

ORIGINAL ARTICLE

Comparative genomics of two newly isolated *Dehalococcoides* strains and an enrichment using a genus microarray

Patrick KH Lee¹, Dan Cheng², Ping Hu³, Kimberlee A West¹, Gregory J Dick^{4,6}, Eoin L Brodie³, Gary L Andersen³, Stephen H Zinder⁵, Jianzhong He² and Lisa Alvarez-Cohen^{1,3}
¹Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA;
²Division of Environmental Science and Engineering, National University of Singapore, Singapore;
³Ecology Department, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA;
⁴Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA
and ⁵Department of Microbiology, Cornell University, Ithaca, NY, USA

Comparative genomics of *Dehalococcoides* strains and an enrichment were performed using a microarray targeting genes from all available sequenced genomes of the *Dehalococcoides* genus. The microarray was designed with 4305 probe sets to target 98.6% of the open-reading frames from strains 195, CBDB1, BAV1 and VS. The microarrays were validated and applied to query the genomes of two recently isolated *Dehalococcoides* strains, ANAS1 and ANAS2, and their enrichment source (ANAS) to understand the genome–physiology relationships. Strains ANAS1 and ANAS2 can both couple the reduction of trichloroethene, *cis*-dichloroethene (DCE) and 1,1-DCE, but not tetrachloroethene and *trans*-DCE with growth, whereas only strain ANAS2 couples vinyl chloride reduction to growth. Comparative genomic analysis showed that the genomes of both strains are similar to each other and to strain 195, except for genes that are within the previously defined integrated elements or high-plasticity regions. Combined results of the two isolates closely matched the results obtained using genomic DNA of the ANAS enrichment. The genome similarities, together with the distinct chlorinated ethene usage of strains ANAS1, ANAS2 and 195 demonstrate that closely phylogenetically related strains can be physiologically different. This incongruence between physiology and core genome phylogeny seems to be related to the presence of distinct reductive dehalogenase-encoding genes with assigned chlorinated ethene functions (*pceA*, *tceA* in strain 195; *tceA* in strain ANAS1; *vcrA* in strain ANAS2). Overall, the microarrays are a valuable high-throughput tool for comparative genomics of unsequenced *Dehalococcoides*-containing samples to provide insights into their gene content and dechlorination functions.

The ISME Journal (2011) 5, 1014–1024; doi:10.1038/ismej.2010.202; published online 13 January 2011

Subject Category: integrated genomics and post-genomic approaches in microbial ecology

Keywords: comparative genomics; microarray; *Dehalococcoides*; systems biology; bioremediation; chlorinated solvents

Introduction

Bacteria belonging to the genus *Dehalococcoides* are known for their ability to reductively dechlorinate the toxic and carcinogenic chlorinated ethenes to the innocuous product ethene (McCarty, 1997; Löffler and Edwards, 2006). This physiological trait

is vital to the bioremediation of many ground water resources in the USA and worldwide, and field-scale studies have previously demonstrated that the presence of *Dehalococcoides* spp. is necessary for conversion of chlorinated ethene to ethene (Ellis *et al.*, 2000; Hendrickson *et al.*, 2002; Major *et al.*, 2002). Bacteria that does not belong to the *Dehalococcoides* genus can also dechlorinate tetrachloroethene (PCE) and trichloroethene (TCE), but no other genus has been shown to reduce dichloroethene (DCE) isomers or vinyl chloride (VC) (Smidt and de Vos, 2004). To expedite the complete remediation of chlorinated ethene-contaminated sites, further understanding of the physiology, biochemistry, phylogeny and ecology of *Dehalococcoides* spp. is warranted.

A number of *Dehalococcoides* strains have been isolated over the years (Maymó-Gatell *et al.*, 1997;

Correspondence: J He, Division of Environmental Science and Engineering, National University of Singapore, Block E2-02-13, 1 Engineering Drive 3, Singapore.

E-mail: jianzhong.he@nus.edu.sg

or L Alvarez-Cohen, 726 Davis Hall, Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720-1710, USA.

E-mail: alvarez@ce.berkeley.edu

⁶Current address: Department of Geological Sciences, University of Michigan, Ann Arbor, MI 48109, USA.

Received 26 August 2010; revised 1 December 2010; accepted 1 December 2010; published online 13 January 2011

Adrian *et al.*, 2000; He *et al.*, 2003b, 2005; Müller *et al.*, 2004; Sung *et al.*, 2006; Bunge *et al.*, 2008; Cheng and He, 2009) and the ability to culture them has facilitated physiological characterization and genome sequencing. Thus, so far, the genomes of four *Dehalococcoides* strains have been sequenced and annotated (strains 195, CBDB1, BAV1 and VS) (Kube *et al.*, 2005; Seshadri *et al.*, 2005; McMurdie *et al.*, 2009). Consistent with their genome annotations, *Dehalococcoides* spp. are obligated to oxidize hydrogen as the electron donor, utilize acetate/CO₂ as main carbon sources and reduce halogenated organics as terminal electron acceptors (Kube *et al.*, 2005; Seshadri *et al.*, 2005; Löffler and Edwards, 2006; McMurdie *et al.*, 2009). The characterized strains of *Dehalococcoides* differ in their usage of halogenated organics, as dictated by the specific reductive dehalogenases (RDases) that they each carry (Löffler and Edwards, 2006). Four RDase-encoding genes with experimentally determined ethene dechlorination functions are currently known (*pceA*, *tceA*, *vcrA* and *bvcA*) (Magnuson *et al.*, 1998, 2000; Krajmalnik-Brown *et al.*, 2004; Müller *et al.*, 2004), making them useful biomarkers for field diagnostic purposes (Lee *et al.*, 2008; van der Zaan *et al.*, 2010). Recent genome comparison of the four sequenced *Dehalococcoides* strains has identified a core genome containing genes that share a high degree of synteny and encode essential metabolic functions (McMurdie *et al.*, 2009). Variations among the individual genomes are mostly concentrated in the high-plasticity regions (HPRs), which are characterized by presumptive insertion sequences, repeated elements, deletions, inversions and phage-like genes (McMurdie *et al.*, 2009). A large number of known and putative RDase-encoding genes are found in the HPRs of each strain, suggesting that horizontal gene transfer is a key mechanism for acquiring novel RDases and distinct dechlorination abilities (McMurdie *et al.*, 2009).

The genome sequence and annotation of the four *Dehalococcoides* strains have established a reference for the genomic characteristics of this genus. Using sequences from the four *Dehalococcoides* genomes, a microarray using unique probe sets to target all identified genes was designed, constructed and validated in this study. Microarray technology has successfully been applied in recent years in comparative genomic analyses of related strains of bacteria in a variety of genera (Poly *et al.*, 2004; Witney *et al.*, 2005; Cooke *et al.*, 2008; Boesten *et al.*, 2009; Castellanos *et al.*, 2009), including *Dehalococcoides* (West *et al.*, 2008). In this study, we report the physiology of two novel *Dehalococcoides* strains that have been isolated from an enrichment culture (ANAS) that has been stably dechlorinating trichloroethene to ethene for over 10 years (Richardson *et al.*, 2002; Holmes *et al.*, 2006). In addition, we perform a comparative genomic analysis of the two isolates (strains ANAS1 and ANAS2) and the enrichment culture using the microarrays, and correlate the

results to physiology to provide insights into the biology of *Dehalococcoides*.

