Microbiology (2014), **160**, 362–372 **DOI** 10.1099/mic.0.073965-0

Single-cell genomics reveals metabolic strategies for microbial growth and survival in an oligotrophic aquifer

Michael J. Wilkins,[1]† David W. Kennedy,[1] Cindy J. Castelle,[2] Erin K. Field,[3] Ramunas Stepanauskas,[3] James K. Fredrickson[1] and Allan E. Konopka[1]

Correspondence
Michael J. Wilkins
wilkins.231@osu.edu

[1]Pacific Northwest National Laboratory, Richland, WA 99354, USA

[2]University of California Berkeley, Berkeley, CA 94720, USA

[3]Bigelow Laboratory for Ocean Sciences, Boothbay, ME 04544, USA

Bacteria from the genus *Pedobacter* are a major component of microbial assemblages at Hanford Site (a largely decommissioned nuclear production complex) in eastern Washington state, USA, and have been shown to change significantly in abundance in response to the subsurface intrusion of Columbia River water. Here we employed single-cell genomics techniques to shed light on the physiological niche of these micro-organisms. Analysis of four *Pedobacter* single amplified genomes (SAGs) from Hanford Site sediments revealed a chemoheterotrophic lifestyle, with the potential to exist under both aerobic and microaerophilic conditions via expression of both aa_3-type and cbb_3-type cytochrome *c* oxidases. These SAGs encoded a wide range of both intra- and extracellular carbohydrate-active enzymes, potentially enabling the degradation of recalcitrant substrates such as xylan and chitin, and the utilization of more labile sugars such as mannose and fucose. Coupled to these enzymes, a diversity of transporters and sugar-binding molecules were involved in the uptake of carbon from the extracellular local environment. The SAGs were enriched in TonB-dependent receptors, which play a key role in uptake of substrates resulting from degradation of recalcitrant carbon. Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas mechanisms for resisting viral infections were identified in all SAGs. These data demonstrate the potential mechanisms utilized for persistence by heterotrophic micro-organisms in a carbon-limited aquifer, and hint at potential linkages between observed *Pedobacter* abundance shifts within the 300 Area (in the south-eastern corner of the site) subsurface and biogeochemical shifts associated with Columbia River water intrusion.

Received 7 October 2013
Accepted 9 December 2013

INTRODUCTION

Within shallow groundwater aquifers, indigenous microbial populations catalyse a broad range of biogeochemical cycles that have implications for domestic, industrial and agricultural water usage. Biologically catalysed processes include organic contaminant degradation, the release of certain toxic metals such as arsenic and the immobilization of some soluble metals such as uranium (Anderson *et al.*, 2003; Islam *et al.*, 2004; Maymó-Gatell *et al.*, 1997; Williams *et al.*, 2011). Relative to surface environments, these subsurface systems are perceived to be relatively stable and undergo less frequent perturbation. However, despite these general observations, dynamic temporal and spatial shifts within microbial populations have been reported in certain aquifers (Ginige *et al.*, 2013; Lin *et al.*, 2012b; Zhou *et al.*, 2012a); the factors responsible for this variability are currently poorly understood. The majority of analyses of microbial populations have focused on the identification of microbial taxa, via a marker gene such as 16S rRNA. In addition, the ability to generate and analyse shotgun metagenomic data has allowed metabolic and physiological traits to be inferred (Hemme *et al.*, 2010; Wrighton *et al.*, 2012). These datasets have proved critical for understanding (1) how microbial populations interact with their environment, and (2) the factors responsible for driving shifts in microbial population structures over space and time. More recently, single-cell genomics has become a widely applicable research tool, enabling direct rather than

†Present address: School of Earth Sciences & Department of Microbiology, The Ohio State University, Columbus, OH 43210, USA.

Abbreviations: CRISPR, clustered regularly interspaced short palindromic repeat; GH, glycoside hydrolase; MDA, multiple displacement amplification; SAG, single amplified genome; TBDR, TonB-dependent receptor.

Four supplementary tables and two supplementary figures are available with the online version of this paper.

 073965 © 2014 SGM *Printed in Great Britain*

inferred examination of the metabolic potential of uncultured microbial cells (Stepanauskas, 2012).

The Hanford Site in eastern Washington state, USA, was operated during the Second World War and Cold War for plutonium production, as part of the US nuclear weapons programme. The site is now mostly decommissioned, and work at the 300 Area (in the south-eastern corner of the site) (Fig. 1) has focused on the role of biogeochemical cycling in contaminant fate and transport (Peretyazhko *et al.*, 2012; Um *et al.*, 2010; Zachara *et al.*, 2013). The unconfined aquifer beneath the 300 Area is an oligotrophic environment characterized by oxic conditions in the upper Hanford Formation (O_2 concentrations are frequently measured at 10 mg l^{-1}) and limited nutrient concentrations (Lin *et al.*, 2012a). Dissolved organic carbon concentrations typically average around 1 mg l^{-1}. Despite these geochemical conditions, biomass concentrations of 10^4 to 10^5 cells ml^{-1} have been measured (J. Stegen, personal communication), with cell doubling times as fast as 2 h, similar to rates reported in other aquifer systems (Ayuso *et al.*, 2010). The aquifer is mainly present within highly transmissive Hanford Formation sediments, with the hydraulic gradient directly affected by the height stage of the nearby Columbia River (Bjornstad *et al.*, 2009). For the majority of the year, groundwater discharges into the Columbia River; however, under high river stage conditions associated with mountain snowpack runoff in spring, flow direction is reversed and Columbia River water intrudes into the subsurface and groundwater elevation increases. Using a 16S rRNA gene pyrosequencing approach, Lin *et al.* (2012b) tracked planktonic microbial communities within Hanford Formation groundwater over a nine-month sampling period. Dynamic shifts in microbial community structure were reported, often associated with seasonal Columbia River water intrusion (Lin *et al.*, 2012b). However, the absence of functional information for specific taxa prohibited a deeper understanding of the physical or biogeochemical controls that led to these temporal dynamics.

Within the data collected by Lin *et al.* (2012b), the relative abundance of *Pedobacter* 16S rRNA gene sequences ranged from ~0 % to ~14 %, suggesting that this genus constitutes a major component of Hanford sediment microbial communities (Fig. 2). Despite the isolation of more than 40

