The ISME Journal (2014) 8, 1831–1842

www.nature.com/ismej

ORIGINAL ARTICLE

Single cell genomic study of *Dehalococcoidetes* species from deep-sea sediments of the Peruvian Margin

Anne-Kristin Kaster[1], Koshlan Mayer-Blackwell[2], Ben Pasarelli[3] and Alfred M Spormann[1,2]
[1]*Department of Chemical Engineering, Stanford University, Stanford, CA, USA;* [2]*Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA and* [3]*Department of Bioengineering, Stanford University, Stanford, CA, USA*

The phylum *Chloroflexi* is one of the most frequently detected phyla in the subseafloor of the Pacific Ocean margins. Dehalogenating *Chloroflexi* (*Dehalococcoidetes*) was originally discovered as the key microorganisms mediating reductive dehalogenation *via* their key enzymes reductive dehalogenases (Rdh) as sole mode of energy conservation in terrestrial environments. The frequent detection of *Dehalococcoidetes*-related 16S rRNA and *rdh* genes in the marine subsurface implies a role for dissimilatory dehalorespiration in this environment; however, the two genes have never been linked to each other. To provide fundamental insights into the metabolism, genomic population structure and evolution of marine subsurface *Dehalococcoidetes sp.*, we analyzed a non-contaminated deep-sea sediment core sample from the Peruvian Margin Ocean Drilling Program (ODP) site 1230, collected 7.3 m below the seafloor by a single cell genomic approach. We present for the first time single cell genomic data on three *d*eep-*s*ea *Chloroflexi* (Dsc) single cells from a marine subsurface environment. Two of the single cells were considered to be part of a local *Dehalococcoidetes* population and assembled together into a 1.38-Mb genome, which appears to be at least 85% complete. Despite a high degree of sequence-level similarity between the shared proteins in the Dsc and terrestrial *Dehalococcoidetes*, no evidence for catabolic reductive dehalogenation was found in Dsc. The genome content is however consistent with a strictly anaerobic organotrophic or lithotrophic lifestyle.
The ISME Journal (2014) **8,** 1831–1842; doi:10.1038/ismej.2014.24; published online 6 March 2014
Subject Category: Geomicrobiology and microbial contributions to geochemical cycles
Keywords: *Chloroflexi*; *Dehalococoocides*; deep-sea sediment; dehalogenation

Introduction

The deep marine subsurface is one of the largest unexplored biospheres on Earth, and is estimated to contain up to 3×10^{29} microbial cells, a number equivalent to the combined microbial biomass of the oceanic water column and terrestrial soil (Kallmeyer *et al.*, 2012). Microbial communities in the subseafloor depend on the supply of energy substrates and growth factors from the overlying surface world (that is, land and ocean) and/or the underlying lithosphere (that is, earth's crust and mantle). The significant abundance of microorganisms in the marine subsurface suggests that they have been involved in global biogeochemical cycles over geological timescales (Parkes *et al.*, 2002, 2005; D'Hondt *et al.*, 2002a, b, 2004). Most of these bacteria and archaea are phylogenetically distinct from previously cultured microorganisms (Biddle and Teske, 2008; Biddle *et al.*, 2008; Lever *et al.*, 2013; Rinke *et al.*, 2013), and consequently their metabolic characteristics as well as ecological roles remain largely unknown.

Molecular environmental surveys have shown that the phylum *Chloroflexi* is particularly widely distributed in many deep-sea sediments, representing up to 80% of the total bacterial 16S rRNA gene sequences at some sites (Parkes *et al.*, 2002; Webster *et al.*, 2004; Inagaki *et al.*, 2006; Webster *et al.*, 2006; Fry *et al.*, 2008; Blazejak and Schippers, 2010). Despite the ubiquity of subsurface *Chloroflexi*, these microorganisms have so far evaded cultivation attempts in the laboratory, and therefore essentially nothing is known about their metabolism or eco-physiology (D'Hondt *et al.*, 2004; Toffin *et al.*, 2004; Batzke *et al.*, 2007; Webster *et al.*, 2011). The few cultured representatives from the phylum *Chloroflexi* cover a wide metabolic spectrum,

Correspondence: AM Spormann, Department of Chemical Engineering, Stanford University, Mail Code 5429, 318 Campus Drive, Stanford, 94305 CA, USA.
E-mail: spormann@stanford.edu
Received 16 October 2013; revised 27 January 2014; accepted 3 February 2014; published online 6 March 2014

including aerobic organotrophs, anoxygenic phototrophs, nitrate reducers and anaerobic halorespirers (Yamada and Sekiguchi, 2009; Tas *et al.*, 2010, Bryant *et al.*, 2012; Krzmarzick *et al.*, 2012; Sorokin *et al.*, 2012).

Among the *Chloroflexi* 16S rRNA gene sequences found in deep-sea subsurface environments are those affiliated with the distinct class-level clade *Dehalococcoidetes*, consisting of *Dehalococcoides mccartyi* and *Dehalogenimonas sp.* (Parkes *et al.*, 2002; Inagaki *et al.*, 2003, 2006; Webster *et al.*, 2004; Biddle *et al.*, 2008, 2011; Blazejak and Schippers, 2010; Siddaramappa *et al.*, 2012; Loeffler *et al.*, 2013). *Dehalococcoidetes* were first isolated from chloroethene-contaminated terrestrial aquifer environments, and are strictly anaerobic, slow-growing and highly niche-adapted *Chloroflexi* that use organohalide respiration *via* reductive dehalogenases (Rdh) as their sole mode of energy conservation. Their small genomes range in size from 1.3 to 1.7 Mb, and these microorganisms depend on a supporting microbial community for acquisition of electron donors and cofactors (for example, H_2 and cobalamin derivatives). So far only nine *Dehalococcoidetes* strains have been cultivated (Maymo-Gatell *et al.*, 1997; Muller *et al.*, 2004; He *et al.*, 2005; Miller *et al.*, 2005; Cheng and He, 2009; Moe *et al.*, 2009; Siddaramappa *et al.*, 2012; Loeffler *et al.*, 2013; Poritz *et al.*, 2013).

While *Dehalococcoidetes* are mostly undetectable by molecular methods in non-contaminated freshwater environments, a pioneering study by Futagami *et al.* (2009) discovered *Dehalococcoidetes* and novel putative reductive dehalogenase genes (*rdh)* in deep-sea sediments from the southeast Pacific off Peru, the eastern equatorial Pacific, the Juan de Fuca Ridge flank off Oregon (USA) and the northwest Pacific off Japan. The abundant detection of both *Dehalococcoidetes* 16S rRNA gene sequences and *rdh* genes in conjunction with the previously identified terrestrial chloroethene-reducing *Dehalococcoidetes* suggested the presence of dissimilatory dehalorespiration in pristine marine subsurface sediments of the Pacific as an ecologically significant microbial activity (Futagami *et al.*, 2009). However *rdh* genes have so far never been linked to *Dehalococcoidetes* from subsurface environments.

We therefore initiated a study probing for a linkage between deep-sea sediment *Dehalococcoidetes* 16S rRNA and *rdh* gene sequences in a non-contaminated deep-sea sediment sample obtained from the Peruvian Margin site 1230 (Ocean Drilling Program (ODP) expedition 201) (Supplementary Figure 1) (D'Hondt *et al.*, 2003b). We employed molecular and single cell genomic approaches as the methods of choice to provide fundamental insights into the *in situ* metabolic lifestyle of *Dehalococcoidetes*-like *Chloroflexi* in these non-contaminated and poorly understood environments.