Materials and methods

Bacterial cultures and isolation

Dehalococcoides ethenogenes 195 (Maymó-Gatell *et al.*, 1997) and *Dehalococcoides* sp. BAV1 (He *et al.*, 2003b) were grown in 100 ml liquid medium, in 160 ml serum bottles, with TCE or VC as previously described (He *et al.*, 2003b, 2007). The enrichment culture ANAS was maintained in a semi-batch mode, and enriched with lactate and TCE as previously described (Richardson *et al.*, 2002; Johnson *et al.*, 2005; Holmes *et al.*, 2006; Lee *et al.*, 2006; West *et al.*, 2008).

Strains ANAS1 and ANAS2 were isolated from the ANAS enrichment culture by dilution-to-extinction series in liquid medium and agar shakes as previously described (He *et al.*, 2003b). After undergoing the first serial dilution (10⁻¹ to 10⁻⁵) with TCE, acetate and hydrogen, the cultures were subjected to ampicillin (50 mg l⁻¹) treatments in the next three consecutive sets of serial dilution to facilitate the isolation process. Subsequently, the ANAS dilution series was amended with either TCE or VC without antibiotics to enrich for the two *Dehalococcoides* strains. Because VC-respiring *Dehalococcoides* tend to be more sensitive to oxygen than TCE-respiring *Dehalococcoides* (Amos *et al.*, 2008), the reducing strength of the medium was modified during the isolation process. Specifically, medium for the VC-grown dilution series was reduced with 0.2 mM L-cysteine, 0.2 mM sodium sulfide and 0.5 mM DL-dithiothreitol, whereas medium containing half of these concentrations was used for the TCE-grown dilution series. The colonies obtained from the VC-fed dilution series were again inoculated into medium with ampicillin (50 mg l⁻¹). This process resulted in the isolation of two new *Dehalococcoides* strains, ANAS1 and ANAS2. The activities of the new strains on PCE, TCE, 1,1-DCE, *trans*-DCE (tDCE), *cis*-DCE (cDCE) and VC (about 2 mM each in the liquid phase) were tested in 160-ml serum bottles in triplicate.

Analytical methods

Headspace samples of ethene and chlorinated ethenes were measured by gas chromatography as described previously (Lee *et al.*, 2006; Cheng *et al.*, 2010).

PCR, qPCR and clone libraries

Universal bacterial primers 8F and 1392R were used to obtain nearly complete 16S rRNA gene sequences from strains ANAS1 and ANAS2. The PCR protocol was as previously described (Löffler *et al.*, 2000). A 16S rRNA gene clone library was constructed using amplicons from each strain according to a previously described method (He *et al.*, 2003a).

Quantitative PCR was used to measure the concentrations of selected genes, and triplicate samples were analyzed using a previously described protocol (Cheng and He, 2009). Primer and probe sequences for the targeted bacterial 16S rRNA gene, *Dehalococcoides* 16S rRNA, *tceA* and *vcrA* genes have previously been published (Holmes *et al.*, 2006; Cheng and He, 2009).

Microarray design

The microarrays targeting all genes from the four sequenced *Dehalococcoides* genomes (195, CBDB1, BAV1 and VS) were produced by Affymetrix (Santa Clara, CA, USA) as prokaryotic midi format (format 100) photolithographic microarray chips. The design strategy was to target each gene within the genomes with a unique probe set that would not cross-hybridize with closely related genes. If a unique probe set could not be found for a gene, multiple genes were targeted by the same probe set. Probe sets consisted of 11 probe pairs (22 total probes) of 25-mer oligonucleotide probes that are distributed along the length of the respective gene. Each probe pair is made up of a perfect-matched probe and a corresponding single-mismatch probe in which the thirteenth nucleotide is a mismatch with the target to control for nonspecific hybridization. The microarrays also contained 24 positive control and 21 negative control probe sets as described previously (West *et al.*, 2008) to facilitate calibration and to resolve background signals.

The input sequences for the microarrays were downloaded from the National Center for Biotechnology Information and included *Dehalococcoides* sequences as well as all the GenBank sequences targeting putative RDases, known hydrogen-producing nitrogenases and hydrogenases, and a variety of enzymes involved in vitamin B₁₂, biotin and methionine biosynthesis pathways. The draft strain VS genome was evaluated for contaminating sequences using a tetra-nucleotide binning method (Dick *et al.*, 2009) with a 5 kb cutoff that resulted in 251 high-confidence *Dehalococcoides* contigs, 63 uncertain contigs and 6 contigs that were possible contaminants. A second independent filtering analysis was performed by aligning all the annotated genes in the strain VS draft genome against the genomes of strains 195, BAV1 and CBDB1, and retaining only those with at least 70% identity over at least 80% of the length of the gene. Only genes that ranked in the high-confidence category in the tetra-nucleotide binning analysis, or passed the alignment filter were included in the microarray design. All RDases in the draft genome of strain VS were manually analyzed. In total, 1455 open-reading frames from the strain VS draft genome were included in the design.

Microarray analysis

To obtain sufficient mass of genomic DNA (gDNA) from strains 195, BAV1, ANAS1 and ANAS2, multiple

100 ml cultures were grown in parallel. gDNA of the ANAS enrichment culture was collected from 30 ml of reactor sample. After the amended electron acceptors were reduced in each culture, cells were collected, pooled and the gDNA was extracted using either the Qiagen (Valencia, CA, USA) DNeasy Blood & Tissue kit or the AllPrep DNA/RNA Mini Kit (Qiagen) according to the manufacturer's recommendations. gDNA of *Anaeromyxobacter dehalogenans* strain 2CP-C (Thomas *et al.*, 2008) was purchased from the American Type Culture Collection (Manassas, VA, USA) to serve as a negative control.

The extracted gDNA from each culture was quantified using the PicoGreen double-stranded DNA quantitation kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. gDNA (1 µg) was prepared for each microarray according to the protocol described previously (West *et al.*, 2008), and triplicate analyses were performed for each culture. Briefly, the gDNA and positive control spike-mix was fragmented, biotin end-labeled and hybridized to the microarray. Following hybridization, microarrays were processed according to section 3 of the Affymetrix GeneChip Expression Analysis technical manual. Data analysis was performed as described previously (West *et al.*, 2008) using Affymetrix GeneChip software (Affymetrix) and the MAS5 algorithm. Each microarray was normalized by scaling the signal intensities of the positive control spike-mix to a target signal intensity of 2500 to allow comparison between microarrays. A gene was considered 'present' in a culture if the probe set across all three replicate samples had signal intensities greater than 140 and a *P*-value less than 0.05. The microarray data collected in this study are available in Supplementary Table S1.

Accession numbers

The 16S rRNA gene sequences of ANAS1 (HM241729) and ANAS2 (HM241730) were deposited in the GenBank. Their corresponding RDase-encoding gene sequences were also deposited under accession numbers HM241731 (*tceA* gene, ANAS1) and HM241732 (*vcrA* gene, ANAS2). The microarray data analyzed in this study (GSE23707) and information regarding the microarray platform (GPL10838) were deposited in the National Center for Biotechnology Information Gene Expression Omnibus database.

Results

Design of the *Dehalococcoides* genus microarray

To query the genomic content of unsequenced *Dehalococcoides*-containing samples, a microarray targeting the sequenced strains of *Dehalococcoides* was designed and constructed. Of the total 6096 genes identified across the four genomes, 6010 or 98.6% are targeted by probe sets on the microarrays (Supplementary Table S2). Over half of the genes

not targeted on the microarray (51 of 86) are from strain VS and were excluded because of the unavailability of a finished genome at the time of design (Supplementary Table S2).