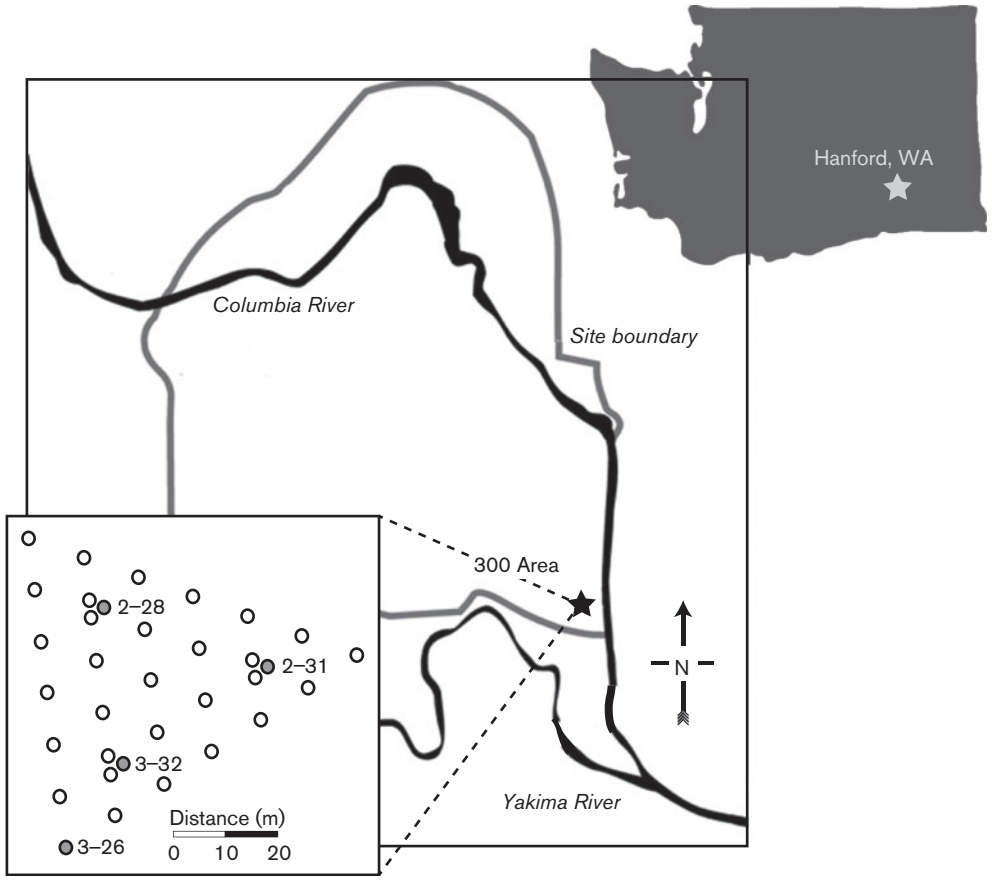

Fig. 1. Map of the US Department of Energy Hanford site, with the 300 Area located in the south-east corner of the site. Inset: the well layout at the Hanford Integrated Field Research Challenge (IFRC) site, where samples were obtained for these analyses (wells displayed in grey).

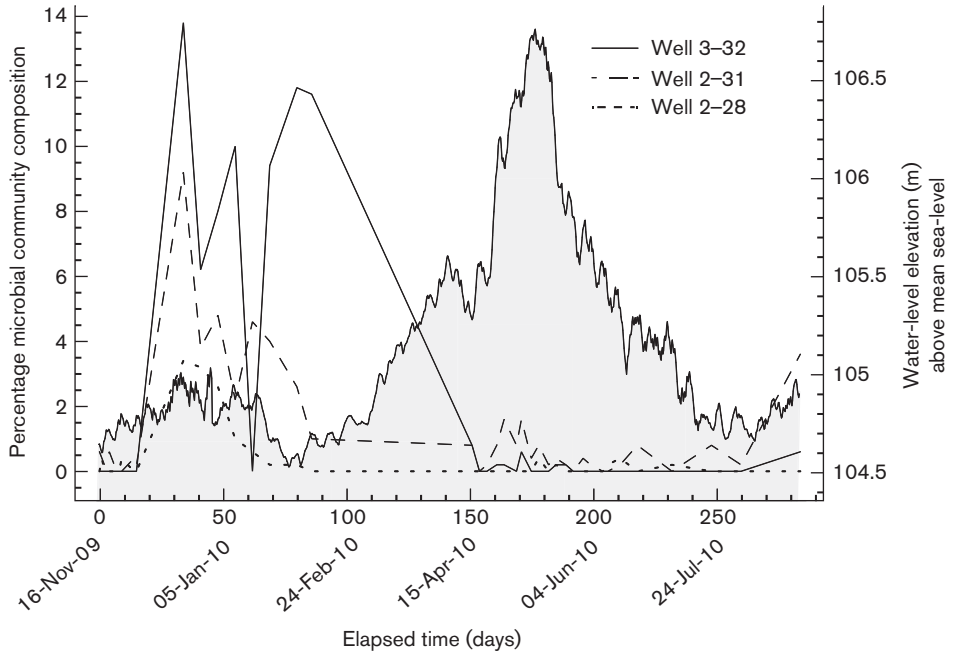

Fig. 2. Temporal abundances of *Pedobacter* 16S rRNA gene pyrotag sequences recovered from groundwater at the Hanford 300 Area. Samples were collected from three wells over a nine-month period. The shaded area indicates water table elevation in a nearby well, 3-26. Water table data modified from Lin *et al.* (2012b).

Pedobacter species from a diverse range of soil and sediment environments (An *et al.*, 2009; Gordon *et al.*, 2009; Hoang *et al.*, 2013; Jeon *et al.*, 2009; Jung *et al.*, 2012; Kwon *et al.*, 2011; Luo *et al.*, 2010; Margesin *et al.*, 2003; Margesin & Zhang, 2013; Muurholm *et al.*, 2007; Oh *et al.*, 2013; Roh *et al.*, 2008; Steyn *et al.*, 1998), little is known about the ecophysiology of these species. All *Pedobacter* species grow under aerobic, heterotrophic conditions, and have been shown to utilize a wide range of carbon compounds in the laboratory, including xylose, mannose, ribose and glucose. Interestingly, many of these species show ability to grow over a wide range of temperatures, from 1 to 37 °C (Margesin *et al.*, 2003; Margesin & Zhang, 2013; Zhou *et al.*, 2012b). Common genus characteristics include the production of the enzymes catalase and alkaline phosphatase. Building on these studies, we hypothesized that *Pedobacter* within the Hanford subsurface would also probably play a key role in carbon cycling. Given the oligotrophic nature of the 300 Area aquifer, we anticipated that these organisms might employ a range of strategies to maximize available substrates. Finally, a better understanding of the physiology and metabolism of these species would be used to better develop hypotheses linking shifts in *Pedobacter* abundances with geochemical trends associated with Columbia River water intrusion in the Hanford subsurface. To this end, we isolated four *Pedobacter* single cells from Hanford aquifer sediments using flow cytometry, and undertook their whole genome amplification. Subsequent sequencing and genome assembly enabled functional potential within these sequences to be inferred, revealing a potential metabolic role for these species in the Hanford subsurface, and aiding our understanding of the temporal dynamics of microbial populations in the Hanford subsurface.

METHODS

Cell selection. Hanford aquifer sediments recovered from a depth of 30 ft (9.1 m) in well 2-25 were mixed with a synthetic Hanford groundwater (Lin *et al.*, 2012a) containing 1 % pyrophosphate to aid cell desorption from mineral surfaces. Sediment was incubated in this material at room temperature under gently shaking conditions for 2 h. Samples of the liquid phase were removed, mixed with glycerol to a final concentration of 20 %, and flash frozen in liquid N_2 before being shipped on dry ice to the Single Cell Genomics Center at Bigelow Laboratory for Ocean Sciences.

FACS. The sample was vortexed vigorously for 30 s and centrifuged at 2000 r.p.m. 270 ***g*** for 1 min. The supernatant was diluted 1 : 10 with synthetic groundwater, stained with 5 μM (final concentration) SYTO-9 (Invitrogen) for 10–30 min and pre-screened through a 40 μm mesh-size cell strainer (BD). Microbial cells were sorted with a MoFlo (Beckman Coulter) flow cytometer equipped with a CyClone robotic arm for droplet deposition into 384-well plates. The cytometer was triggered on side scatter. Sorting gate was delineated using side scatter, green fluorescence (SYTO-9) and red fluorescence signals. The 'single 1 drop' mode was used for maximal sort purity, which ensures the absence of non-target particles within the target cell drop and the adjacent drops. Under these sorting conditions, sorted drops contain a few tens of picolitres of sample surrounding the target cell (Sieracki *et al.*, 2005), so non-target DNA is very low or absent. The accuracy of 10 μm fluorescent bead deposition into the 384-well plates was verified by microscopically examining the presence of beads in the plate wells. Of the two to three

plates examined each sort day, <2 % of wells were found not to contain a bead and only <0.5 % of wells were found to contain more than one bead, indicating very high purity of single cells. In addition, we verified the lack of DNA contamination in the sheath fluid and in sheath fluid lines by performing real-time multiple displacement amplification with the processed sheath fluid as the template.