Materials and methods

DNA extraction and screening

DNA was extracted and screened from 31 deep-sea sediment samples originating from the eastern equatorial Pacific, the southeast Pacific off Peru (Peruvian Margin, ODP Leg 201), the northeast Pacific at the Juan de Fuca Ridge flank off Oregon, USA (IODP Expedition 301), the northwest Pacific off Japan (JAMSTEC Chikyu Shakedown Expedition CK06-06) and the Nankai Trough Forearc Basin off Japan (ODP Expedition 315) (Supplementary Table 1). DNA was extracted from 5 to 8 g of sample using various methods, ranging from the method described by Futagami *et al.* (2009) to phenol-chloroform extraction and/or using the FastDNA SPIN Kit for Soil (MP Biomedicals, Solon, OH, USA). The best method to extract DNA, while avoiding excessive co-extraction of humic acids, was found to be a modified FastDNA SPIN Kit for Soil extraction protocol (for details, see Supplementary Information). One microliter of extracted DNA was amplified by multiple displacement amplification (MDA) (Lasken, 2007) in three independent reactions using the Repli-g Mini Kit for single cells (Qiagen, Hilden, Germany). Amplified DNA was then recovered using the QiAamp kit (Qiagen). DNA concentration for PCR screening ranged from 50 to 100 ng μl^{-1}. 16S rRNA genes were amplified using broad eubacterial primers 27F (5′-AGAGTTTGATCMTGGCTCAG-3′) and 1391R (5′-GACGGGCRGTGWGTRCA-3′) (Lane, 1991), *rdh* genes using primers RFF2 (5′-SHMGBMGWGATTTY ATGAARR-3′) and B1R (5′-CHADHAGCCAYTCRTA CCA-3′) (Krajmalnik-Brown *et al.*, 2004; Futagami *et al.*, 2009) and *ssrA* genes using primers tmRNA_ Dehalococcoidetes_U_dhg_f_deg (5′-GGGANGCGTG NNTTCGAC-3′) and rec02R (5′-TGCGGWATGCC VATGTGG-3′) (McMurdie *et al.*, 2011). PCR was performed with 0.5 μM of each primer in 50 μl volumes. For details of PCRs, see Supplementary Table 2. PCR products were cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA) and subsequently Sanger sequenced.

Nanoliter-qPCR screening

A panel of novel qPCR primer pairs was developed (Mayer-Blackwell and Spormann, 2014) (Supplementary Table 3) to detect a large fraction of known reductive dehalogenase genes (*rdhA*) at a single annealing temperature and buffer composition. Non-redundant full-length and near full-length *rdhA* gene sequences curated in the protein family Pfam PF13486 as of July 2012 (Punta *et al.*, 2012) were clustered based on the percent identity (PID) using an all-vs-all blastp (Altschul *et al.*, 1990). Sequences considered ranged between 350 and 700 amino acids. Specific assays were designed for 50 reference sequences, each with at least one known high PID homolog. Thousand of

candidate primer pairs were generated using primer3 (Rozen and Skaletski, 2000) and filtered based on complementarity to at least three distinct sequences sharing high PID to the reference sequence. Where possible, lack of complementarity to homologs with lower PID was used as additional criteria in assay selection. Our overall collection of assays included two types of PCR primer sets. The first one was highly specific to an *rdh* reference sequence and those homologs with high PID (usually >90% at the amino-acid level) to the reference. The second type was of broader specificity to extend the number of sequences matched to include the reference sequence and as many homolog sequences as possible, regardless of the level of PID with the reference.

Over 600 candidate assays were physically tested in a nanoliter-qPCR platform (WaferGen Biosystems Inc., Fremont, CA, USA) against a collection of 500 bp linear DNA standards (Integrated DNA Technology, Coralville, IA, USA) that were diluted to approximate concentrations of 20 000, 2000, 200 and 20 copies per 100 nl reaction well. Of those 600 assays tested, 170 assays were selected on the basis of high PCR efficiency (>90%) and the sensitivity to detect target at <10 gene copies per reaction. In all, this validated panel of assays covered 233 of 389 distinct sequences in the *dehalogenase protein family*, including a third of the sequences previously recovered from the most comprehensive deep-sea sediment survey of reductive dehalogenases to date (Futagami *et al.*, 2009). The other sequences not targeted by the panel had no high PID homologs, making them less-suitable candidates for assay development.

DNA samples were applied to nanoliter-qPCR chip at concentration ranging from 1 to 50 ng μl^{-1} (Supplementary Table 4). In separate wells, positive control standards for a larger fraction of assays were supplied at a concentration of 500 copies per reaction well. A chip wide positive control consisting of three *Mus musculus* gene fragments was spiked into the master mix at a concentration of 400 copies per reaction to test whether inhibitory compound—typical in DNA extractions from sediment environments—affected the rate of DNA amplification. The spike-in amplified uniformly across the samples suggesting that inhibitory effects were minimal if at all present. The reactions were carried out according to the instrument manual specifications with Roche (Basel, Switzerland) Lightcycler 480 SYBR Master Mix. The PCR program was as follows: 95 °C for 3 min, then 40 cycles of: 95 °C for 1 min and 60 °C for 1:10 min.

Single cell sorting via Fluorescent Activated Cell Sorting and amplification of single cell DNA using MDA

A frozen deep-sea sediment sample of the Peruvian Margin drill site 1230 (ODP 201), collected 7.3 meters below seafloor (mbsf) and stored at −80 °C without glycerol preservation for 8 years, was used for single cell genome analysis. Physical isolation of the single cells was performed by Fluorescent Activated Cell Sorting in two 384-well plates (630 single cells, 6 positive controls and 132 negative controls). The sample processing was performed at the Bigelow Laboratory Single Cell Genomics Center (www.bigelow.org/scgc). Single cells were lysed, and the DNA was amplified by MDA as described previously (Swan *et al.*, 2011). In all, 250 wells showed good amplification with a Cp value of <10 h (~40%). DNA was screened with broad eubacterial (27F-M13: 5′-AGRGTTYGATYM TGGCTCAG-3′/907R_degen-M13: 5′-CCGTCAATTC MTTTRAGTTT-3′) and archaeal (Arc_344F-M13: 5′-ACGGGGYGCAGCAGGCGCGA-3′/Arch_915R-M13R: 5′-GTGCTCCCCCGCCAATTCCT-3′) 16S rRNA primers (Lane, 1991; Shapiro, 2000; Zhu *et al.*, 2005) and Sanger sequenced. Analysis with the RDB (Ribosomal Database) (Cole *et al.*, 2009) yielded 33 hits (5.2% of all single cells sorted, 13.2% of successful MDA reactions), including three *Chloroflexi* single cells. The single *Chloroflexi* cells showed a 16S rRNA sequence by Sanger most similar to *Dehalogenimonas* (Figure 1). The first MDA products yielded 500–900 ng

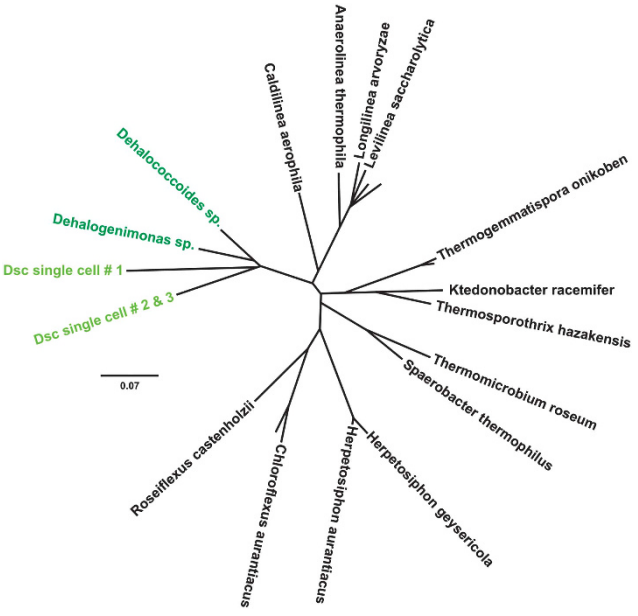

Figure 1 Phylogenetic tree of cultured and sequenced *Chloroflexi* including three single Dsc cells from a deep-sea sediment sample of the Peru Margin 1230 site collected 7.3 mbsf. The tree was generated using the Geneious tree builder software (Auckland, New Zealand) (gap cost matrix of 51% similarity (5.0/−3.0), gap open penalty of 12, gap extension penalty of 3, global alignment, Jukes Cantor genetic distance model and neighbor joining as tree building method with no outgroup). Two of the isolated Dsc single cells showed an identical 16S rRNA sequence. All three single cells belong to the class *Dehalococcoidetes*, including *Dehalococcoides* and *Dehalogenimonas sp.*The closest cultivated isolate was *Dehalogenimonas lykanthroporepellens* strain BL-DC-9, which had 83% sequence identity with the 16S rRNA gene of Dsc1 and 87% sequence identity with DscP2.

of DNA after clean up with the QIAamp DNA kit (Qiagen). The first MDA products of the three single cells were re-amplified in a second MDA. To avoid additional bias, the second MDA was performed in four separate reactions that were subsequently combined at the end.