In the design of the microarrays, unique probe sets targeting genes from only a single genome were sought to allow strain differentiation. In total, 4305 probe sets were designed to represent 6010 *Dehalococcoides* genes, including 68.5% unique probes targeting genes from a single genome (Figure 1). Genes in strains BAV1 and CBDB1 are targeted by a large number of common probe sets (1203), whereas strains 195 and VS are represented by a relatively large number of unique probe sets (1279 and 1190, respectively) (Figure 1). The limited number of unique probe sets for strains BAV1 and CBDB1 mostly target genes within the HPRs or integrated elements (IEs).

Microarray validation

To evaluate the hybridization efficiency and quantitative response of the designed probe sets, gDNA (1 µg) of strains 195 and BAV1 was applied to the microarrays as positive controls, resulting in 100% (1629 of 1629) detection for strain 195 genes and 99.7% (1430 of 1435) detection for strain BAV1 (Figure 2). Analytical reproducibility of the positive controls was also high as linear regression of the signal intensities of the positive probe sets between replicate samples yielded $r^2 > 0.94$ and a slope of ~ 1.0 (Supplementary Figures S1 and S2), and the average coefficient of variation for a probe set across replicates was $< 3.3\%$. As a negative control, gDNA of *A. dehalogenans* strain 2CP-C, a δ -Proteobacterium capable of reductive dechlorination of chlorophenols and distantly related to *Dehalococcoides* (Thomas *et al.*, 2008), was applied. Hybridization results showed only six probe sets responding positively, including five tRNAs and a

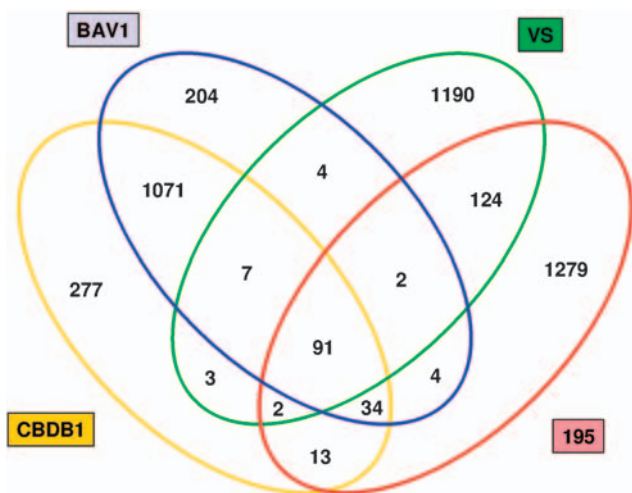


Figure 1 A Venn diagram showing the genome target distribution for the 4305 probe sets on the microarray.

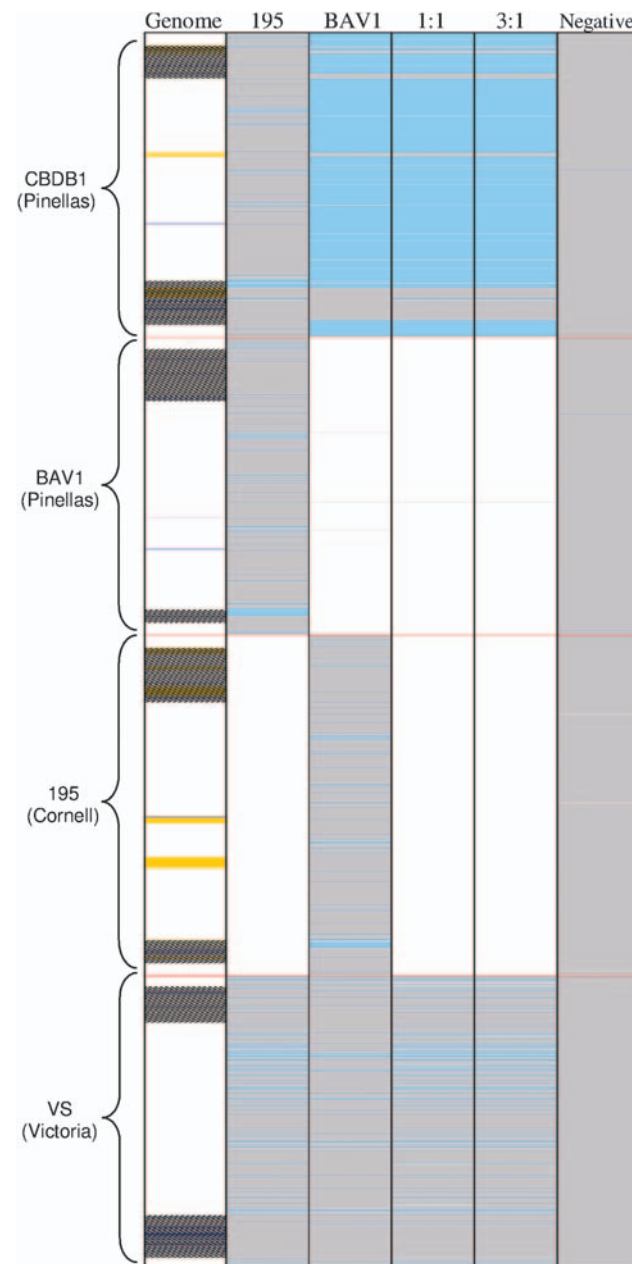


Figure 2 Microarray results for the positive controls (gDNA of strain 195 alone, strain BAV1 alone, strains 195 and BAV1 at 1:1 and 3:1 ratios by mass) and negative controls (gDNA of *Anaeromyxobacter dehalogenans* strain 2CP-C) from triplicate analyses are represented as genes in the four sequenced genomes (marked on the left along with the subgroup classification and separated by a red horizontal line). Each column represents a sample as indicated on the top and each row represents a gene where all the genes that are targeted by the microarrays (except those that cross hybridize) are depicted and arranged according to their physical location in the genome from top to bottom. Genes that are considered present are colored white or light blue and those that are absent are colored light gray. Because an unique probe set is not available for each gene, the color light blue is used to indicate genes that are present and targeted by the same probe sets as genes in strain 195 (in the strain 195 alone and negative samples), strain BAV1 (in the strain BAV1 alone sample) and strains 195 and/or BAV1 (in the 1:1 and 3:1 samples). The 'genome' column on the left depicts annotations that are of interest, including the two HPRs of each strain (dark shade), IEs (I to IX) of strain 195 and IEs (I, III and VII) of strain CBDB1 (orange), putative RDases (dark blue), and the *pceA*, *tceA*, *vcrA* and *bvcA* genes (purple).

525-bp hypothetical gene in strain CBDB1 (*cbdb_A1030*) (Figure 2). Cross-species hybridization with tRNAs is not surprising because of sequence conservation among bacteria. BLAST analysis against the strain 2CP-C genome returned no close homolog for *cbdb_A1030*. The probe sets for *cbdb_A1030* also responded positively when gDNA of strains 195 and BAV1 were hybridized, suggesting that this probe set is prone to nonspecific hybridization.