Multiple displacement amplification (MDA). Bacterial cells were deposited into 384-well plates containing 0.6 µl per well of TE buffer. Plates were stored at −80 °C until further processing. The cells were lysed and their DNA was denatured using cold KOH (Raghunathan *et al.*, 2005). Genomic DNA from the lysed cells was amplified using MDA (Dean *et al.*, 2002) in a final volume of 10 µl. The MDA reactions contained 2 U Repliphi polymerase μl^{-1} (Epicentre), 1× reaction buffer (Epicentre), 0.4 mM each dNTP (Epicentre), 2 mM DTT (Epicentre), 50 mM random hexamers with the two 3′-terminal nucleotide bonds phosphorothioated (IDT) and 1 µM SYTO-9 (Invitrogen) (all final concentration). The MDA reactions were run at 30 °C for 12–16 h, and then inactivated by 15 min incubation at 65 °C. The amplified genomic DNA was stored at −80 °C until further processing. We refer to the MDA products originating from individual cells as SAGs. The instruments and the reagents were decontaminated for DNA prior to sorting and MDA setup, as previously described (Stepanauskas & Sieracki, 2007; Woyke *et al.*, 2011). Cell sorting and MDA setup were performed in a high-efficiency particulate absorption-filtered environment. As a quality control, the kinetics of each MDA reaction was monitored by measuring the SYTO-9 fluorescence using FLUOstar Omega (BMG). The critical point (Cp) was determined for each MDA reaction as the time required to produce half of the maximal fluorescence. The Cp is inversely correlated with the amount of DNA template (Zhang *et al.*, 2006). Of the 384 wells, 315 were dedicated for single cells, 66 were used as negative controls (no droplet deposition) and three received 10 cells each (positive controls). The Cp values were significantly lower in one-cell wells than in zero-cell wells ($P<0.05$; Wilcoxon two-sample test) in each microplate.

16S rRNA gene amplification. The MDA products were diluted 50-fold in sterile TE buffer. Then, 0.5 µl aliquots of the dilute MDA products served as templates in 5 µl real-time PCR screens targeting bacterial small subunit rRNA genes using primers 27F′ and 907R (Casamayor *et al.*, 2000; Lane, 1991). Forward (5′-GTAAAACGAC-GGCCAGT-3′) or reverse (5′-CAGGAAACAGCTATGACC-3′) M13 sequencing primer was appended to the 5′ end of each PCR primer to aid direct sequencing of the PCR products. All PCRs were performed using LightCycler 480 SYBR Green I Master mix (Roche) in a LightCycler 480 II real-time thermal cycler (Roche). The real-time PCR kinetics and the amplicon melting curves served as proxies detecting successful target gene amplification. New, 20 µl PCRs were set up for the PCR-positive SAGs and the amplicons were sequenced from both ends using M13 targets and Sanger technology by Beckman Coulter Genomics. Eleven of 17 PCR-positive SAGs were identified as *Pedobacter* species. Of these, four were chosen for further analysis.

SAG genome sequencing and assembly. SAG genomic DNA was sheared using an S220 focused ultrasonicator (Covaris) and gel-fractionated for 450 bp fragments. Illumina sequencing libraries were prepared using TruSeq reagents and protocols (Illumina). Then, 100 × 2 bp paired-end reads were sequenced using the Illumina HiSeq 2000 platform (Illumina). Twelve indexed SAG libraries were multiplexed, in equal proportions, into one lane of a flowcell. The amount of sequence data generated varied from 15 709 164 to 18 935 063 paired-end reads per SAG (see Table S1). For each sample, sequence data were quality trimmed (Q<0) and reads with greater than 5 % ambiguities were removed using NGS Toolkit (Patel & Jain, 2012). The sequence data were then digitally normalized to reduce over-represented and under-represented sequences (Swan *et al.*, 2011; Woyke *et al.*, 2011). Normalized reads were then assembled in SPAdes v.2.2.1 (Bankevich *et al.*, 2012) using the following settings: data flagged as single-cell, error correction max iterations 2 and iterative k-mer sizes 21,33,55. Assembled contigs shorter than 2000 bp were removed. Tetramer frequencies were extracted from the remaining contigs and principal components analysis was then used to extract the most important components of this high-dimensional feature matrix (Swan *et al.*, 2011; Woyke *et al.*, 2009). Contigs representing extremes on the first eight PCs were examined for their closest BLASTX hits against the NCBI non-redundant database to identify potential contaminant contigs that were then removed from the final assemblies. Between 3 and 6 % of base pairs per SAG were removed due to possible contamination. Statistics on the final assemblies can be found in Table S1 (available in the online Supplementary Material).

Single-cell sorting, whole-genome amplification, SAG identification and whole-genome assemblies and quality control were performed at the Bigelow Laboratory Single Cell Genomics Center (SCGC; www.bigelow.org/scgc). Genomic sequencing library preparation and sequencing were performed at the Oregon State University's Center for Genome Research (cgrb.oregonstate.edu). Our previous studies demonstrate the reliability of our methodology, with insignificant levels of DNA contamination (Stepanauskas & Sieracki, 2007; Swan *et al.*, 2011; Woyke *et al.*, 2009, 2011).

Genomic analyses. Gene calling and annotation of assembled SAGs were performed using the RAST server (http://rast.nmpdr.org/) (Aziz *et al.*, 2008). Protein localization information was determined using the PSORTb 3.0 software (http://www.psort.org/psortb/) (Yu *et al.*, 2010), while protein family information was obtained via Pfam searches (http://pfam.sanger.ac.uk/) (Punta *et al.*, 2012). Genome size was estimated using single-copy gene lists published by Raes *et al.* (2007). Orthologous proteins in other sequenced *Pedobacter* species were inferred using the SPOCS tool (Curtis *et al.*, 2013). Clustered regularly interspaced short palindromic repeat (CRISPR)-Cas sequences were identified using the CRISPRFinder tool (http://crispr.u-psud.fr/) (Grissa *et al.*, 2007). 3D protein structure predictions were generated by SWISS-MODEL based on protein alignment and secondary structure prediction (Schwede *et al.*, 2003). SWISS-MODEL is an automated protein homology-modelling server (Bordoli *et al.*, 2009). The alignment mode was utilized for a first approach based on a user-defined target–template alignment. Conservation of key catalytic residues and the secondary structure for each model were manually inspected. The four Hanford *Pedobacter* SAGs are publicly available at the US Department of Energy Integrated Microbial Genomes (IMG) site: http://img.jgi.doe.gov/.

RESULTS AND DISCUSSION

We analysed four single amplified genomes (SAGs) obtained from Hanford Site sediment: SCGC AC-286-M16, SCGC AC-286-M17, SCGC AC-286-P17 and SCGC AC-288-E13 (hereafter M16, M17, P17 and E13). Phylogenetic analysis of 16S rRNA gene sequences from the Hanford SAGs placed these organisms within the genus *Pedobacter*. Assembly size of these SAGs ranged from 0.8 to 2.4 Mb (Table S1). Genome completeness was estimated using conserved single copy gene identifications (Raes *et al.*, 2007), suggesting 20–75 % genome recovery. The SAGs exhibited 97.4–99.0 % similarity to each other at the 16S rRNA gene sequence level. Linking to common characteristics of cultured *Pedobacter* species, genes encoding catalase were detected in two SAGs, but no evidence of alkaline phosphatase was identified in the genomic data.

Results for the utilization of carbon compounds are presented below.

Carbohydrate utilization

Genomic reconstructions indicated that the Hanford *Pedobacter* SAGs are chemoheterotrophic, obligately aerobic bacteria, with the ability to degrade and oxidize a range of organic carbon compounds, and survive under both oxic and microaerophilic conditions. To survive in the relatively oligotrophic Hanford environment, the *Pedobacter* SAGs encode a wide range of mechanisms for degrading both complex and simple carbon substrates, and transporting these compounds into the cell. Carbohydrate-active enzymes accounted for approximately 5 % of the ORFs per SAG; a similar percentage is reported in the sequenced *Pedobacter* strains *Pedobacter saltans* DSM 12145 and *Pedobacter* sp. BAL39. Nearly half of these enzymes were classified within glycoside hydrolase (GH) families, and had functions associated with the degradation of polymers (xylan and chitin), as well as the metabolism of monosaccharides such as rhamnose, mannose and fucose (Fig. 3, Table 1). Other carbohydrate-active enzymes contained carbohydrate-binding modules, while others were classified within the polysaccharide lyase family. Protein localization inferences suggested that a subset of these enzymes are secreted into the extracellular environment where they can catalyse the depolymerization of particulate or mineral-adsorbed organic material. This pool of secreted enzymes may be larger; for each SAG, ~30–60 % of enzymes with unknown cellular localizations (Fig. 3) contained a signal peptide, suggesting a role in secretion from the cell.