Sequencing with Illumina HighSeq and PacBio RS
The first MDA products of the single cells were sequenced separately on an Illumina HighSeq platform (San Diego, CA, USA) using Nextera library preparation with an average yield of 15 000 Mb and 150 000 000 reads with 2×100 bp read length. The second MDA products were sequenced using the PacBio *RS* Magbead CLR sequencing technique (Menlo Park, CA, USA), resulting in a mean read length of over 2.5 Kb and ~100 Mb raw sequence data. Sequencing was carried out according to the manufacturer's instructions and resulted in 12 Mb raw sequence data for single cell 1 and 190 Mb for single cells 2 and 3.

Assembly and bioinformatics
After quality assessment, trimming and/or normalization of the sequencing reads, bioinformatics tools were used to conduct the assembly, open reading frame calling and annotation of the genes (see Supplementary Information). Analysis with PRINSEQ (http://prinseq.sourceforge.net/) revealed a high exact duplication rate (>90%) in all three raw Illumina data sets. Analysis of the reads showed differences in coverage of up to 4 orders of magnitude, indicating large bias in the first MDA reaction (Supplementary Figure 2). Different strategies were applied to assemble the reads of the individual cells and improve the assembly with gap-closing strategies (Supplementary Table 4) (Bankevich *et al.*, 2012; Koren *et al.*, 2012). Single cells 2 and 3 were assembled together since they showed almost 100% identity at the nucleotide level after individual assembly (Figure 2). At this stage, a 0.32-Mb assembly was contained in 126 contigs for single cell 1 and a 1.38-Mb assembly in 327 contigs for the co-assembly of single cells 2 and 3.

Genome annotation, estimation of completeness and size
Assembled contigs were submitted to the Integrated Microbial Genomes database annotation pipeline (IMG, version 4.1) (Markowitz *et al.*, 2010) and to the Rapid Annotations using Subsystems Technology pipeline (RAST, version 4.0) (Aziz *et al.*, 2008) in 2013. Some computationally assigned annotations were manually changed based on the inspection of evidence for the assigned annotations, orthologs in related genomes and gene neighborhoods. Pathways were predicted using RAST, IMG and KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa and Goto, 2000; Aziz *et al.*, 2008; Markowitz *et al.*, 2010). Nucleotide and amino-acid sequences of genes were blasted (Altschul *et al.*, 1990; Altschul and Koonin, 1998) as query sequences against the NCBI databases (using an *e*-value threshold of 10^{-5}). Gene annotations described in Results section together with the fasta nucleotide and amino-acid sequences of Dcs1 and DscP2 (for *d*eep-*s*ea *Chloroflexi* (Dsc) *p*opulation) including annotations are listed in Supplementary Documents 1 (Dcs1.faa), 2 (Dcs1.fna), 3 (DscP2.faa) and 4 (DscP2.fna), respectively.

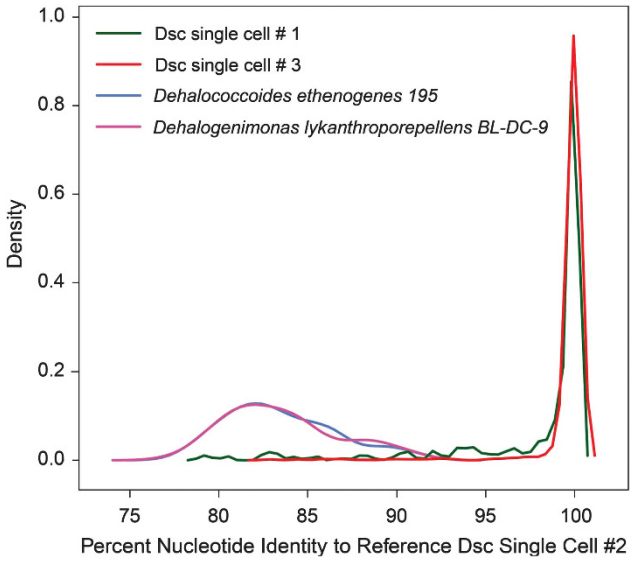

Figure 2 Nucleotide identity comparison of the individual assemblies of single cells Dsc1, 2 and 3. An all-vs-all blastn was conducted on predicted coding nucleotide sequences (RAST, Aziz *et al* 2008) from each isolate, using Dsc single cell 2 as the reference genome. The default blastall parameters were used with an *e*-value cutoff of 0.001. A custom python script was used to sort the best single alignment (if any) for each predicted coding gene in the reference genome and comparison genomes. Alignments shorter than 50 bp in length were excluded. A density plot of the percent identity for each alignment is visualized using the R statistical environment.

Sequence accession
The genomic data (contigs longer than 200 bp) are present as BioProjects with the accession numbers PRJNA222231 for Dsc1 and PRJNA227210 for DscP2 and has been deposited at GenBank under the accession JARM00000000 for Dsc1 and JARN00 000000 for DscP2. The version described in this paper is version JARN01000000 and JARM01 000000, respectively.

Results and discussion

Single cell source
We analyzed total DNA extracted from several deep-sea sediments from the Peruvian Margin and other Pacific sites for the presence of *Dehalococcoidetes* 16S rRNA and *rdh* genes by PCR and nanoliter-qPCR

(Supplementary Tables 1 and 4). On the basis of these screening results, we identified a sample obtained from the Peruvian Margin trench site 1230, collected at 7.3 mbsf as promising for further investigation by single cell genomics. Consequently, cells from this site were extracted, sorted by Fluorescent Activated Cell Sorting in 384-well plates, lysed and amplified *via* MDA. A flow diagram of the procedure is shown in Supplementary Figure 3. Out of 630 single cells sorted we identified three *Chloroflexi* cells, designated as Dsc (*D*eep-*s*ea *Chloroflexi*) 1, 2 and 3.

Site 1230 is located on the lower slope of the Peru Trench at 5086 m water depth. Sediments of this area are part of an accretionary wedge just landward of the Peru Trench (Suess and von Huene, 1988) (Supplementary Figure 1). The surface waters at the Peruvian Margin are part of Peru's upwelling system and are biologically highly productive, with a total organic content of ~2–3 wt% (D'Hondt *et al.*, 2003a,b). The upper 20 m of the sediment column contains a narrow suboxic zone with a steep sulfate gradient, and sulfate reduction is the dominant mineralization process. The gradient is nearly linear indicating that most of the net sulfate reduction takes place at the sulfate/methane interface (Iversen and Jorgensen, 1985; Borowski *et al.*, 1996, 2000; Niewohner *et al.*, 1998), and methane builds up steeply at the sulfate boundary. There was also visual evidence of methane hydrates (Kvenvolden *et al.*, 1990; Kvenvolden and Kastner, 1990; Suess *et al.*, 1990) (Supplementary Figure 4).

Our sample from 7.3 mbsf is located in the transition zone between the sulfate-rich surface layers and the sulfate-depleted, methanogenic deeper sediment layers. Microbiological sampling of site 1230 showed that methanogenic activity and steep sulfate depletion appeared to be driven by methane ascending from deeper deposits. This suggests microbial communities with activities in hydrate-bearing sediments rich in organic material. Acetate and formate are generated as fermentation products and are used as substrates by sulfate-reducing or methanogenic prokaryotes. Methanogenesis, microbial sulfate reduction and anaerobic methane oxidation appear to have major roles in this particular deep-sea sediment (D'Hondt *et al.*, 2003a,b).

Previous quantification of 16S rRNA gene copies at 1 m depth from the Peruvian Margin showed *Chloroflexi* numbers to be nearly equivalent to the total bacterial counts (Blazejak and Schippers, 2010). A metagenomic study of site 1229 (Supplementary Figure 1) also revealed high levels of *Chloroflexi* genes at 1, 16, 32 and 50 mbsf (Biddle *et al.*, 2008, 2011). It has also been speculated that subsurface *Chloroflexi* are active since their 16S rRNA sequences could be recovered from the surface of an *in situ* colonization experiment of polished and sterilized rock chips placed within a borehole, indicating recent growth (Orcutt *et al.*, 2011). Data from a 16S rRNA gene clone library that we prepared from the sample collected at site 1230 7.3 mbsf indicated that *Chloroflexi* may constitute >70% of the detected bacterial phyla (data not shown). On the basis of these cumulative findings and the data collected by the Shipboard Scientific Parties in 1988 and 2003, site 1230 provides an excellent opportunity for assessing the nature of *Chloroflexi* and their metabolic potential in hydrate-bearing sediments rich in organic material and under high hydrostatic pressure.