Not surprisingly, some minor cross-hybridization also occurred between *Dehalococcoides* strains, with 71 and 80 probe sets, targeting other strains, responded positively to the gDNA of strain 195 and strain BAV1, respectively (Supplementary Table S3). Overall, the total number of probe sets that exhibited nonspecific hybridization was relatively few in comparison with the total number of probe sets on the microarrays (151 of 4305 or 3.5%), and all cross-hybridizing probe sets were removed from subsequent analyses in this study.

The quantitative response of the microarrays was investigated in a mixed ratio experiment ranging from 1:1 to 3:1 mixture of strains 195 and BAV1 gDNA. Signal intensities increased linearly with increasing masses ($r^2 > 0.96$) (Supplementary Figure S3), and the gDNA mixtures had no effect on detection of genes in the two respective genomes (Figure 2).

Physiology of *Dehalococcoides* sp. strains ANAS1 and ANAS2

ANAS is an enrichment culture that dechlorinates TCE to ethene with lactate provided as electron donor and as carbon source. Previous studies have indicated that there are two distinct *Dehalococcoides* populations within this culture (Holmes *et al.*, 2006; Lee *et al.*, 2006). Dilution-to-extinction

methods were used in this study to isolate the two unique *Dehalococcoides* strains from ANAS, with strain ANAS1 resulting from the TCE dilution series and strain ANAS2 resulting from the VC series. Strain ANAS1 respire TCE to VC, whereas strain ANAS2 respire VC to ethene (Figure 3). Restriction digestion with enzymes *HhaI*, *RsaI* and *MspI* of the 72 clones obtained from a 16S rRNA gene clone library of the respective cultures provided identical restriction patterns, indicating the purity of the isolates. Confocal microscopy further corroborated that only one cell type was present (Supplementary Figure S4). Sequencing results showed that the 16S rRNA gene sequences of strains ANAS1 and ANAS2 are identical to each other and to those of strain 195 in the Cornell subgroup over 1354 bases.

PCR analyses targeting the known chlorinated ethene RDase genes (*pceA*, *tceA*, *bvcA* and *vcrA*) showed that strain ANAS1 only contains the *tceA* gene, whereas strain ANAS2 only contains the *vcrA* gene. This result is consistent with previous studies of the ANAS culture that detected the *tceA* and *vcrA* genes at different concentrations within the enrichment (Holmes *et al.*, 2006; Lee *et al.*, 2006). The *tceA* gene of strain ANAS1 is 98% (716/725 b.p.s.) identical with the *tceA* gene of strain 195, and the *vcrA* gene of strain ANAS2 is 99% (1475/1482 b.p.s.) identical with that of strain VS.

To further confirm that only a single population was present within each putative isolate, quantitative PCR targeting genes encoding bacterial 16S rRNA, *Dehalococcoides* 16S rRNA, *tceA* and *vcrA* were used to demonstrate that total bacterial cells of strains ANAS1 and ANAS2 closely match the *Dehalococcoides* cell numbers as well as the quantity of the respective functional RDase-encoding genes (Figure 4). In contrast, *Dehalococcoides* cells represented only a fraction of the

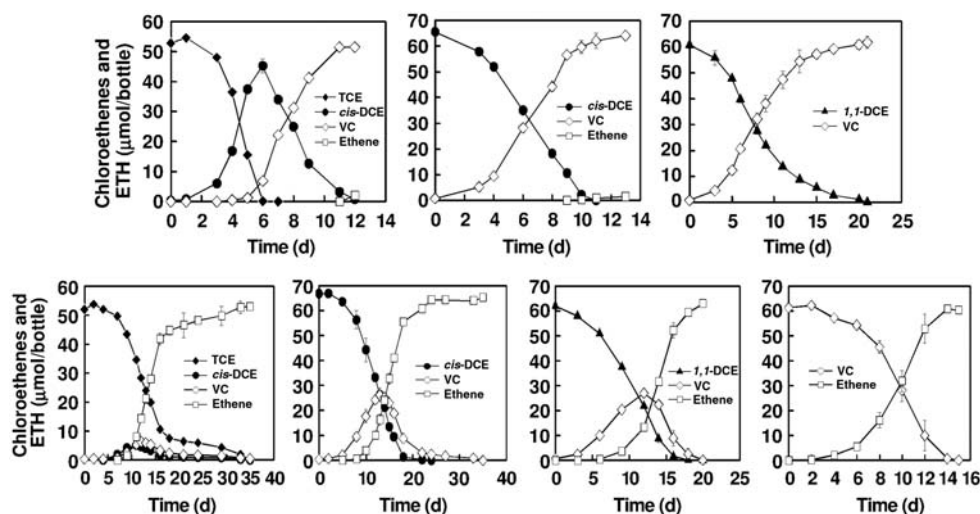


Figure 3 Reductive dechlorination profiles of strain ANAS1 (top panel) and strain ANAS2 (bottom panel) when provided with each growth-supporting chlorinated ethene as the starting electron acceptor. Error bars represent s.d.'s of triplicate samples. d, days; ETH, ethene.

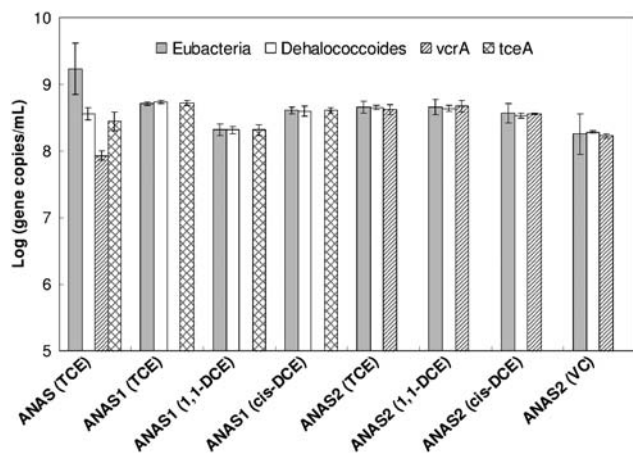


Figure 4 Quantification of the bacterial 16S rRNA gene, *Dehalococcoides* 16S rRNA gene, *vcrA* gene and *tceA* gene for the ANAS enrichment culture, strain ANAS1 and strain ANAS2 after the provided growth-supporting chlorinated ethenes (indicated in brackets) had been consumed. Error bars represent s.d.'s of triplicate analyses.

total bacterial populations in the ANAS enrichment culture, as did both the RDase-encoding genes (Figure 4).

Strain ANAS1 couples growth with the reduction of TCE, cDCE and 1,1-DCE to VC at average rates of 110, 60 and 30 $\mu\text{mol l}^{-1}$ per day, respectively (Figure 3), whereas PCE, tDCE and VC do not support growth. Strain ANAS2 couples growth with the reduction of TCE, cDCE 1,1-DCE and VC (43, 41, 38 and 51 $\mu\text{mol l}^{-1}$ per day, respectively) but not PCE and tDCE (Figure 3). Strains ANAS1 and ANAS2 exhibited different degradation patterns with significant buildup of cDCE before generation of VC for strain ANAS1, and complete reduction to ethene for strain ANAS2 (Figure 3).