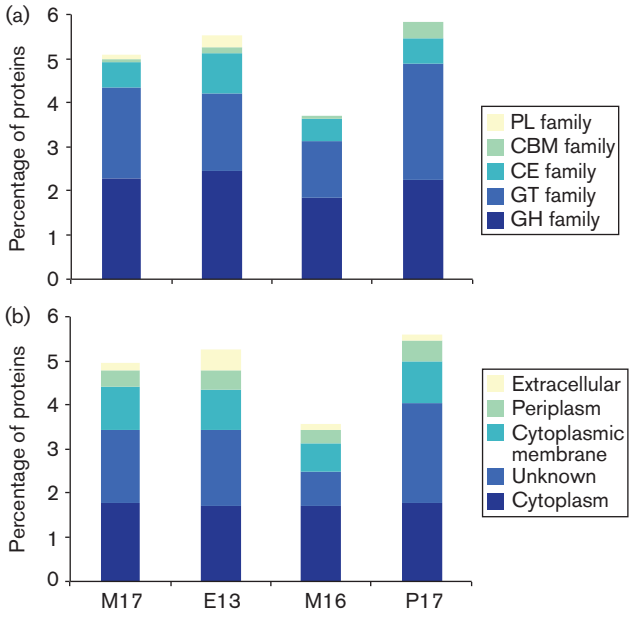

Fig. 3. Distribution of carbohydrate-active enzymes by enzyme class (a) and predicted cellular localization (b) across the four SAGs. PL, polysaccharide lyase; CBM, carbohydrate binding module; CE, carbohydrate esterase; GT, glycosyltransferase; GH, glycoside hydrolase.

Following the breakdown of more recalcitrant carbon substrates, *Pedobacter* micro-organisms must transport labile carbohydrates into the cell for metabolism. This may be achieved via a series of mechanisms that enable efficient carbon uptake in the oligotrophic Hanford environment. The *Pedobacter* SAGs contain large numbers of ABC-transporter systems, permease proteins, and TonB-dependent receptor (TBDR) proteins. While TBDRs have traditionally been implicated in Fe(III) and vitamin B_{12} uptake (Chimento *et al.*, 2005; Postle & Kadner, 2003), recent work has elucidated a role for TBDRs in the transport of sugars across the outer membrane of Gram-negative bacteria (Blanvillain *et al.*, 2007; Neugebauer *et al.*, 2005). Blanvillain *et al.* (2007) identified a small but phylogenetically diverse group of bacteria that over-represented TNBRs within their genomes. As a common characteristic of this group, all shared the ability to degrade complex carbon sources. The *Pedobacter* SAGs fall within this group, with 18 TNBRs (proteins containing Pfam domains PF07715 and PF00593) encoded within the most complete M17 SAG genome. The TBDR/Mbp genome size ratio of 7.5 places SAG M17 clearly within the 'over-represented' group (Blanvillain *et al.*, 2007). For comparison, across the microbial groups studied by Blanvillain *et al.* (2007) 27 % of bacteria had no TBDR proteins, and 43 % possessed between one and 13 TBDRs. Interestingly, over-representation of TNBRs is also observed in two additional sequenced *Pedobacter* taxa, *Pedobacter* sp. BAL39 and *P. saltans*, suggesting that this trend may apply to *Pedobacter* generally. Although specific substrates for many of these transport systems cannot be elucidated from genomic sequence alone, genes encoding multiple periplasmic maltose/maltodextrin binding proteins were identified, together with fucose permease proteins that were predicted to localize to the cytoplasmic membrane.

All the enzymes necessary for the degradation and utilization of xylan were detected within the SAGs (Fig. 4), including endo-1,4-xylanases for the breakdown of xylan into xylose subunits, and xylose isomerase for the conversion of xylose into xylulose. An encoded extracellular solute binding protein has ~30 % amino acid similarity to the xylo-oligosaccharide binding XynE protein in *Geobacillus stearothermophilus* (Shulami *et al.*, 2007), suggesting a role for this protein in xylan/xylose uptake into the cell. Xylulose generated via these reactions is subsequently metabolized via the pentose phosphate and Embden–Meyerhof–Parnas pathways. Xylan degrading potential was also detected in *P. saltans* genomic data, but not *Pedobacter* sp. BAL39.

Protein orthologue searches across multiple *Pedobacter* genomes were performed to identify patterns and distributions of carbohydrate-active enzymes; results indicated that only five GH enzymes were unique to the Hanford SAGs,

Table 1. Carbohydrate-active enzymes encoded across four *Pedobacter* SAGs from the Hanford 300 Area subsurface

Defined activity	Pfam domain	Carbohydrate Active enZYmes (CAZY) family	Hanford SAG			
			M17	E13	M16	P17
Cellulose degradation						
β-Glucosidase	PF00933	GH3	1	1	0	0
Cellulase	PF00150	GH5	2	1	0	1
Endo-1,6-β-glucanase	PF02055	GH30	2	0	0	0
Debranching enzymes						
1,4-α-Glucan branching enzyme	PF00128	GH13	1	2	2	0
α-Glucosidase	PF01055, PF01261, PF04909	GH31, GH97	2	1	1	0
α-L-Arabinofuranosidase	PF04616	GH43	2	2	2	2
α-L-Rhamnosidase	PF05592	GH78	3	1	2	0
Endohemicellulases						
Endo-1,4-β-xylanase	PF00331, PF04616	GH10, GH43	1	3	0	0
Amylolytic enzymes						
α-Amylase	PF00128	GH13, GH70	1	1	1	1
Amylomaltase	PF02446, PF00128	GH13, GH77	1	1	1	1
Glucoamylase	PF00723	GH15	1	0	2	0
Oligosaccharide-degrading enzymes						
α-1,2-Mannosidase	PF07971	GH92	4	0	0	1
α-1,6-Mannanase	PF03663	GH76	3	0	1	0
α-L-Fucosidase	PF00107, PF01120	GH29	1	3	0	1
β-Galactosidase	PF00703, PF01301, PF02836, PF02929	GH2, GH35	5	2	0	0
β-Glucanase	PF00722	GH16	0	1	1	0
β-Hexosaminidase	PF00728, PF00933	GH3, GH20	1	1	3	2
β-Mannosidase	PF00703	GH2	1	0	0	1
β-Xylosidase	PF04616	GH43	0	4	0	0
Glucuronyl hydrolase	PF05147	GH88	0	1	0	0
Other glycoside hydrolases	PF02055	GH30	0	1	0	1
Predicted glycoside hydrolase	PF00759, PF04041, PF06439, PF06441, PF07944	GH9, GH125, GH127, GH130	4	3	2	2
Other degrading enzymes						
β-Lactamase	PF00144	GH3	4	1	4	0
Chitinase	PF00704	GH18	1	2	0	1
Glucosaminidase	PF00144, PF01476	GH3, GH73	1	2	0	2
Malto-oligosyltrehalose synthase	PF00128	GH13	1	0	1	0
Malto-oligosyltrehalose trehalohydrolase	PF00128	GH13	1	1	1	1
Maltose phosphorylase	PF03632	GH65	0	1	0	0

while 23 different kinds of GH enzymes found in SAG genomes had orthologues in two other sequenced *Pedobacter* taxa, *P. saltans* and *Pedobacter* sp. BAL39 (Table S2). These species were isolated from soil (Steyn *et al.*, 1998) and seawater (Pinhassi *et al.*, 1997), respectively, and further support the inference that the capacity for complex carbon and sugar degradation is a conserved trait across these species. In this study, however, the relatively high number of both carbohydrate-active enzymes and mechanisms for carbon substrate uptake encoded within the Hanford SAGs further supports the hypothesis that these mechanisms enable heterotrophic bacteria to thrive within the oligotrophic Hanford subsurface.