Genome statistics and analysis

Phylogenetic 16S rRNA gene sequence analysis grouped Dsc1, 2 and 3 within the previously termed 'subphylum II' clade of the *Chloroflexi* (Inagaki *et al.*, 2006). All nine known *Dehalococcoidetes* species (*Dehalococcoides mccartyi* strain BAV1, CBDB1, VS, GT, DCMB5, GY50, BTF08, *Dehalococcoides ethenogenes* 195 and *Dehalogenimonas lykanthroporepellens* BL-DC-9) share about 83% sequence identity at the 16S rRNA gene level with Dsc1 and 87% sequence identity with Dsc2 and 3.

The 16S rRNA gene of Dsc2 and 3 had an 81.7% pairwise nucleotide identity with Dsc 1 (ClustalW alignment with IUB cost matrix, gap open cost 15 and gap extent cost 6.66) (Figure 1). Comparing the Sanger sequenced 16S rRNA gene of Dsc2 and 3 revealed that they were identical over a range of 1364 nucleotides, and all other shared genes showed almost 100% pairwise identity on the nucleotide level after individual assembly (Figure 2). Therefore, we considered Dsc2 and 3 to be part of a coherent local Dsc population and combined the individual genomic sequence reads into one assembly, designated as DscP2 (for *D*eep-*s*ea *Chloroflexi p*opulation derived from Dsc*2* and 3).

For Dcs1, a 0.32-Mb and for DscP2, a combined 1.38-Mb genome were assembled, respectively. On the basis of size of contigs (bp), number of tRNAs and reference marker genes (from http://phylosift.wordpress.com/tutorials/scripts-markers/ and *Chloroflexi* marker genes, obtained from IMG), we conservatively estimated that the genomes of Dcs1 and DscP2 are 35% and 85% complete, respectively. Due to the bias of the amplification reaction (Supplementary Figure 2) (Lasken 2007, 2012; Lasken and Stockwell, 2007; Marcy *et al.*, 2007) it is unlikely that the genomes can be closed with the current data. The genome of Dcs1 was assembled into 126 contigs, ranging in size from 109 to 42 543 bp. In all, 42% of these contigs were longer than 1000 bp and 11% were longer than 10 000 bp. The N50 weighted median statistic of genome completeness was calculated as 12 440 bp. DscP2 was assembled into 327 contigs, ranging in size from 94 to 51 893 bp. In all, 52% of these contigs were longer than 500 bp, 37% were longer than 1000 bp and 12% were longer than 10 000 bp. The N50 weighted median statistic of genome completeness

was calculated as 19 540 bp (Supplementary Table 5) (http://prinseq.sourceforge.net/; Earl *et al.*, 2011). Despite being unfinished, the assembled genome of DscP2 still appears to provide a reasonable representation of a DscP species. Genome statistics are shown in Table 1.

The most common contaminants in single cell genome studies are *Delftia*, *Pseudomonas* and *Ralstonia*, which stem from the reagents used in the MDA reaction; others can be *Propionibacterium* and *Lactobacillus* (Woyke *et al.*, 2011). No genes from these known common contaminants were found in our genomic data. Another source of contamination may be in the form of free DNA associated with the sample that is sorted and carried through during the Fluorescent Activated Cell Sorting procedure with the single cell. We therefore performed an additional manual screening of the data. While there are no clear rules on the identification and removal of contamination (that is, phage or horizontal gene transfer may be difficult to discriminate from contamination), a principle component analysis of the tetramer frequency and GC content of the scaffolds showed no suspicious open reading frames (Markowitz *et al.*, 2010). In addition, the absence of eukaryotic or viral genes in the scaffolds strongly indicated that the genetic material was not significantly, if at all, contaminated.

Blastp and blastn analyses (Altschul *et al.*, 1990) revealed a coherent phylogenetic composition of the DscP2 genome. Over 75% of the DscP2 genes had the best Blast hit to genes of known *Chloroflexi* with a sequence identity of larger than 90%. Forty-eight percent of all encoded protein sequences had a 60% or better pairwise identity to proteins found in *Dehalogenimonas*. Twenty-three percent of all protein-encoding genes showed the best hit to phylogenetically unassigned bacteria. For Dcs1, 51% of the genes had the best closest hit to genes in the phyla *Chloroflexi* with a sequence identity over 90%; 47% of the hits were to phylogenetically unassigned bacteria (Supplementary Tables 6 and 7). More than 100 genes in Dcs1 and 500 genes in DscP2 are encoding hypothetical proteins. Tetramer frequency analysis using the IMG software suggested that horizontal gene transfer of some genes might have occurred into Dcs1 and DspP2.

Predicted metabolic properties

Despite the grouping of the 16S rRNA gene sequence of Dcs1 and DscP2 into the class of *Dehalococcoidetes*, we only found one gene putatively encoding a conserved domain of a reductive dehalogenase in DscP2 (DscP2_00865) (Supplementary Figure 5). However, no associated membrane-bound anchor proteins or transcriptional regulators, which are commonly associated with respiration of organohalide compounds, were found in the genomic data (McMurdie *et al.*, 2009). The amino-acid sequence of this iron–sulfur cluster protein shows only 27% identity to a putative reductive dehalogenase found in *Dehalococcoides ethenogenes* 195 and *Dehalococcoides mccartyi BAV1*, and has 40–50% identity to putative reductive dehalogenases found in *Firmicutes*, such as *Desulfitobacterium* and *Dehalobacter* and γ-Proteobacteria, for example, *Shewanella* (*e*-values range from 10^{-7} to 10^{-4}). Extended searches for reductive dehalogenases genes using PCR and nanoliter-qPCR with primers targeting these genes on the MDA-derived DNA as template did also not provide evidence for the presence of *rdh* (Supplementary Table 4). This apparent absence of *rdh* genes in a genome derived from a deep-sea sediment *Dehalococcoidetes* is intriguing, considering that some marine subsurface *Dehalococcoidetes*-affiliated bacteria were hypothesized to perform reductive dehalogenation (Adrian *et al.*, 2009; Futagami *et al.*, 2009; Valentine, 2010; Durbin and

Table 1 Genome statistics of DscP1 and DscP2

	Dsc1		*DscP2*	
Estimation of completeness	*35%* *Number*	*% of Total*	*85%* *Number*	*% of Total*
DNA, total number of bases	324 175	100	1 380 382	100
DNA coding number of bases	266 905	82	1 178 768	85
DNA G + C number of bases	147 712	46	626 849	45
DNA scaffolds	126	100	327	100
Genes total number	493	100	1857	100
Protein coding genes	461	94	1807	97
RNA genes	32	6.5	50	2.7
tRNA genes	25	5	47	2.5
Protein coding genes with function prediction	255	52	1240	67
Protein coding genes without function prediction	206	42	567	31

Abbreviations: Dsc, deep-sea *Chloroflexi*; DscP, *deep-sea Chloroflexi p*opulation; MDA, multiple displacement amplification.
Due to the bias of the MDA reaction it is unlikely that the genomes can be closed with the current data. It is estimated based on the size of contigs (bp), number of tRNAs and reference marker genes that the completeness of the genomes is ~35% for Dsc1 and 85% for DscP2.