Typical of other *Dehalococcoides* spp. (He *et al.*, 2007), strains ANAS1 and ANAS2 are non-motile disks of $\sim 1.0 \mu\text{m}$ in diameter (Supplementary Figure S4), with doubling times under optimal conditions (pH 7.2, temperature 30 °C) of ~ 20 h and yields on TCE of $4.7 (\pm 0.17) \times 10^8$ 16S rRNA gene copies per $\mu\text{mol Cl}^-$ and $2.5 (\pm 0.19) \times 10^8$ 16S rRNA gene copies per $\mu\text{mol Cl}^-$, respectively. Furthermore, strains ANAS1 and ANAS2 require hydrogen as the electron donor and acetate, and CO_2 as carbon sources (lactate, pyruvate, propionate and formate cannot be used). Neither strain can use 1,1-dichloroethane, chloroform, carbon tetrachloride, PCBs (Aroclor 1260 and CB-155), 2,4,6-trichlorophenol, pentachlorophenol, sulfate, sulfite, nitrate or nitrite as electron acceptors.

Genomic analysis of strains ANAS1 and ANAS2

The gDNA of the ANAS enrichment culture was previously analyzed using a microarray that only targets genes of strain 195, and the results indicated that the genomes of the *Dehalococcoides* population within that culture are similar to that of strain 195,

except for the IEs (West *et al.*, 2008). However, analysis of the enrichment culture did not allow for resolution of genes between the individual strains. Therefore, in this study, gDNA of strains ANAS1 and ANAS2 was independently hybridized to the *Dehalococcoides* genus microarrays to resolve the genomic content of the respective strains. Surprisingly, despite their physiological differences, both *Dehalococcoides* strains had similar genomic content that showed strong relatedness to strain 195 (83.1% for ANAS1 and 81.3% for ANAS2), with relatively few genes showing greater similarity to other *Dehalococcoides* strains (Figure 5). In fact, strains ANAS1 and ANAS2 had only 61 and 64 probe sets, respectively, which were positively detected for targets in *Dehalococcoides* strains other than 195 (Supplementary Table S4). This combined result was reproduced when gDNA of the ANAS enrichment culture was analyzed using the genus microarray (Figure 5), with 99% of the probe sets agreeing, confirming that hybridization results are unaffected by the presence of non-targeted organisms in the enrichment. The 1% of the inconsistent probe sets exhibited low-average signal intensities (< 300 U) in both the enrichment and the isolates (Supplementary Table S5), suggesting that poor hybridization may have caused the inconsistency.

Nearly all of the strain 195 genes not detected in strains ANAS1 and ANAS2 are located within the two HPRs (Figure 5) or more specifically, within the nine previously defined IEs (Seshadri *et al.*, 2005) (Supplementary Tables S6 and S7). The *pceA* gene that is within HPR1 is absent in both strains. Notable genes of strain 195 that are associated with the hydrogenase complexes, nitrogenases, the Wood-Ljungdal carbon fixation pathway, acetate assimilation, gluconeogenesis, the citric acid cycle and central metabolism (Seshadri *et al.*, 2005; Tang *et al.*, 2009) are present in strains ANAS1 and ANAS2.

Comparison between results for strains ANAS1 and ANAS2 showed that 97.7% of all the probe sets had a consistent present or absent detection, indicating that the number of probe sets that were strain specific (60 in strain ANAS1; 36 in strain ANAS2) is relatively few (Supplementary Table S8). One of the most notable differences between the two strains is the presence of the *tceA* gene, located within IE (I) of strain 195, in strain ANAS1 that is clearly absent in strain ANAS2. Furthermore, only the *vcrABC* genes (DhcVS_1289-91) of strain VS, but not any of the immediately neighboring genes and only a few of the genes within that HPR, are present in strain ANAS2 (Figure 5). Other differences between the two strains include genes that encode hypothetical proteins, putative RDases and genes that are within the HPRs or IEs of strain 195 or other sequenced genomes (Supplementary Table S8).

When the gDNA of the ANAS enrichment culture was hybridized to the strain 195-only microarrays, the tryptophan operon (DET1481-88) and the

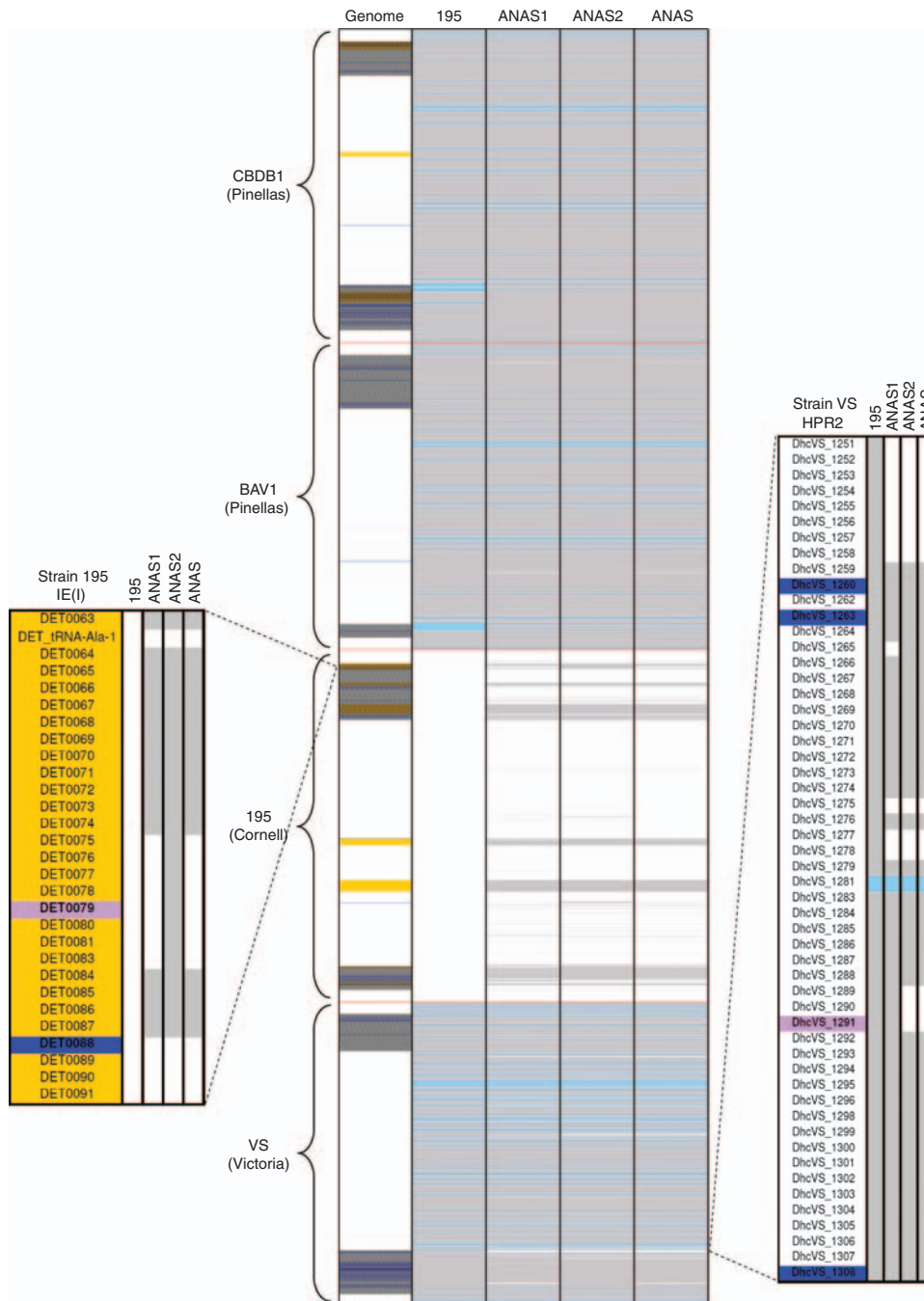


Figure 5 Microarray results for strains ANAS1, ANAS2 and the ANAS enrichment culture are represented as genes in the four sequenced genomes. Results of strain 195 are included as a reference. The figure legend is the same as Figure 2 and the color light blue is exclusively used to indicate genes that are targeted by the same probe sets as genes in strain 195. The magnified figure on the left shows the IE (I) of strain 195 (in which the *tceA* gene (DET0079) is located), and the magnified figure on the right shows the beginning section of the HPR2 of strain VS (in which the tryptophan operon (DhcVS_1251-58) and *vcrA* gene (DhcVS_1291) are located). The locus tags in bold highlight the RDase-encoding genes.