Orthologue searches also revealed a large number of hypothetical genes unique to the Hanford SAGs (Table S3); of 549 orthologous genes unique to the SAGs (i.e. present across two or more SAGs, but not the two sequenced *Pedobacter* species), 400 (~72 %) were annotated as hypothetical genes. These trends were further pronounced within patterns of genes present in only one genome; of 374 non-orthologous genes in Hanford SAG M17, 232 (~62 %) were annotated as hypothetical genes. While these suites of genes may play key roles in shaping ecology and physiology within these species, the role for this 'biological dark matter' requires detailed experimental analyses (Kolker *et al.*, 2005).

Defence mechanisms

Genes encoding orthologous CRISPR-associated Cas proteins (Cas1, Cas2 and Cas9) were detected across three of the SAGs (M17, M16 and E13) but were absent from

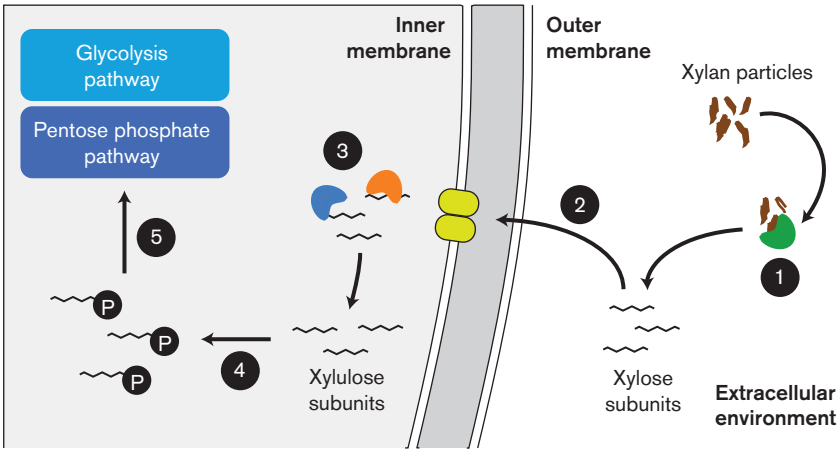

Fig. 4. Inferred mechanism for extracellular xylan degradation and subsequent uptake in Hanford *Pedobacter* SAGs. (1) Activity of extracellular xylanase. (2) Binding of xylose subunits by protein with similarity to xylo-oligosaccharide binding protein XynE, and transport into cytoplasm via ABC transporter. (3) Activity of α-*N*-arabinofuranosidase and xylose isomerase. (4) Phosphorylation of D-xylulose by xylulose kinase. (5) Metabolism of D-xylulose by pentose phosphate and Embden–Meyerhof–Parnas pathways.

P. saltans and *Pedobacter* sp. BAL39, indicating the development of viral defence mechanisms for these species within the Hanford aquifer. From the operon structure in all three SAGs (*cas9–cas1–cas2*) we can infer that these Cas proteins form a type II-A or II-B system (Makarova *et al.*, 2011), characterized by the presence of Cas9, which generates CRISPR RNA and cleaves target DNA. CRISPR regions were identified in SAG contigs, with numbers of spacers ranging from four to 47. Across the three SAGs, 70 different spacer sequences were identified (Table S4); only four were shared across all three SAGs, while 40 were shared between SAGs M16 and E13. SAG M17 contained 20 unique sequences (Fig. 5). Of these shared spacer sequences between M16 and E13, the majority are present in a conserved order across the two SAGs. Single nucleotide polymorphisms are responsible for unique spacer sequences in M16 and E13 (Fig. 5, Table S4). From an evolutionary perspective, the pattern and ordering of spacer distribution suggest that SAGs M16 and E13 have a far more recent common ancestor than M17; the high number of unique spacers in M17 indicates that the lineage for this SAG incorporated mostly different foreign DNA into genomic material, relative to the other SAGs considered in this study. From an environmental perspective, the presence and distribution of these genes in the Hanford single cells suggest that viral infection is an active process within the aquifer, and may be a factor controlling microbial population dynamics.

Central metabolism

Following the uptake of carbon substrates into the cell, central metabolism in the Hanford SAGs is inferred to proceed through a complete tricarboxylic acid cycle, with acetyl-CoA generated from pyruvate via a pyruvate dehydrogenase complex, or from acetate via the combined action of acetate kinase and phosphate acetyltransferase. The presence of a 14-subunit NADH dehydrogenase complex, succinate dehydrogenase, F_0F_1-type ATP synthase and cytochrome *c* oxidase is evidence of an electron transport chain for generation of ATP via oxidative phosphorylation. Present in all the SAGs studied here are copies of lactate dehydrogenase (EC 1.1.1.28), which catalyses the reversible reaction between pyruvate and lactate. The presence of this gene offers two metabolic possibilities, the first being that lactate is imported into the cell from the local environment, and converted to pyruvate for biosynthetic

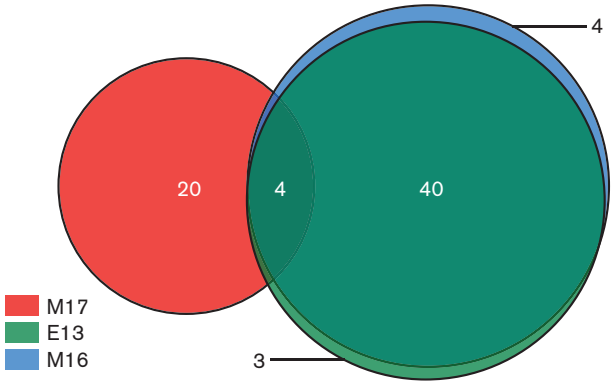

Fig. 5. CRISPR spacer distribution across three Hanford *Pedobacter* SAGs. Values refer to individual spacer sequences present in each genome, and illustrate the high similarity in spacer composition between E13 and P16, and the dissimilarity of spacer sequences in M17.

reactions. The absence of annotated lactate permeases in the SAGs would appear to discount the possibility of lactate uptake into the cell, although compound-specific permeases are difficult to discern by annotation alone. The second possibility is that the Hanford *Pedobacter* are facultatively aerobic, with the ability to ferment carbon under anaerobic conditions, such as might be found in microenvironments in the Hanford subsurface. As part of a fermentative lifestyle, lactate dehydrogenase is a critical enzyme, coupling the oxidation of NADH to the reduction of pyruvate. However, other genes that are frequently present in fermenters are not found within the Hanford SAGs; pyruvate ferredoxin oxidoreductase or pyruvate formate lyase both convert pyruvate to acetyl-CoA without the generation of NADH, and as such are favoured by fermenters (Wrighton *et al.*, 2012). Within the Hanford SAGs, however, the only mechanism identified for converting pyruvate to acetyl-CoA is via pyruvate dehydrogenase, which generates reducing power as NADH. While some fermentative micro-organisms are able to dispose of this reducing power via the action of hydrogenases, none of these enzymes are detected within the SAGs. However, note that other fermentative micro-organisms that use pyruvate dehydrogenase as the primary mechanism for conversion of pyruvate to acetyl-CoA have been identified within metagenomic data (Di Rienzi *et al.*, 2013). These observations have not been confirmed experimentally. In light of these observations, and the metabolic repertoire of cultured *Pedobacter* species (e.g. *Pedobacter bauzanensis*) that have not been reported to ferment glucose (Muurholm *et al.*, 2007; Zhang *et al.*, 2010), it is more likely that lactate is utilized as a carbon source for biosynthesis within the SAGs. However, the difficulty in identifying fermentative metabolism in chemoheterotrophic bacteria means that in this instance we cannot positively identify a role for lactate dehydrogenase in the SAGs.