Teske, 2011; Wagner *et al.*, 2012). In terrestrial *Dehalococcoidetes* strains, the *rdh* operon is embedded in horizontally acquired genomic islands that integrated at the single-copy tmRNA gene, *ssrA*. The genomes of most sequenced *Dehalococcoidetes* contain two high-plasticity regions around the origin of replication that harbor the vast majority of putative *rdh* (Kube *et al.*, 2005; McMurdie *et al.*, 2009). While the absence of *rdh* genes, as found in this study, is not an unambiguous proof for a non-halorespiring metabolism, the presence of *rdh* genes in a *Dehalococcoidetes*-like structure in DscP2 is unlikely, based on the estimated 85% or greater completeness of the genome and the >30% genome occupation of high-plasticity region in *Dehalococcoidetes. rdhAB* and genes believed to be involved in assembly and maturation (*rdhF-I*) or regulation (*rdhD*, *R*) comprise between 3.5% and 8.6% of *Dehalococcoidetes* genomes by length (McMurdie *et al.*, 2009). Thus, although the two *Dsc* genomes show most similarity to genomes of previously sequenced *Dehalococcoides* and *Dehalogenimonas sp.*, Dsc1 and DscP2 appear considerably different in terms of overall gene content and do not harbor canonical *rdh* genes similar to those found in terrestrial *Dehalococcoidetes* strains.

Our finding is interesting in the light of a recent report on the spatial distribution of dehalogenation activity in the Nankai Trough plate-subduction zone of the northwest Pacific off the Kii Peninsula (Japan). Incubation experiments with slurries of sediment collected at various depths and locations showed that degradation of several organohalides occurred in the shallow sedimentary basin and *rdhA* genes were detected in the sediments. Interestingly, DNA fragments obtained from those positive enrichment cultures showed best BLAST hits to known *Dehalococcoides*. However, no functionally known dehalogenation-related gene such as *rdhA* was found. This indicates the need to improve the molecular approach to assess functional genes for reductive dehalogenation and to probe other classes of enzymes responsible for organohalide respiration (Futagami *et al.*, 2013).

Table 2 Key metabolic genes found in DscP2

Energy conserving pathway	*EC number*	*Gene number DscP2_0*
Carbon fixation pathways		
2-Oxoacid:acceptor oxidoreductase, gamma subunit	1.2.7.3	1608
3-Isopropylmalate dehydratase small and large subunit	4.2.1.33	1842, 1843 and 0331, 0335
5,10-Methylenetetrahydrofolate reductase (NAD(P))	1.5.1.20	1779
Acetate-CoA ligase	6.2.1.1	1730
Acetyl-CoA acetyltransferase	2.3.1.9	0601, 1197
CO dehydrogenase/CO-methylating acetyl-CoA synthase complex, beta and delta subunit	2.3.1.169	1776, 1403
Dihydrolipoamide dehydrogenase	1.8.1.4	1365, 1528
Enoyl-CoA hydratase/carnithine racemase	4.2.1.55	920
Formate dehydrogenase, alpha and delta subunit	1.2.1.2	0586, 1042
Formyltetrahydrofolate synthetase	6.3.4.3	1405, 1568
Fumarase, class I alpha subunit	4.2.1.2	1663
Glucose-6-phosphate isomerase/mannose-6-phosphate isomerase, archaeal	5.3.1.9/ 5.3.1.8	1799
Hydrolyase, beta region	4.2.1.2	1661
Isocitrate dehydrogenase	1.1.1.42/1.1.1.41	1844, 0336
Isopropylmalate/homocitrate/citramalate synthases		1845, 0330, 0337
Phosphoenolpyruvate synthase	2.7.9.2.	1066
Pyruvate:ferredoxin oxidoreductase	1.2.7.1	1432, 0683
Pyruvate/oxaloacetate carboxyltransferase	6.4.1.1	1665
Ribose 5-phosphate isomerase B	5.3.1.6	1056
Ribulose-5-phosphate 3-epimerase	5.1.3.1	1556
Succinyl-CoA synthetase, alpha and beta subunit	6.2.1.5	0599, 0660 and 1195, 1196, 1198
Transketolase	2.2.1.1	1242, 1243
Triosephosphate isomerase	5.3.1.1	882
Methane metabolism		
Heterodisulfide reductase subunits		1404, 1424, 1443
Coenzyme F_{420}-reducing hydrogenase, alpha, delta and gamma subunit	1.12.99.-	1166,1167, 1173
Archaeal/vacuolar-type H+-ATPase subunit ABCDGEFI	3.6.3.14	1180, 1181, 1182, 1183, 1184, 1185, 0692, 0694
5,10-Methylenetetrahydrofolate reductase (NAD(P))	1.5.1.20	1779
Nitrogen metabolism		
Glutamate dehydrogenase (NAD/NADP)	1.4.1.3	1781
Glutamate synthase (NADPH)	1.4.1.14	1852, 1168, 1291
Nitrite reductase (NAD(P)H) large subunit	1.7.1.4	1429, 1439
Nitrogen-fixing NifU domain protein		0590, 0617, 1049
Nitroreductase		0133, 0980
Sulfur metabolism		
Cysteine desulfurase NifS	2.8.1.7	598, 1050
Putative dissimilatory sulfite reductase		1814
Serine O-acetyltransferase	2.3.1.30	1652
Molybdenum cofactor biosynthesis protein MoaA, B, C		0184, 0775, 0776

Abbreviations: DscP, deep-sea *Chloroflexi p*opulation; IMG, Integrated Microbial Genomes; RAST, Rapid Annotations using Subsystems Technology.
We attempted to predict other central metabolic pathways from the genome annotations by IMG (Markowitz *et al.*, 2010) and RAST (Aziz *et al.*, 2008), and found that no complete gene set encoding enzymes for a known canonical energy conserving pathways were present.

Interestingly, the assembled genome of DscP2 revealed the presence of a predicted haloacid dehalogenase (HAD) gene (DscP2_01004). This finding is similar to that recently made in the genome of a *Dehalococcoidetes* isolated from a metagenome derived from floodplain sediments deposited by the Colorado River (Hug *et al.*, 2013). This class of enzymes catalyzes the hydrolytic dehalogenation of halogenated organic acids. The HAD and Rdh families are none-homologous, mechanistically different, and have evolved independently. The HAD may confer the ability to utilize iodated, chlorinated and/or brominated compounds, which occur naturally in sediments, by converting halogenated organic compounds into halogen-free organic compounds, which can then be used in canonical carbon degradation pathways. Intriguingly, also genes encoding for putative HAD superfamily enzymes are found (DscP2_01698, 0889), one being located next to the putative *rdh* (DscP2_0864). HAD superfamily enzymes are also found in known *Dehalococcoidetes* strains (Moe *et al.*, 2009; Loeffler *et al.*, 2013).

With the apparent absence of putative reductive dehalogenases in the genomes, we attempted to predict central catabolic pathways for DscP2 from the genome annotations. Therefore, we analyzed pathways using RAST, IMG and KEGG (Kanehisa and Goto, 2000; Aziz *et al.*, 2008; Markowitz *et al.*, 2010). IMG annotation seemed in general more precise when compared with the NCBI database than the ones from RAST; however, all potentially interesting proteins were blasted against the NCBI database as well, to confirm annotations and to avoid over-annotation. In addition, the absence or presence of certain gene encoding proteins was confirmed by performing BLAST analysis against the genome of DscP2. However, we could not identify a complete gene set encoding enzymes for a known canonical energy-conserving pathway (Table 2).

A number of genes encoding putative enzymes found to be involved in methanogenesis were found, such as heterodisulfide reductase (Hdr) subunits. HdrA is very well conserved in many anaerobic microorganisms from terrestrial and aquatic environments (Kaster *et al.*, 2011b). The finding of heterodisulfide reductase-like enzymes in strict anaerobes, such as sulfate reducers, acetogens and methanogens (Stojanowic *et al.*, 2003; Strittmatter *et al.*, 2009; Kaster *et al.*, 2011a; Callaghan *et al.*, 2012), supports the notion that Dsc *sp.* may have a primarily strict anaerobic metabolism. In addition, an archaeal type H^+ ATPase (subunits ABCDEFI) was present.

Multiple genes and operons encoding for hydrogenase accessory proteins were identified (*hypABCDEF*) as well as genes encoding for the 'periplasmic' [NiFe]-hydrogenases, F_{420}-reducing hydrogenase subunits alpha, delta and gamma. A putative F_{390} synthetase was identified, which in some methanogens synthesize 8-hydroxyadenylylated coenzyme F_{420} (coenzyme F_{390}) (Vermeij *et al.*, 1996). The function of coenzyme F_{390} is yet unclear, although studies suggest that it is involved as a response regulator for sensing changing cellular hydrogen concentrations (Vermeij *et al.,* 1997). However, no gene encoding a membrane-bound energy-conserving hydrogenase was found. The DscP2 genome apparently also lacks a standard electron transport chain including cytochromes, quinones or methanophenazine, but contains ferredoxins, flavodoxins and iron–sulfur flavoprotein, as well as iron transporters and many iron-sulfur assembly proteins.