neighboring peptide-ABC transporter (DET1490-94) were not detected (West *et al.*, 2008). The absence of these strain 195 genes was confirmed using the genus microarray, whereas the tryptophan operon of strain VS (DhcVS_1251-58) was detected in ANAS, ANAS1 and ANAS2 (Figure 5). The strain VS peptide-ABC transporter genes (DhcVS_1266-70) were not detected in the three cultures. Other

detected genes not associated with strain 195 mostly encode hypothetical proteins, and complete operons were rarely found.

In addition to the functional *tceA* and *vcrA* genes, other putative RDase-encoding genes were found within strains ANAS1 and ANAS2 (Supplementary Table S9). In total, strain ANAS1 matched probe sets for 7 of the 101 putative RDase-encoding genes

identified in the sequenced genomes with six being associated with strain 195 and one from strain VS (DET0079, DET0088, DET0173, DET0180, DET1535, DET1545 and DhcVS_1314). Strain ANAS2 matched probe sets for five putative RDase genes (DET0088, DET0173, DET0180, DET1545 and DhcVS_1291), with only the *vcrA* gene from strain VS distinct from those in strain ANAS1. Neither strain contains RDase-encoding genes identical with those from strains CBDB1 or BAV1.

Discussion

A *Dehalococcoides* genus microarray targeting 98.6% of all genes from the four sequenced *Dehalococcoides* genomes was constructed and validated in this study. This high-throughput tool will facilitate the genomic study of *Dehalococcoides* strains that have not been sequenced as well as mixed populations and environmental samples that potentially contain multiple *Dehalococcoides* strains. However, although the microarray technology is a powerful tool for genomic analysis, it is important to remember that only known sequences can be targeted and gene arrangement cannot be inferred from the output data.

On the basis of multiple alignments of either the 16S rRNA gene (Hendrickson *et al.*, 2002) or the core genome of each sequenced strain (McMurdie *et al.*, 2009), strains CBDB1 and BAV1 are clustered in a subgroup called Pinellas, whereas strains 195 and VS are in the subgroups called Cornell and Victoria, respectively. Because strains CBDB1 and BAV1 have highly similar sequences (McMurdie *et al.*, 2009), the ability to design unique probe sets for these strains was reduced; hence, many of their genes are targeted by common probe sets (Figure 1). In contrast, the ability to resolve strains 195 and VS using unique probes is high because of the greater dissimilarity between these respective strains (Figure 1). Validation of the constructed microarrays using gDNA from strains 195 and BAV1 demonstrated high analytical reproducibility, satisfactory quantitative response and minimal numbers of false negatives (Figure 2 and Supplementary Figure S3).

The *Dehalococcoides* genus microarrays were used in this study to query the genomes of the two recently isolated *Dehalococcoides* strains, ANAS1 and ANAS2, and to compare with the genome of their source enrichment ANAS to deduce genome–physiology relationships. The close tracking between the sum of *tceA* and *vcrA* genes with the *Dehalococcoides* 16S rRNA gene in ANAS (Figure 4), as well as the presence of only one of these two RDase-encoding genes in the respective isolates confirmed that these are the two functionally important *Dehalococcoides* strains in the enrichment (Figure 4). In addition, the genes detected in the ANAS enrichment were an extremely close match to the sum of those detected in ANAS1 and

ANAS2 (Figure 5), further supporting that these two isolates represent the *Dehalococcoides* population in that culture. Physiological characterizations of the individual strains showed relatively similar functions in that they both couple TCE, cDCE and 1,1-DCE with growth but not PCE or tDCE (Figure 3), with strain ANAS2 exhibiting a relatively slower dechlorination rate. The main observed difference between the two strains is in VC dechlorination by strain ANAS2 (Figure 3). Although the genomes of strains ANAS1 and ANAS2 are relatively similar to each other and to strain 195, both strains lack the PCE dechlorination activity of strain 195 (Maymó-Gatell *et al.*, 1997). These results together demonstrate that phylogenetically closely related strains can be physiologically different (Figures 3 and 5), resembling the situation for strains CBDB1, BAV1 and other members of the Pinellas group, and suggest an important role for horizontal gene transfer of RDase-encoding genes in determining the physiology of different strains.

The differences in PCE and VC usage between strains 195, ANAS1 and ANAS2 are consistent with the enzymatic function of the known RDases. That is, the *pceA* gene present in strain 195 encodes for PCE reduction (Magnuson *et al.*, 1998), a capability observed only in that strain, whereas the *vcrA* gene (Müller *et al.*, 2004) is only detected in strain ANAS2, the only strain of the three capable of VC metabolism. Further, the *tceA* gene detected in both strains 195 and ANAS1 encodes for reduction of TCE and isomers of DCE, but has little activity on VC (Magnuson *et al.*, 1998, 2000), consistent with the observed capabilities of those strains. Whether an RDase other than *TceA* or *VcrA* is responsible for TCE dechlorination in strain ANAS2 is not presently known. It was noted previously that characterized chlorinated ethene RDase-encoding genes are not necessarily restricted to the *Dehalococcoides* subgroup from which the gene was first identified but can also occur in other subgroups (Sung *et al.*, 2006). Indeed, *Dehalococcoides* strains GT and FL2 are both members of the Pinellas subgroup, yet strain GT contains the *vcrA* gene (first identified in strain VS of the Victoria subgroup) (Müller *et al.*, 2004; Sung *et al.*, 2006) and strain FL2 contains the *tceA* gene (first identified in strain 195 of the Cornell subgroup) (Magnuson *et al.*, 1998; He *et al.*, 2005). Physiologically, as is consistent with the presence of the functional RDases, strain GT possesses the ability to metabolically dechlorinate VC, whereas strain FL2 does not (He *et al.*, 2005; Sung *et al.*, 2006). In this study of strains ANAS1 and ANAS2, both the *tceA* and *vcrA* genes were found to be associated with members of the Cornell subgroup (Figure 5).

