Drill* **8**: 41–45.

Freifeld, B.M., Trautz, R.C., Kharaka, Y.K., Phelps, T.J., Myer, L.R., Hovorka, S.D., and Collins, D.J. (2005) The U-tube: a novel system for acquiring borehole fluid samples from a deep geologic CO_2 sequestration experiment. *J Geophys Res Solid Earth* **110**: B10203.

Gibbins, J., and Chalmers, H. (2008) Carbon capture and storage. *Energy Policy* **36**: 4317–4322.

Hyatt, D., Chen, G.-L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

Kanehisa, M., and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**: 27–30.

Kawai, M., Futagami, T., Toyoda, A., Takaki, Y., Nishi, S., Hori, S., *et al*. (2014) High frequency of phylogenetically diverse reductive dehalogenase-homologous genes in deep subseafloor sedimentary metagenomes. *Front Microbiol* **5**: 80.

King, G.M., and Weber, C.F. (2007) Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat Rev Microbiol* **5**: 107–118.

Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Lavalleur, H.J., and Colwell, F.S. (2013) Microbial characterization of basalt formation waters targeted for geological carbon sequestration. *FEMS Microbiol Ecol* **85**: 62–73.

Mu, A., Boreham, C., Leong, H.X., Haese, R.R., and Moreau, J.W. (2014) Changes in the deep subsurface microbial biosphere resulting from a field-scale CO_2 geosequestration experiment. *Front Microbiol* **5**: 209.

Mu, A., and Moreau, J.W. (2015) The geomicrobiology of CO_2 geosequestration: a focused review on prokaryotic community responses to field-scale CO_2 injection. *Front Microbiol* **6**: 263.

Oelgeschläger, E., and Rother, M. (2008) Carbon monoxide-dependent energy metabolism in anaerobic bacteria and archaea. *Arch Microbiol* **190**: 257–269.

Parshina, S.N., Kijlstra, S., Henstra, A.M., Sipma, J., Plugge, C.M., and Stams, A.J.M. (2005) Carbon monoxide conversion by thermophilic sulfate-reducing bacteria in pure culture and in co-culture *with Carboxydothermus hydrogenoformans*. *Appl Microbiol Biot* **68**: 390–396.

Parshina, S.N., Sipma, J., Henstra, A.M., and Stams, A.J.M. (2010) Carbon monoxide as an electron donor for the biological reduction of sulphate. *Int J Microbiol* **2010**: 319527.

Paterson, L., Boreham, C., Bunch, M., Dance, T., Ennis-King, J., Freifeld, B., *et al*. (2013) Overview of the CO2CRC Otway residual saturation and dissolution test. *Energy Proced* **37**: 6140–6148.

Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–1428.

Phillips, A.J., Lauchnor, E., Eldring, J.J., Esposito, R., Mitchell, A.C., Gerlach, R., *et al*. (2012) Potential CO_2 leakage reduction through biofilm-induced calcium carbonate precipitation. *Environ Sci Technol* **47**: 142–149.

Ragsdale, S.W. (2004) Life with carbon monoxide. *Crit Rev Biochem Mol* **39**: 165–195.

Ragsdale, S.W., and Kumar, M. (1996) Nickel-containing carbon monoxide dehydrogenase/acetyl-CoA synthase. *Chem Rev* **96**: 2515–2540.

Meyer, O., and Rajagopalan, K.V. (1984) Molybdopterin in carbon monoxide oxidase from carboxydotrophic bacteria. *J Bacteriol* **157**: 643–648.

Slobodkin, A.I., Jeanthon, C., L'Haridon, S., Nazina, T., Miroshnichenko, M., and Bonch-Osmolovskaya, E. (1999) Dissimilatory reduction of Fe(III) by thermophilic bacteria and archaea in deep subsurface petroleum reservoirs of Western Siberia. *Curr Microbiol* **39**: 99–102.

Soboh, B., Linder, D., and Hedderich, R. (2002) Purification and catalytic properties of a CO-oxidizing: H_2-evolving enzyme complex from *Carboxydothermus hydrogenoformans*. *Eur J Biochem* **269**: 5712–5721.

Sokolova, T.G., Kostrikina, N.A., Chernyh, N.A., Tourova, T. P., Kolganova, T.V., and Bonch-Osmolovskaya, E.A. (2002) Carboxydocella thermautotrophica gen. nov., sp. nov., a novel anaerobic, CO-utilizing thermophile from a Kamchatkan hot spring. *Int J Syst Evol Micr* **52**: 1961–1967.

Techtmann, S.M., Colman, A.S., and Robb, F.T. (2009) "That which does not kill us only makes us stronger": the role of carbon monoxide in thermophilic microbial consortia. *Environ Microbiol* **11**: 1027–1037.

Toshchakov, S.V., Lebedinsky, A.V., Sokolova, T.G., Zavarzina, D.G., Korzhenkov, A.A., Teplyuk, A.V., *et al*. (2018) Genomic insights into energy metabolism of *Carboxydocella thermautotrophica* coupling hydrogenogenic CO oxidation with the reduction of Fe(III) minerals. *Front Microbiol* **9**: 1759.

Warnecke, F., and Hess, M. (2009) A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. *J Biotechnol* **142**: 91–95.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1 Supporting Information