Some genes encoding for enzymes of the Wood-Ljungdahl pathway were discovered: one encoding a formate dehydrogenase and a formyltetrahydrofolate synthetase; a carbonmonoxide dehydrogenase/CO-methylating acetyl-CoA synthase complex (alpha and beta subunit) and a methyltetrahydrofolate reductase were found (clustered with a methyltetrahydrofolate methyltransferase). An acetyl-CoA synthetase was present as well. However, genes encoding for a methenyltetrahydrofolate cyclohydrolase and methylenetetrahydrofolate dehydrogenase were apparently absent. Genes encoding subunits of the carboxylating pyruvate:ferredoxin oxidoreductase were found and may provide a link between the reductive acetyl-CoA pathway (Wood-Ljungdahl pathway) and other anabolic pathways. The genes found above are present only in strictly anaerobic prokaryotes, and are also present in terrestrial *Dehalococcoidetes*. Since some genes found in the Wood-Ljungdahl pathway can also be used to synthesize acetate from H_2 and CO_2, it does not allow us to distinguish whether DscP2 could be an acetate oxidizer, a homoacetogen, or assimilate CO_2 *via* the Wood-Ljungdahl pathway (Klenk *et al.*, 1997; Anderson *et al.*, 2011; Berg, 2011).

There is also some evidence for anabolic nitrogen metabolism. Genes encoding for a nitrite and nitro reductase were found, as well as several nitrogen-fixing NifU protein domains. In addition, a gamma-glutamyltranspeptidase and a putative glutathione transporter for potential utilization of glutathione as a sulfur source were found as well as an enzyme containing a domain of a putative dissimilatory sulfite reductase.

Recently, a partial genome of a *Dehalococcoidetes* species found in the Aarhus Bay in Denmark was reported (DEH-J10) (Wasmund *et al.*, 2013), sharing 92% 16S rRNA identity with Dcs1 and 85% with DscP2. However, analysis and comparison of the overall amino-acid sequence content showed that the DEH-J10 genome is significantly different from Dsc1 and DscP2 (Supplementary Figure 6). Also, in DEH-J10 no evidence for a dehalorespiring lifestyle was discovered. The data indicated that this genome likely confers metabolic versatility to the organism and the authors suggested that DEH-J10 could employ the beta-oxidation pathway to use various organics as a source for carbon and reducing equivalents.

Concluding remarks

Single cell genomics is a highly promising approach to obtain some physiological and genomic insights into uncultured microorganisms from rare samples in the biosphere. Our data on the deep-sea sediment *Dehalococcoidetes* (Dsc1 and DscP2) suggest that these microorganisms have an anaerobic core metabolism that is consistent with a strictly anaerobic lifestyle and which does not involve canonical electron transfer pathways. Although single cell genomics rarely retrieves complete genomes, it is nevertheless currently the only way of linking information about metabolic potentials of an organism that cannot be cultured. Our identification of three partial Dsc single cell genomes is especially remarkable, in the sense that *Chloroflexi* single amplified genomes from different environments are extremely difficult to obtain, despite their numerical abundance (Tanja Woyke, JGI, personal communication). This fact may be due to inefficient lysis of this type of bacteria and/or their usually high GC content, both factors making a successful amplification *via* MDA less likely.

Assuming that our DscP2 genome is reasonably complete, although not closed, it is tempting to speculate on the core lifestyle of non-dehalogenating *Dehalococcoidetes* and their adaptation to dehalorespiration. Terrestrial *Dehalococcoidetes* isolated from chloroethene-contaminated sites exhibit clear signatures of strong niche adaptation to reductive dehalogenation as sole catabolic metabolism. This is supported by the observation that, most, if not all, *rdh* genes are horizontally transferred and located within two high-plasticity regions, which are essentially specialized integration sites for acquisition of genomic islands (McMurdie *et al*., 2007, 2009, 2011). The conventional hypothesis on the lifestyle has been that under non-contaminated conditions, *Dehalococcoidetes* use naturally occurring organohalogens as the electron acceptor, and that exposure to anthropogenic chlorinated solvents selects for genetic variants that adapted involving horizontal gene transfer to utilization of anthropogenic organohalogens (McMurdie *et al*., 2007, 2011). This hypothesis implies that *Dehalococcoidetes* from non-contaminated environments such as from pristine deep-sea sediments would reveal a non-adapted dehalogenating lifestyle. In this study, we found no evidence for the presence of *rdh* genes or organohalogen respiration in Dsc1 or DscP2, which confirms previous observations (Wasmund *et al*., 2013). Thus, these findings collectively suggest that reductive dehalogenation of *Dehalococcoidetes*-like *Chloroflexi* is a recently acquired trait that became beneficial and selectable because reductively dehalogenating enzymes integrated mechanistically easily into some distinctive, strictly anaerobic non-dehalogenating core metabolism similar to that inferred for DscP2. The Rdh domain protein found in DscP2 might also be a precursor that could be an ancestor of a fully functional Rdh, once halogenated compounds are introduced.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This project was funded by the NSF Center for Dark Energy Biosphere Investigations (C-DEBI) and a fellowship to AK by C-DEBI and the Deutsche Forschungsgemeinschaft. We thank Jennifer Biddle (UC Delaware), Taiki Futagami and Fumio Inagaki (Kochi Institute for Core sample Research, Japan) for deep-sea sediment samples. We thank Ramunas Stepanauskas from the Bigelow Laboratory Single Cell Genomics Center for single cell sorting. We also thank Sergey Koren from the National Biodefense Analysis and Countermeasures Center for bioinformatics support.

References

Adrian L, Reinhardt R, Kube M. (2009). Genomics as a tool to analyze bioremediation potentials and functional diversification in subsurface environments. *Geochim Cosmochim Acta* **73**: A12.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Altschul SF, Koonin EV. (1998). Iterated profile searches with PSI-BLAST - a tool for discovery in protein databases. *Trends Biochem Sci* **23**: 444–447.

Anderson I, Risso C, Holmes D, Lucas S, Copeland A, Lapidus A *et al*. (2011). Complete genome sequence of Ferroglobus placidus AEDII12DO. *Stand Genomic Sci* **5**: 50–60.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA *et al*. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS *et al*. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455–477.

Batzke A, Engelen B, Sass H, Cypionka H. (2007). Phylogenetic and physiological diversity of cultured deep-biosphere bacteria from equatorial Pacific Ocean and Peru Margin sediments. *Geomicrobiol J* **24**: 261–273.

Berg IA. (2011). Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. *Appl Environ Microbiol* **77**: 1925–1936.

Biddle JF, Fitz-Gibbon S, Schuster SC, Brenchley JE, House CH. (2008). Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proc Natl Acad Sci USA* **105**: 10583–10588.

Biddle JF, Teske AP. (2008). A genetic view of diversity beneath the seafloor. *Geochim Cosmochim Acta* **72**: A83.

Biddle JF, White JR, Teske AP, House CH. (2011). Metagenomics of the subsurface Brazos-Trinity Basin

(IODP site 1320): comparison with other sediment and pyrosequenced metagenomes. *ISME J* **5**: 1038–1047.

Blazejak A, Schippers A. (2010). High abundance of JS-1- and Chloroflexi-related Bacteria in deeply buried marine sediments revealed by quantitative, real-time PCR. *FEMS Microbiol Ecol* **72**: 198–207.

Borowski WS, Paull CK, Ussler W. (1996). Marine pore-water sulfate profiles indicate in situ methane flux from underlying gas hydrate. *Geology* **24**: 655–658.

Borowski WS, Hoehler TM, Alperin MJ, Rodriguez NM, Paull CK. (2000). Significance of anaerobic methane oxidation in methane-rich sediments overlying the Blake Ridge gas hydrates. In: Paull CK, Matsumoto R, Wallace PJ, Dillon WP (eds) *Proceedings of the Ocean Drilling Program, Scientific Results* vol. 164. Ocean Drilling Program: College Station, TX, pp 87–99.