The nitrogenase operon, present in strain 195 but not the other sequenced strains (Seshadri *et al.*, 2005; McMurdie *et al.*, 2009) and previously demonstrated to be functional (Lee *et al.*, 2009), seems prevalent in the Cornell subgroup as strains ANAS1 and ANAS2 both carry the nitrogen-fixing

genes. If the nitrogenase operon is in fact restricted to the Cornell subgroup, then genes of this operon can serve as phylogenetic biomarkers for that subgroup. Genes of strain 195 that are absent in strains ANAS1 and ANAS2 are concentrated within the defined HPRs or IEs (Seshadri *et al.*, 2005; McMurdie *et al.*, 2009) (Figure 5). Analysis of the four sequenced strains has indicated that the HPRs are hotspots for genomic rearrangements and are the locations of many strain-specific genes (McMurdie *et al.*, 2009). Therefore, it is not surprising that strains ANAS1 and ANAS2 lack many of the strain 195 genes located in these regions. Conversely, it is expected that these two strains should each possess their own unique strain-specific genes within their respective HPRs near the origin of replication. These strain-specific sequences and their genomic arrangements will have to be determined by *de novo* sequencing. Strain-specific genes located in genomic islands that interrupt the synteny of the stable core genome of a genus have also been reported in many other bacteria, including the marine cyanobacterium *Prochlorococcus* (Rocap *et al.*, 2003; Coleman *et al.*, 2006) and the sulfate-reducing *Desulfovibrio* (Walker *et al.*, 2009). Genes in these islands can encode functions that influence fitness (Rocap *et al.*, 2003; Coleman *et al.*, 2006; Walker *et al.*, 2009), and similarly for *Dehalococcoides*, the putative RDase-encoding genes within genomic islands often determine the type of environmental contaminants that can be degraded by specific strains.

Only a small number of *Dehalococcoides* genes outside of strain 195 were detected in strains ANAS1 and ANAS2 (Figure 5). Of metabolic interest is the tryptophan operon, located in the HPR2 of all the sequenced strains (McMurdie *et al.*, 2009) that code for proteins that synthesize an amino acid that is important in the membrane-anchoring proteins of RDases (Krajmalnik-Brown *et al.*, 2004; Kube *et al.*, 2005). In comparing the four sequenced genomes, it was noted that the tryptophan operon of strain 195 is unusually similar to that of the Pinellas strains, and that sequences in that region were phylogenetically incongruent with the rest of the core genomes (McMurdie *et al.*, 2009). However, ANAS1, ANAS2 and ANAS contain tryptophan operons similar to strain VS of the Victoria subgroup (Figure 5), suggesting that the evolutionary history of this operon includes exchange between more than one *Dehalococcoides* subgroup.

A large number of putative RDase-encoding genes are present in the sequenced strains (McMurdie *et al.*, 2009) and other *Dehalococcoides*-containing cultures (Hölscher *et al.*, 2004; Waller *et al.*, 2005). Some of the RDases are unique to each strain, whereas others possess homologs in one or more strains (McMurdie *et al.*, 2009). For strains ANAS1 and ANAS2, five to seven of the previously sequenced RDases were found, but the total number of RDases in each strain might be higher if novel

RDases are present. DET0180 and DET1545 are two putative RDases found in both strains that also have orthologous RDases in all four sequenced strains (McMurdie *et al.*, 2009). The substrates of these putative RDases are not currently known, but previous transcriptional analyses have indicated that DET0173, DET0180, DET1545 (common in ANAS1 and ANAS2) and DET1535 might be important at later growth stages (Rahm *et al.*, 2006; Johnson *et al.*, 2008).

In conclusion, a comparative genomics analysis was performed on two newly isolated *Dehalococcoides* strains along with their source enrichment culture using a microarray targeting genes of all four sequenced *Dehalococcoides* genomes to identify important genomic features in unsequenced strains. Integration of the microarray results with the physiology of the cultures demonstrated that strains that are phylogenetically related on the genome level can be physiologically incongruent, suggesting that the dechlorination functions of *Dehalococcoides* cultures are independent of phylogenetic affiliation but are dictated by a small number of RDase-encoding genes. Therefore, in contaminated environments, the presence and expression of the key RDase-encoding genes are essential biomarkers for monitoring the physiology of *Dehalococcoides* and the overall cleanup process.

Acknowledgements

We thank Ryan West for his programming expertise. This research was supported by the Strategic Environmental Research and Development Program through grant ER-1587 and the NIEHS Superfund Basic Research Project ES04705-19. DC and JZH were supported by the Singapore National Research Foundation for its Competitive Research Program Funding Scheme through grant NRF-CRP5-2009-05. GJD was supported by DOE Genomics: GTL project Grant No. DE-FG02-05ER64134 (Office of Science) to Jillian F Banfield.

References

- Adrian L, Szewzyk U, Wecke J, Görisch H. (2000). Bacterial dehalorespiration with chlorinated benzenes. *Nature* **408**: 580–583.
- Amos BK, Ritalahti KM, Cruz-Garcia C, Padilla-Crespo E, Löffler FE. (2008). Oxygen effect on *Dehalococcoides* viability and biomarker quantification. *Environ Sci Technol* **42**: 5718–5726.
- Boesten RJ, Schuren FHJ, de Vos WM. (2009). A *Bifidobacterium* mixed-species microarray for high resolution discrimination between intestinal bifidobacteria. *J Microbiol Methods* **76**: 269–277.
- Bunge M, Wagner A, Fischer M, Andreesen JR, Lechner U. (2008). Enrichment of a dioxin-dehalogenating *Dehalococcoides* species in two-liquid phase cultures. *Environ Microbiol* **10**: 2670–2683.
- Castellanos E, Aranaz A, Gould KA, Linedale R, Stevenson K, Alvarez J *et al.* (2009). Discovery of stable and variable differences in the *Mycobacterium*