Oxygen reduction

Dissolved O_2 concentrations in 300 Area Hanford Formation groundwater are typically at saturation (~10 mg l^{-1}). Despite this, the heterogeneous nature of the Hanford subsurface can result in the development of diffusion-dominated microenvironments (Nguyen *et al.*, 2012), where microaerophilic conditions may exist. Genomic reconstructions of *Pedobacter* SAGs suggest that the organisms are unable to grow anaerobically, consistent with their detection only from the oxic Hanford Formation. However, genes encoding both aa_3-type and cbb_3-type cytochrome *c* oxidases were identified across the SAGs, and two additional sequenced *Pedobacter* taxa, *P. saltans* and *Pedobacter* sp. BAL39. The aa_3-type cytochrome *c* oxidases are typically found in aerobic bacteria, and have an O_2 substrate affinity (K_m) approximately six- to eightfold higher than the cbb_3-type oxidases (Morris & Schmidt, 2013). The cbb_3-type oxidases have therefore been implicated in aerobic respiration under microaerophilic conditions; indeed, an O_2 K_m value of 7 nM was calculated for a cbb_3 oxidase protein in *Bradyrhizobium japonicum* (Preisig *et al.*, 1996). Multiple lines of evidence were used to identify the aa_3-type cytochrome *c* oxidase; protein modelling of subunit I identified the six histidines (H67, H260, H309, H310, H395 and H397) involved in binding the haem-copper binuclear centre, while the detection of a binuclear copper centre (CuA) in subunit II suggests that this oxygen reductase is a cytochrome *c* oxidase rather than a quinol oxidase (as this centre is absent in quinol oxidases) (Fig. S1). Further analysis of subunit II was unable to identify a *c*-type haem motif as retrieved in caa_3-type oxidases. *P. saltans* also encodes an alternative complex III (ACIII) that functionally replaces the bc_1 complex (complex III in the oxidative phosphorylation chain) (Pereira *et al.*, 2007). As observed in *Rhodothermus marinus* (Refojo *et al.*, 2010), ACIII genes in *P. saltans* cluster with the low oxygen affinity aa_3-type cytochrome *c* oxidase and thus probably form a functional association under high oxygen concentrations.

The cbb_3-type oxidase was identified via analysis of the CcoN, CcoO and CcoP subunits required for catalytic activity. The CcoN subunit corresponds to the catalytic subunit homologous to the subunit I of the aa_3 type oxidase, while the CcoO and CcoP subunits are membrane-bound mono- (or di-) haem cytochromes *c*. The CcoN and CcoO subunits were fused into one subunit in the Hanford SAGs; similar observations have been reported in members of the *Bacteroides*, *Planctomycetes* and *Verrucomicrobia* (Ducluzeau *et al.*, 2008). A single *c*-type haem-binding motif ($C_{204}xxC_{207}H_{208}$) was detected in the CcoP subunits, while protein modelling identified haem *c* (Cys_{550} X_2 C_{553} His_{554} Met_{622}) signatures from the CcoO subunits. Finally, conserved histidine residues binding the haem-copper binuclear centre including haems *b* and b_3 (H344 and H346, and H60, H206, H256 and H257) were identified in the CcoN subunits (Buschmann *et al.*, 2010) (Fig. S2). Potential for microaerophilic growth in *Pedobacter* species is further supported by the presence of genes encoding a cytochrome *bd*-type quinol oxidase in SAG M16, *P. saltans* and *Pedobacter* sp. BAL39. Previous studies with this enzyme have reported an exceptionally high affinity for oxygen (Puustinen *et al.*, 1991), and highest expression levels under microaerophilic growth conditions (Fu *et al.*, 1991; Rice & Hempfling, 1978).

These comparative analyses suggest that potential for microaerophilic growth may be a conserved trait across *Pedobacter* species, with this respiratory versatility probably aiding growth of these species across the wide range of environments from which they have been isolated. For the SAGs analysed here, these data suggest that *Pedobacter* strains growing within the Hanford Formation may have the ability to survive over a range of O_2 concentrations, utilizing different terminal oxidases for fully aerobic or microaerophilic conditions. Coupled to the potential utilization of a wide range of carbon substrates, this trait increases the metabolic flexibility of Hanford *Pedobacter* strains and could enable active metabolism across a range of microenvironments in the 300 Area aquifer.

Potential links to local hydrology

Temporal relative abundances for *Pedobacter* species in the Hanford 300 Area subsurface were generally highest between December and April, before the effects of snow-melt resulted in Columbia River stage change, river water intrusion into the subsurface and a water table rise of ~1.5 m (Fig. 2) (Lin *et al.*, 2012b). Biogeochemical shifts (e.g. nutrients, carbon influx) associated with river water intrusion into the subsurface may therefore have had a deleterious effect on *Pedobacter* populations. It has been hypothesized that changes in organic carbon associated with river water intrusion may drive some microbial community shifts; given that dissolved organic carbon concentrations in Columbia River water (~2 mg l^{-1}) are generally double those for Hanford groundwater, one might anticipate increases in abundances of r-strategist (fast growth rate) bacteria as more carbon is introduced into the subsurface system. This was clearly not the case for *Pedobacter* populations identified by Lin *et al.* (2012b), suggesting that they are K-strategists that succeed under resource-limited conditions (Freilich *et al.*, 2010). Although originally applied to animals and plants, the r/K selection theory may play a role in understanding microbial dynamics in environments experiencing nutrient and resource fluxes, such as the 300 Area aquifer during Columbia River water intrusion. Increased flux of organic carbon or the introduction of a *specific* carbon compound during this event may offer an opportunity for fast-growing bacterial species to out-compete potential K-strategists such as *Pedobacter* species. The ability of specific bacterial groups to rapidly increase in abundance in response to changing organic carbon flux has been previously demonstrated in an analogous environment (Wilkins *et al.*, 2013). Better characterizing chemical (carbon, sulfur, nitrogen species) shifts associated with Columbia River water intrusion into the Hanford aquifer will therefore be essential for identifying geochemical drivers of microbial community change. Other potential variables linked to river intrusion that might negatively impact indigenous micro-organisms include the introduction of foreign microbial species or predators into the ecosystem, or as-yet-unidentified factors.

Here we were able to isolate single cells from complex mixtures of gravel, silt and clay sediments using single-cell manipulation techniques. Amplification and sequencing of genomic DNA from targeted species allowed the physiology and ecology of *Pedobacter* cells to be inferred. These results revealed a relatively versatile lifestyle for these species within the Hanford aquifer, with roles in carbon cycling and the potential colonization of microaerophilic microenvironments, such as might be found in heterogeneous subsurface aquifers.

ACKNOWLEDGEMENTS

This research was supported by the US Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of the Subsurface Biogeochemistry Research Program's Scientific Focus Area (SFA) at Pacific North-west National Laboratory (PNNL). PNNL is operated for DOE by Battelle under contract DE-AC06-76RLO 1830. We thank two anonymous reviewers for their helpful comments and suggestions.

REFERENCES

An, D.-S., Kim, S.-G., Ten, L. N. & Cho, C.-H. (2009). *Pedobacter daechungensis* sp. nov., from freshwater lake sediment in South Korea. *Int J Syst Evol Microbiol* **59**, 69–72.

Anderson, R. T., Vrionis, H. A., Ortiz-Bernad, I., Resch, C. T., Long, P. E., Dayvault, R., Karp, K., Marutzky, S., Metzler, D. R. & other authors (2003). Stimulating the in situ activity of *Geobacter* species to remove uranium from the groundwater of a uranium-contaminated aquifer. *Appl Environ Microbiol* **69**, 5884–5891.

Ayuso, S. V., Lopez-Archilla, A. I., Montes, C. & Guerrero, M. C. (2010). Microbial activities in a coastal, sandy aquifer system (Donana Natural Protected Area, SW Spain). *Geomicrobiol J* **27**, 409–423.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M. & other authors (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S. & other authors (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455–477.

Bjornstad, B., Horner, J., Vermeul, V., Lanigan, D. & Thorne, P. (2009). Borehole completion and conceptual hydrogeologic model for the IFRC Well Field, 300 Area, Hanford Site. Richland, WA: Pacific Northwest National Laboratory.

Blanvillain, S., Meyer, D., Boulanger, A., Lautier, M., Guynet, C., Denancé, N., Vasse, J., Lauber, E. & Arlat, M. (2007). Plant carbohydrate scavenging through tonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. *PLoS ONE* **2**, e224.

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. & Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* **4**, 1–13.

Buschmann, S., Warkentin, E., Xie, H., Langer, J. D., Ermler, U. & Michel, H. (2010). The structure of cbb3 cytochrome oxidase provides insights into proton pumping. *Science* **329**, 327–330.