Bryant DA, Liu ZF, Li T, Zhao FQ, Costas AMG, Klatt CG *et al.* (2012). Comparative and functional genomics of anoxygenic green bacteria from the taxa Chlorobi, Chloroflexi, and Acidobacteria. *Adv Photosynth Resp* **33**: 47–102.

Callaghan AV, Morris BEL, Pereira IAC, McInerney MJ, Austin RN, Groves JT *et al.* (2012). The genome sequence of Desulfatibacillum alkenivorans AK-01: a blueprint for anaerobic alkane oxidation. *Environ Microbiol* **14**: 101–113.

Cheng D, He J. (2009). Isolation and characterization of '*Dehalococcoides*' sp. strain MB, which dechlorinates tetrachloroethene to trans-1,2-dichloroethene. *Appl Environ Microbiol* **75**: 5910–5918.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ *et al.* (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141–D145.

D'Hondt S, Jorgensen BB, Blake R, Dickens G, Hindrichs K, Holm N *et al.* (2002a). Microbial activity in deeply buried marine sediments. *Geochim Cosmochim Acta* **66**: A163.

D'Hondt S, Rutherford S, Spivack AJ. (2002b). Metabolic activity of subsurface life in deep-sea sediments. *Science* **295**: 2067–2070.

D'Hondt S, Jørgensen BB, Miller DJ. The Leg 201 Scientific Party. (2003a). ODP Leg 201 explores microbial life in deeply buried marine sediments off Peru. *JOIDES J* **29**: 11–15.

D'Hondt S, Jorgensen BB, Miller DJ, Batzke A, Blake R, Cragg BA *et al.* (2004). Distributions of microbial activities in deep subseafloor sediments. *Science* **306**: 2216–2221.

D'Hondt SL, Jørgensen BB, Miller DJ. (2003b). *Proceedings of the Ocean Drilling Program, Initial Reports,* vol. 201. Ocean Drilling Program: College Station, TX, USA.

Durbin AM, Teske A. (2011). Microbial diversity and stratification of South Pacific abyssal marine sediments. *Environ Microbiol* **13**: 3219–3234.

Earl D, Bradnam K St, John J, Darling A, Lin DW, Fass J *et al.* (2011). Assemblathon 1: a competitive assessment of de novo short read assembly methods. *Genome Res* **21**: 2224–2241.

Fry JC, Parkes RJ, Cragg BA, Weightman AJ, Webster G. (2008). Prokaryotic biodiversity and activity in the deep subseafloor biosphere. *FEMS Microbiol Ecol* **66**: 181–196.

Futagami T, Morono Y, Terada T, Kaksonen AH, Inagaki F. (2009). Dehalogenation activities and distribution of reductive dehalogenase homologous genes in marine subsurface sediments. *Appl Environ Microbiol* **75**: 6905–6909.

Futagami T, Morono Y, Terada T, Kaksonen AH, Inagaki F. (2013). Distribution of dehalogenation activity in subseafloor sediments of the Nankai Trough subduction zone. *Philos Trans R Soc Lond B Biol Sci* **368**: 20120249.

He J, Sung Y, Krajmalnik-Brown R, Ritalahti KM, Löffler FE. (2005). Isolation and characterization of *Dehalococcoides* sp. strain FL2, a trichloroethene (TCE)- and 1,2-dichloroethene-respiring anaerobe. *Environ Microbiol* **7**: 1442–1450.

Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR *et al.* (2013). Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* **1**: 22.

Inagaki F, Suzuki M, Nealson KH, Horikoshi K, D'Hondt SL, Jorgensen BB *et al.* (2003). Subseafloor microbial diversity in the Peru Margin (ODP Leg. 201). *Geochim Cosmochim Acta* **67**: A171.

Inagaki F, Nunoura T, Nakagawa S, Teske A, Lever M, Lauer A *et al.* (2006). Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments, on the Pacific Ocean Margin. *Proc Natl Acad Sci USA* **103**: 2815–2820.

Iversen N, Jorgensen BB. (1985). Anaerobic methane oxidation rates at the sulfate methane transition in marine-sediments from Kattegat and Skagerrak (Denmark). *Limnol Oceanogr* **30**: 944–955.

Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, D'Hondt S. (2012). Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc Natl Acad Sci USA* **109**: 16213–16216.

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

Kaster AK, Goenrich M, Seedorf H, Liesegang H, Wollherr A, Gottschalk G *et al.* (2011a). More than 200 genes required for methane formation from H(2) and CO(2) and energy conservation are present in Methanothermobacter marburgensis and Methanothermobacter thermautotrophicus. *Archaea* **2011**: 973848.

Kaster AK, Moll J, Parey K, Thauer RK. (2011b). Coupling of ferredoxin and heterodisulfide reduction *via* electron bifurcation in hydrogenotrophic methanogenic archaea. *Proc Natl Acad Sci USA* **108**: 2981–2986.

Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA *et al.* (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus. *Nature* **390**: 364–370.

Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G *et al.* (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat Biotechnol* **30**: 693–700.

Krajmalnik-Brown R, Holscher T, Thomson IN, Saunders FM, Ritalahti KM, Loffler FE. (2004). Genetic identification of a putative vinyl chloride reductase in Dehalococcoides sp. strain BAV1. *Appl Environ Microbiol* **70**: 6347–6351.

Krzmarzick MJ, Crary BB, Harding JJ, Oyerinde OO, Leri AC, Myneni SCB *et al.* (2012). Natural niche for organohalide-respiring Chloroflexi. *Appl Environ Microbiol* **78**: 393–401.

Kube M, Beck A, Zinder SH, Kuhl H, Reinhardt R, Adrian L. (2005). Genome sequence of the chlorinated

compound-respiring bacterium Dehalococcoides species strain CBDB1. *Nat Biotechnol* **23**: 1269–1273.

Kvenvolden KA, Frank TJ, Golan-Bac M. (1990). Hydrocarbon gases in tertiary and quaternary sediments offshore Peru—results and comparisons. In Suess E, von Huene R *et al.* (eds) *Proceedings of the Ocean Drilling Program, Scientific Results* vol. 112. Ocean Drilling Program: College Station, TX, pp 505–515.

Kvenvolden KA, Kastner M. (1990). Gas hydrates of the Peruvian outer continental margin. In Suess E, von Huene R *et al.* (eds) *Proceedings of the Ocean Drilling Program, Scientific Results* vol. 112. Ocean Drilling Program: College Station, TX, pp 517–526.

Lane D. (1991). 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds) *Nucleic Acid Techniques in Bacterial Systematics*. John Wiley: Chichester, UK.

Lasken RS. (2007). Single-cell genomic sequencing using Multiple Displacement Amplification. *Curr Opin Microbiol* **10**: 510–516.

Lasken RS, Stockwell TB. (2007). Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC Biotechnol* **7**: 19.

Lasken RS. (2012). Genomic sequencing of uncultured microorganisms from single cells. *Nat Rev Microbiol* **10**: 631–640.

Lever MA, Rouxel O, Alt JC, Shimizu N, Ono SH, Coggon RM *et al.* (2013). Evidence for microbial carbon and sulfur cycling in deeply buried ridge flank basalt. *Science* **339**: 1305–1308.

Loeffler FE, Yan J, Ritalahti KM, Adrian L, Edwards EA, Konstantinidis KT *et al.* (2013). Dehalococcoides mccartyi gen. nov., sp nov., obligately organohalide-respiring anaerobic bacteria relevant to halogen cycling and bioremediation, belong to a novel bacterial class, Dehalococcoidia classis nov., order Dehalococcoidales ord. nov and family Dehalococcoidaceae fam. nov., within the phylum Chloroflexi. *Int J Syst Evol Microbiol* **63**: 625–635.

Marcy Y, Ishoey T, Lasken RS, Stockwell TB, Walenz BP, Halpern AL *et al.* (2007). Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS Genet* **3**: 1702–1708.

Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y *et al.* (2010). The integrated microbial genomes system: an expanding comparative analysis resource. *Nucleic Acids Res* **38**: D382–D390.

Mayer-Blackwell K, Spormann AM. (2014). Quantitative and massively parallel assessment of reductive dehalogenase genes (submitted).

Maymo-Gatell X, Chien Y, Gossett JM, Zinder SH. (1997). Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. *Science* **276**: 1568–1571.