- avium* subsp. *paratuberculosis* type I, II, and III genomes by pan-genome microarray analysis. *Appl Environ Microbiol* **75**: 676–686.
- Cheng D, Chow WL, He J. (2010). A *Dehalococcoides*-containing co-culture that dechlorinates tetrachloroethene to *trans*-1,2-dichloroethene. *ISME J* **4**: 88–97.
- 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.
- Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, DeLong EF *et al.* (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**: 1768–1770.
- Cooke FJ, Brown DJ, Fookes M, Pickard D, Ivens A, Wain J *et al.* (2008). Characterization of the genomes of a diverse collection of *Salmonella enterica* serovar typhimurium definitive phage type 104. *J Bacteriol* **190**: 8155–8162.
- Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP *et al.* (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10**: R85.
- Ellis DE, Lutz EJ, Odom JM, Buchanan Jr RJ, Bartlett CL, Lee MD *et al.* (2000). Bioaugmentation for accelerated *in situ* anaerobic bioremediation. *Environ Sci Technol* **34**: 2254–2260.
- He J, Ritalahti KM, Aiello MR, Löffler FE. (2003a). Complete detoxification of vinyl chloride by an anaerobic enrichment culture and identification of the reductively dechlorinating population as a *Dehalococcoides* species. *Appl Environ Microbiol* **69**: 996–1003.
- He J, Holmes VF, Lee PKH, Alvarez-Cohen L. (2007). Influence of vitamin B₁₂ and cocultures on the growth of *Dehalococcoides* isolates in defined medium. *Appl Environ Microbiol* **73**: 2847–2853.
- He J, Ritalahti KM, Yang KL, Koenigsberg SS, Löffler FE. (2003b). Detoxification of vinyl chloride to ethene coupled to growth of an anaerobic bacterium. *Nature* **424**: 62–65.
- 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.
- Hendrickson ER, Payne JA, Young RM, Starr MG, Perry MP, Fahnstock S *et al.* (2002). Molecular analysis of *Dehalococcoides* 16S ribosomal DNA from chloroethene-contaminated sites throughout North America and Europe. *Appl Environ Microbiol* **68**: 485–495.
- Holmes VF, He J, Lee PKH, Alvarez-Cohen L. (2006). Discrimination of multiple *Dehalococcoides* strains in a trichloroethene enrichment by quantification of their reductive dehalogenase genes. *Appl Environ Microbiol* **72**: 5877–5883.
- Hölscher T, Krajmalnik-Brown R, Ritalahti KM, von Wintzingerode F, Görisch H, Löffler FE *et al.* (2004). Multiple nonidentical reductive-dehalogenase-homologous genes are common in *Dehalococcoides*. *Appl Environ Microbiol* **70**: 5290–5297.
- Johnson DR, Brodie EL, Hubbard AE, Andersen GL, Zinder SH, Alvarez-Cohen L. (2008). Temporal transcriptomic microarray analysis of '*Dehalococcoides ethenogenes*' strain 195 during the transition into stationary phase. *Appl Environ Microbiol* **74**: 2864–2872.
- Johnson DR, Lee PKH, Holmes VF, Fortin AC, Alvarez-Cohen L. (2005). Transcriptional expression of the *tceA* gene in a *Dehalococcoides*-containing microbial enrichment. *Appl Environ Microbiol* **71**: 7145–7151.
- Krajmalnik-Brown R, Hölscher T, Thomson IN, Saunders FM, Ritalahti KM, Löffler FE. (2004). Genetic identification of a putative vinyl chloride reductase in *Dehalococcoides* sp. strain BAV1. *Appl Environ Microbiol* **70**: 6347–6351.
- 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.
- Lee PKH, He J, Zinder SH, Alvarez-Cohen L. (2009). Evidence for nitrogen fixation by '*Dehalococcoides ethenogenes*' strain 195. *Appl Environ Microbiol* **75**: 7551–7555.
- Lee PKH, Johnson DR, Holmes VF, He J, Alvarez-Cohen L. (2006). Reductive dehalogenase gene expression as a biomarker for physiological activity of *Dehalococcoides* spp. *Appl Environ Microbiol* **72**: 6161–6168.
- Lee PKH, Macbeth TW, Sorenson Jr KS, Deeb RA, Alvarez-Cohen L. (2008). Quantifying genes and transcripts to assess the *in situ* physiology of '*Dehalococcoides*' spp. in a trichloroethene-contaminated groundwater site. *Appl Environ Microbiol* **74**: 2728–2739.
- Löffler FE, Edwards EA. (2006). Harnessing microbial activities for environmental cleanup. *Curr Opin Biotechnol* **17**: 274–284.
- Löffler FE, Sun Q, Li J, Tiedje JM. (2000). 16S rRNA gene-based detection of tetrachloroethene-dechlorinating *Desulfuromonas* and *Dehalococcoides* species. *Appl Environ Microbiol* **66**: 1369–1374.
- Magnuson JK, Romine MF, Burris DR, Kingsley MT. (2000). Trichloroethene reductive dehalogenase from *Dehalococcoides ethenogenes*: sequence of *tceA* and substrate range characterization. *Appl Environ Microbiol* **66**: 5141–5147.
- Magnuson JK, Stern RV, Gossett JM, Zinder SH, Burris DR. (1998). Reductive dechlorination of tetrachloroethene to ethene by a two-component enzyme pathway. *Appl Environ Microbiol* **64**: 1270–1275.
- Major DW, McMaster ML, Cox EE, Edwards EA, Dworatzek SM, Hendrickson ER *et al.* (2002). Field demonstration of successful bioaugmentation to achieve dechlorination of tetrachloroethene to ethene. *Environ Sci Technol* **36**: 5106–5116.
- Maymó-Gatell X, Chien YT, Gossett JM, Zinder SH. (1997). Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. *Science* **276**: 1568–1571.
- McCarty PL. (1997). Microbiology—breathing with chlorinated solvents. *Science* **276**: 1521–1522.
- 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.
- Müller 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.
- Poly F, Threadgill D, Stintzi A. (2004). Identification of *Campylobacter jejuni* ATCC 43431-specific genes by whole microbial genome comparisons. *J Bacteriol* **186**: 4781–4795.
- Rahm BG, Morris RM, Richardson RE. (2006). Temporal expression of respiratory genes in an enrichment

- culture containing *Dehalococcoides ethenogenes*. *Appl Environ Microbiol* **72**: 5486–5491.
- Richardson RE, Bhupathiraju VK, Song DL, Goulet TA, Alvarez-Cohen L. (2002). Phylogenetic characterization of microbial communities that reductively dechlorinate TCE based upon a combination of molecular techniques. *Environ Sci Technol* **36**: 2652–2662.
- Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA *et al.* (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.
- Seshadri R, Adrian L, Fouts DE, Eisen JA, Phillippy AM, Methe BA *et al.* (2005). Genome sequence of the PCE-dechlorinating bacterium *Dehalococcoides ethenogenes*. *Science* **307**: 105–108.
- Smidt H, de Vos WM. (2004). Anaerobic microbial dehalogenation. *Annu Rev Microbiol* **58**: 43–73.
- Sung Y, Ritalahti KM, Apkarian RP, Löffler FE. (2006). Quantitative PCR confirms purity of strain GT, a novel trichloroethene-to-ethene-respiring *Dehalococcoides* isolate. *Appl Environ Microbiol* **72**: 1980–1987.
- Tang YJ, Yi S, Zhuang WQ, Zinder SH, Keasling JD, Alvarez-Cohen L. (2009). Investigation of carbon metabolism in '*Dehalococcoides ethenogenes*' strain 195 by use of isotopomer and transcriptomic analyses. *J Bacteriol* **191**: 5224–5231.
- Thomas SH, Wagner RD, Arakaki AK, Skolnick J, Kirby JR, Shimkets LJ *et al.* (2008). The mosaic genome of *Anaeromyxobacter dehalogenans* strain 2CP-C suggests an aerobic common ancestor to the delta-proteobacteria. *PLoS ONE* **3**: e2103.
- van der Zaan B, Hannes F, Hoekstra N, Rijnaarts H, de Vos WM, Smidt H *et al.* (2010). Correlation of *Dehalococcoides* 16S rRNA and chloroethene-reductive dehalogenase genes with geochemical conditions in chloroethene-contaminated groundwater. *Appl Environ Microbiol* **76**: 843–850.
- Walker CB, Stolyar S, Chivian D, Pinel N, Gabster JA, Dehal PS *et al.* (2009). Contribution of mobile genetic elements to *Desulfovibrio vulgaris* genome plasticity. *Environ Microbiol* **11**: 2244–2252.
- Waller AS, Krajmalnik-Brown R, Löffler FE, Edwards EA. (2005). Multiple reductive-dehalogenase-homologous genes are simultaneously transcribed during dechlorination by *Dehalococcoides*-containing cultures. *Appl Environ Microbiol* **71**: 8257–8264.
- West KA, Johnson DR, Hu P, DeSantis TZ, Brodie EL, Lee PKH *et al.* (2008). Comparative genomics of '*Dehalococcoides ethenogenes*' 195 and an enrichment culture containing unsequenced '*Dehalococcoides*' strains. *Appl Environ Microbiol* **74**: 3533–3540.
- Witney AA, Marsden GL, Holden MTG, Stabler RA, Husain SE, Vass JK *et al.* (2005). Design, validation, and application of a seven-strain *Staphylococcus aureus* PCR product microarray for comparative genomics. *Appl Environ Microbiol* **71**: 7504–7514.

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