Casamayor, E. O., Schäfer, H., Bañeras, L., Pedrós-Alió, C. & Muyzer, G. (2000). Identification of and spatio-temporal differences between microbial assemblages from two neighboring sulfurous lakes: comparison by microscopy and denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **66**, 499–508.

Chimento, D. P., Kadner, R. J. & Wiener, M. C. (2005). Comparative structural analysis of TonB-dependent outer membrane transporters: implications for the transport cycle. *Proteins* **59**, 240–251.

Curtis, D. S., Phillips, A. R., Callister, S. J., Conlan, S. & McCue, L. A. (2013). SPOCS: software for predicting and visualizing orthology/paralogy relationships among genomes. *Bioinformatics* **29**, 2641–2642.

Dean, F. B., Hosono, S., Fang, L. H., Wu, X. H., Faruqi, A. F., Bray-Ward, P., Sun, Z. Y., Zong, Q. L., Du, Y. F. & other authors (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A* **99**, 5261–5266.

Di Rienzi, S. C., Sharon, I., Wrighton, K. C., Koren, O., Hug, L. A., Thomas, B. C., Goodrich, J. K., Bell, J. T., Spector, T. D. & other authors (2013). The human gut and subsurface harbor non-photosynthetic Cyanobacteria. *eLife* **2**, e01102.

Ducluzeau, A.-L., Ouchane, S. & Nitschke, W. (2008). The cbb3 oxidases are an ancient innovation of the domain bacteria. *Mol Biol Evol* **25**, 1158–1166.

Freilich, S., Kreimer, A., Meilijson, I., Gophna, U., Sharan, R. & Ruppin, E. (2010). The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Res* **38**, 3857–3868.

Fu, H. A., Iuchi, S. & Lin, E. C. C. (1991). The requirement of ArcA and Fnr for peak expression of the cyd operon in *Escherichia coli* under microaerobic conditions. *Mol Gen Genet* **226**, 209–213.

Ginige, M. P., Kaksonen, A. H., Morris, C., Shackelton, M. & Patterson, B. M. (2013). Bacterial community and groundwater quality changes in an anaerobic aquifer during groundwater recharge with aerobic recycled water. *FEMS Microbiol Ecol* **85**, 553–567.

Gordon, N. S., Valenzuela, A., Adams, S. M., Ramsey, P. W., Pollock, J. L., Holben, W. E. & Gannon, J. E. (2009). *Pedobacter nyackensis* sp. nov., *Pedobacter alluvionis* sp. nov. and *Pedobacter borealis* sp. nov., isolated from Montana flood-plain sediment and forest soil. *Int J Syst Evol Microbiol* **59**, 1720–1726.

Grissa, I., Vergnaud, G. & Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* **35** (Web Server issue), W52–W57.

Hemme, C. L., Deng, Y., Gentry, T. J., Fields, M. W., Wu, L., Barua, S., Barry, K., Tringe, S. G., Watson, D. B. & other authors (2010). Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *ISME J* **4**, 660–672.

Hoang, V.-A., Kim, Y.-J., Nguyen, N. L., Min, J.-W. & Yang, D.-C. (2013). *Pedobacter ginsengiterrae* sp. nov., isolated from soil of a ginseng field. *Int J Syst Evol Microbiol* **63**, 1273–1279.

Islam, F. S., Gault, A. G., Boothman, C., Polya, D. A., Charnock, J. M., Chatterjee, D. & Lloyd, J. R. (2004). Role of metal-reducing bacteria in arsenic release from Bengal delta sediments. *Nature* **430**, 68–71.

Jeon, Y., Kim, J. M., Park, J. H., Lee, S. H., Seong, C.-N., Lee, S.-S. & Jeon, C. O. (2009). *Pedobacter oryzae* sp. nov., isolated from rice paddy soil. *Int J Syst Evol Microbiol* **59**, 2491–2495.

Jung, Y.-T., Lee, S.-Y., Choi, W.-C., Oh, T.-K. & Yoon, J.-H. (2012). *Pedobacter boryungensis* sp. nov., isolated from soil. *Int J Syst Evol Microbiol* **62**, 13–17.

Kolker, E., Picone, A. F., Galperin, M. Y., Romine, M. F., Higdon, R., Makarova, K. S., Kolker, N., Anderson, G. A., Qiu, X. & other authors (2005). Global profiling of *Shewanella oneidensis* MR-1: expression of hypothetical genes and improved functional annotations. *Proc Natl Acad Sci U S A* **102**, 2099–2104.

Kwon, S.-W., Son, J.-A., Kim, S.-J., Kim, Y.-S., Park, I.-C., Bok, J.-I. & Weon, H.-Y. (2011). *Pedobacter rhizosphaerae* sp. nov. and *Pedobacter soli* sp. nov., isolated from rhizosphere soil of Chinese cabbage (*Brassica campestris*). *Int J Syst Evol Microbiol* **61**, 2874–2879.

Lane, D. J. (1991). 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematics*, pp. 115–175. Edited by E. Stackebrandt & M. Goodfellow. Chichester: Wiley.

Lin, X., Kennedy, D., Peacock, A., McKinley, J., Resch, C. T., Fredrickson, J. & Konopka, A. (2012a). Distribution of microbial biomass and potential for anaerobic respiration in Hanford Site 300 Area subsurface sediment. *Appl Environ Microbiol* **78**, 759–767.

Lin, X., McKinley, J., Resch, C. T., Kaluzny, R., Lauber, C. L., Fredrickson, J., Knight, R. & Konopka, A. (2012b). Spatial and temporal dynamics of the microbial community in the Hanford unconfined aquifer. *ISME J* **6**, 1665–1676.

Luo, X., Wang, Z., Dai, J., Zhang, L., Li, J., Tang, Y., Wang, Y. & Fang, C. (2010). *Pedobacter glucosidilyticus* sp. nov., isolated from dry riverbed soil. *Int J Syst Evol Microbiol* **60**, 229–233.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J. M., Wolf, Y. I. & other authors (2011). Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9**, 467–477.

Margesin, R. & Zhang, D.-C. (2013). *Pedobacter ruber* sp. nov., a psychrophilic bacterium isolated from soil. *Int J Syst Evol Microbiol* **63**, 339–344.

Margesin, R., Spröer, C., Schumann, P. & Schinner, F. (2003). *Pedobacter cryoconitis* sp. nov., a facultative psychrophile from alpine glacier cryoconite. *Int J Syst Evol Microbiol* **53**, 1291–1296.

Maymó-Gatell, X., Chien, Y., Gossett, J. M. & Zinder, S. H. (1997). Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. *Science* **276**, 1568–1571.

Morris, R. L. & Schmidt, T. M. (2013). Shallow breathing: bacterial life at low O_2. *Nat Rev Microbiol* **11**, 205–212.

Muurholm, S., Cousin, S., Päuker, O., Brambilla, E. & Stackebrandt, E. (2007). *Pedobacter duraquae* sp. nov., *Pedobacter westerhofensis* sp. nov., *Pedobacter metabolipauper* sp. nov., *Pedobacter hartonius* sp. nov. and *Pedobacter steynii* sp. nov., isolated from a hard-water rivulet. *Int J Syst Evol Microbiol* **57**, 2221–2227.

Neugebauer, H., Herrmann, C., Kammer, W., Schwarz, G., Nordheim, A. & Braun, V. (2005). ExbBD-dependent transport of maltodextrins through the novel MalA protein across the outer membrane of *Caulobacter crescentus*. *J Bacteriol* **187**, 8300–8311.

Nguyen, H. D., Cao, B., Mishra, B., Boyanov, M. I., Kemner, K. M., Fredrickson, J. K. & Beyenal, H. (2012). Microscale geochemical gradients in Hanford 300 Area sediment biofilms and influence of uranium. *Water Res* **46**, 227–234.

Oh, H.-W., Kim, B.-C., Park, D.-S., Jeong, W.-J., Kim, H., Lee, K. H. & Kim, S. U. (2013). *Pedobacter luteus* sp. nov., isolated from soil. *Int J Syst Evol Microbiol* **63**, 1304–1310.