McMurdie PJ, Behrens SF, Holmes S, Spormann AM. (2007). Unusual codon bias in vinyl chloride reductase genes of Dehalococcoides species. *Appl Environ Microbiol* **73**: 2744–2747.

McMurdie PJ, Behrens SF, Müller JA, Göke J, Ritalahti KM, Wagner R *et al.* (2009). Localized plasticity in the streamlined genomes of vinyl chloride respiring *Dehalococcoides*. *PLoS Genet* **5**: e1000714.

McMurdie PJ, Hug LA, Edwards EA, Holmes S, Spormann AM. (2011). Site-specific mobilization of vinyl chloride respiration islands by a mechanism common in Dehalococcoides. *BMC Genomics* **12**: 287.

Miller GS, Milliken CE, Sowers KR, May HD. (2005). Reductive dechlorination of tetrachloroethene to trans-dichloroethene and cis-dichloroethene by PCB-dechlorinating bacterium DF-1. *Environ Sci Technol* **39**: 2631–2635.

Moe WM, Yan J, Nobre MF, da Costa MS, Rainey FA. (2009). *Dehalogenimonas lykanthroporepellens* gen. nov., sp. nov., a reductively dehalogenating bacterium isolated from chlorinated solvent-contaminated groundwater. *Int J Syst Evol Micr* **59**: 2692–2697.

Muller JA, Rosner BM, von Abendroth G, Meshulam-Simon G, McCarty PL, Spormann AM. (2004). Molecular identification of the catabolic vinyl chloride reductase from Dehalococcoides sp strain VS and its environmental distribution. *Appl Environ Microbiol* **70**: 4880–4888.

Niewohner C, Hensen C, Kasten S, Zabel M, Schulz HD. (1998). Deep sulfate reduction completely mediated by anaerobic methane oxidation in sediments of the upwelling area off Namibia. *Geochim Cosmochim Acta* **62**: 455–464.

Orcutt BN, Bach W, Becker K, Fisher AT, Hentscher M, Toner BM *et al.* (2011). Colonization of subsurface microbial observatories deployed in young ocean crust. *ISME J* **5**: 692–703.

Parkes RJ, Cragg BA, Wellsbury P. (2002). Recent studies on bacterial populations and processes in subseafloor sediments: a review (vol 8, pg 11, 2000). *Hydrogeol J* **10**: 346–346.

Parkes RJ, Webster G, Cragg BA, Weightman AJ, Newberry CJ, Ferdelman TG *et al.* (2005). Deep sub-seafloor prokaryotes stimulated at interfaces over geological time. *Nature* **436**: 390–394.

Poritz M, Goris T, Wubet T, Tarkka MT, Buscot F, Nijenhuis I *et al.* (2013). Genome sequences of two dehalogenation specialists - Dehalococcoides mccartyi strains BTF08 and DCMB5 enriched from the highly polluted Bitterfeld region. *FEMS Microbiol Lett* **343**: 101–104.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C *et al.* (2012). The Pfam protein families database. *Nucleic Acids Res* **40**: D290–D301.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF *et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–437.

Rozen S, Skaletski H. (2000). Primer3 on the WWW for general users and for biologist programmers. In Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press: Totowa, NJ, pp 365–386.

Shapiro HM. (2000). Microbial analysis at the single-cell level: tasks and techniques. *J Microbiol Methods* **42**: 3–16.

Siddaramappa S, Challacombe JF, Delano SF, Green LD, Daligault H, Bruce D *et al.* (2012). Complete genome sequence of Dehalogenimonas lykanthroporepellens type strain (BL-DC-9(T)) and comparison to 'Dehalococcoides' strains. *Stand Genomic Sci* **6**: 251–264.

Sorokin DY, Lucker S, Vejmelkova D, Kostrikina NA, Kleerebezem R, Rijpstra WIC *et al.* (2012). Nitrification expanded: discovery, physiology and genomics of a nitrite-oxidizing bacterium from the phylum Chloroflexi. *ISME J* **6**: 2245–2256.

Stojanowic A, Mander GJ, Duin EC, Hedderich R. (2003). Physiological role of the F-420-non-reducing

hydrogenase (Mvh) from Methanothermobacter marburgensis. *Arch Microbiol* **180**: 194–203.

Strittmatter AW, Liesegang H, Rabus R, Decker I, Amann J, Andres S *et al.* (2009). Genome sequence of Desulfobacterium autotrophicum HRM2, a marine sulfate reducer oxidizing organic carbon completely to carbon dioxide. *Environ Microbiol* **11**: 1038–1055.

Suess E, von Huene R. (1988). *Proceedings ofthe Ocean Drilling Program, Initial Reports,* vol. 112. Ocean Drilling Program: College Station, TX, USA.

Suess E, von Huene R, Emeis KC, Burgois J, del G Cruzado J, De Wever P *et al.* (1990). *Proceedings of the Ocean Drilling Program, Scientific Results,* vol. 112. Ocean Drilling Program: College Station, TX, USA.

Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D *et al.* (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**: 1296–1300.

Tas N, van Eekert MH, de Vos WM, Smidt H. (2010). The little bacteria that can - diversity, genomics and ecophysiology of 'Dehalococcoides' spp. in contaminated environments. *Microb Biotechnol* **3**: 389–402.

Toffin L, Webster G, Weightman AJ, Fry JC, Prieur D. (2004). Molecular monitoring of culturable bacteria from deep-sea sediment of the Nankai Trough, Leg 190 Ocean Drilling Program. *FEMS Microbiol Ecol* **48**: 357–367.

Valentine D. (2010). Emerging topics in marine methane biogeochemistry. *Annu Rev Mar Sci* **3**: 147–171.

Vermeij P, van der Steen RJ, Keltjens JT, Vogels GD, Leisinger T. (1996). Coenzyme F390 synthetase from Methanobacterium thermoautotrophicum Marburg belongs to the superfamily of adenylate-forming enzymes. *J Bacteriol* **178**: 505–510.

Vermeij P, Pennings JL, Maassen SM, Keltjens JT, Vogels GD. (1997). Cellular levels of factor 390 and methanogenic enzymes during growth of Methanobacterium thermoautotrophicum deltaH. *J Bacteriol* **179**: 6640–6648.

Wagner A, Cooper M, Ferdi S, Seifert J, Adrian L. (2012). Growth of Dehalococcoides mccartyi strain CBDB1 by reductive dehalogenation of brominated benzenes to benzene. *Environ Sci Technol* **46**: 8960–8968.

Wasmund K, Schreiber L, Lloyd KG, Petersen DG, Schramm A, Stepanauskas R *et al.* (2013). Genome sequencing of a single cell of the widely distributed marine subsurface Dehalococcoidia, phylum Chloroflexi. *ISME J* **8**: 383–397.

Webster G, Parkes RJ, Fry JC, Weightman AJ. (2004). Widespread occurrence of a novel division of bacteria identified by 16S rRNA gene sequences originally found in deep marine sediments. *Appl Environ Microbiol* **70**: 5708–5713.

Webster G, Parkes RJ, Cragg BA, Newberry CJ, Weightman AJ, Fry JC. (2006). Prokaryotic community composition and biogeochemical processes in deep subseafloor sediments from the Peru Margin. *FEMS Microbiol Ecol* **58**: 65–85.

Webster G, Sass H, Cragg BA, Gorra R, Knab NJ, Green CJ *et al.* (2011). Enrichment and cultivation of prokaryotes associated with the sulphate-methane transition zone of diffusion-controlled sediments of Aarhus Bay, Denmark, under heterotrophic conditions. *FEMS Microbiol Ecol* **77**: 248–263.

Woyke T, Sczyrba A, Lee J, Rinke C, Tighe D, Clingenpeel S *et al.* (2011). Decontamination of MDA reagents for single cell whole genome amplification. *PLoS ONE* **6**: e26161.

Yamada T, Sekiguchi Y. (2009). Cultivation of uncultured chloroflexi subphyla: significance and ecophysiology of formerly uncultured Chloroflexi 'Subphylum I' with natural and biotechnological relevance. *Microbes Environ* **24**: 205–216.

Zhu F, Massana R, Not F, Marie D, Vaulot D. (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol Ecol* **52**: 79–92.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)