Patel, R. K. & Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* **7**, e30619.

Pereira, M. M., Refojo, P. N., Hreggvidsson, G. O., Hjorleifsdottir, S. & Teixeira, M. (2007). The alternative complex III from *Rhodothermus marinus* – a prototype of a new family of quinol:electron acceptor oxidoreductases. *FEBS Lett* **581**, 4831–4835.

Peretyazhko, T. S., Zachara, J. M., Kukkadapu, R. K., Heald, S. M., Kutnyakov, I. V., Resch, C. T., Arey, B. W., Wang, C. M., Kovarik, L. & other authors (2012). Pertechnetate (TcO_4^-) reduction by reactive ferrous iron forms in naturally anoxic, redox transition zone sediments from the Hanford Site, USA. *Geochim Cosmochim Acta* **92**, 48–66.

Pinhassi, J., Zweifel, U. L. & Hagström, A. (1997). Dominant marine bacterioplankton species found among colony-forming bacteria. *Appl Environ Microbiol* **63**, 3359–3366.

Postle, K. & Kadner, R. J. (2003). Touch and go: tying TonB to transport. *Mol Microbiol* **49**, 869–882.

Preisig, O., Zufferey, R., Thöny-Meyer, L., Appleby, C. A. & Hennecke, H. (1996). A high-affinity *cbb*$_3$-type cytochrome oxidase terminates the symbiosis-specific respiratory chain of *Bradyrhizobium japonicum*. *J Bacteriol* **178**, 1532–1538.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G. & other authors (2012). The Pfam protein families database. *Nucleic Acids Res* **40** (Database issue), D290–D301.

Puustinen, A., Finel, M., Haltia, T., Gennis, R. B. & Wikström, M. (1991). Properties of the two terminal oxidases of *Escherichia coli*. *Biochemistry* **30**, 3936–3942.

Raes, J., Korbel, J. O., Lercher, M. J., von Mering, C. & Bork, P. (2007). Prediction of effective genome size in metagenomic samples. *Genome Biol* **8**, R10.

Raghunathan, A., Ferguson, H. R., Jr, Bornarth, C. J., Song, W. M., Driscoll, M. & Lasken, R. S. (2005). Genomic DNA amplification from a single bacterium. *Appl Environ Microbiol* **71**, 3342–3347.

Refojo, P. N., Teixeira, M. & Pereira, M. M. (2010). The alternative complex III of *Rhodothermus marinus* and its structural and functional association with *caa*$_3$ oxygen reductase. *Biochim Biophys Acta* **1797**, 1477–1482.

Rice, C. W. & Hempfling, W. P. (1978). Oxygen-limited continuous culture and respiratory energy conservation in *Escherichia coli*. *J Bacteriol* **134**, 115–124.

Roh, S. W., Quan, Z.-X., Nam, Y.-D., Chang, H.-W., Kim, K.-H., Kim, M.-K., Im, W.-T., Jin, L., Kim, S.-H. & other authors (2008). *Pedobacter agri* sp. nov., from soil. *Int J Syst Evol Microbiol* **58**, 1640–1643.

Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* **31**, 3381–3385.

Shulami, S., Zaide, G., Zolotnitsky, G., Langut, Y., Feld, G., Sonenshein, A. L. & Shoham, Y. (2007). A two-component system regulates the expression of an ABC transporter for xylo-oligosaccharides in *Geobacillus stearothermophilus*. *Appl Environ Microbiol* **73**, 874–884.

Sieracki, M., Poulton, N. & Crosbie, N. (2005). Automated isolation techniques for microalgae. Chapter 7, 101-116. In *Algal Culturing Techniques*. Edited by R. Anderson. New York: Elsevier Academic.

Stepanauskas, R. (2012). Single cell genomics: an individual look at microbes. *Curr Opin Microbiol* **15**, 613–620.

Stepanauskas, R. & Sieracki, M. E. (2007). Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci U S A* **104**, 9052–9057.

Steyn, P. L., Segers, P., Vancanneyt, M., Sandra, P., Kersters, K. & Joubert, J. J. (1998). Classification of heparinolytic bacteria into a new genus, *Pedobacter*, comprising four species: *Pedobacter heparinus* comb. nov., *Pedobacter piscium* comb. nov., *Pedobacter africanus* sp. nov. and *Pedobacter saltans* sp. nov. Proposal of the family Sphingobacteriaceae fam. nov. *Int J Syst Bacteriol* **48**, 165–177.

Swan, B. K., Martinez-Garcia, M., Preston, C. M., Sczyrba, A., Woyke, T., Lamy, D., Reinthaler, T., Poulton, N. J., Masland, E. D. P. & other authors (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**, 1296–1300.

Um, W., Zachara, J. M., Liu, C., Moore, D. A. & Rod, K. A. (2010). Resupply mechanism to a contaminated aquifer: a laboratory study of U(VI) desorption from capillary fringe sediments. *Geochim Cosmochim Acta* **74**, 5155–5170.

Wilkins, M. J., Wrighton, K. C., Nicora, C. D., Williams, K. H., McCue, L. A., Handley, K. M., Miller, C. S., Giloteaux, L., Montgomery, A. P. & other authors (2013). Fluctuations in species-level protein expression occur during element and nutrient cycling in the subsurface. *PLoS ONE* **8**, e57819.

Williams, K. H., Long, P. E., Davis, J. A., Wilkins, M. J., N'Guessan, A. L., Steefel, C. I., Yang, L., Newcomer, D. R., Spane, F. A. & other authors (2011). Acetate availability and its influence on sustainable bioremediation of uranium-contaminated groundwater. *Geomicrobiol J* **28**, 519–539.

Woyke, T., Xie, G., Copeland, A., González, J. M., Han, C., Kiss, H., Saw, J. H., Senin, P., Yang, C. & other authors (2009). Assembling the marine metagenome, one cell at a time. *PLoS ONE* **4**, e5299.

Woyke, T., Sczyrba, A., Lee, J., Rinke, C., Tighe, D., Clingenpeel, S., Malmstrom, R., Stepanauskas, R. & Cheng, J.-F. (2011). Decontamination of MDA reagents for single cell whole genome amplification. *PLoS ONE* **6**, e26161.

Wrighton, K. C., Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., VerBerkmoes, N. C., Wilkins, M. J., Hettich, R. L., Lipton, M. S. & other authors (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**, 1661–1665.

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester, M. & other authors (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608–1615.

Zachara, J. M., Long, P. E., Bargar, J., Davis, J. A., Fox, P., Fredrickson, J. K., Freshley, M. D., Konopka, A. E., Liu, C. & other authors (2013). Persistence of uranium groundwater plumes: contrasting mechanisms at two DOE sites in the groundwater–river interaction zone. *J Contam Hydrol* **147**, 45–72.

Zhang, K., Martiny, A. C., Reppas, N. B., Barry, K. W., Malek, J., Chisholm, S. W. & Church, G. M. (2006). Sequencing genomes from single cells by polymerase cloning. *Nat Biotechnol* **24**, 680–686.

Zhang, D.-C., Schinner, F. & Margesin, R. (2010). *Pedobacter bauzanensis* sp. nov., isolated from soil. *Int J Syst Evol Microbiol* **60**, 2592–2595.

Zhou, Y., Kellermann, C. & Griebler, C. (2012a). Spatio-temporal patterns of microbial communities in a hydrologically dynamic pristine aquifer. *FEMS Microbiol Ecol* **81**, 230–242.

Zhou, Z., Jiang, F., Wang, S., Peng, F., Dai, J., Li, W. & Fang, C. (2012b). *Pedobacter arcticus* sp. nov., a facultative psychrophile isolated from Arctic soil, and emended descriptions of the genus *Pedobacter*, *Pedobacter heparinus*, *Pedobacter daechungensis*, *Pedobacter terricola*, *Pedobacter glucosidilyticus* and *Pedobacter lentus*. *Int J Syst Evol Microbiol* **62**, 1963–1969.

Edited by: W. Röling