# Supplemental Information

## Insights into origins and function of the unexplored majority of the reductive dehalogenase gene family as a result of genome assembly and ortholog group classification

Olivia Molenda, Luz A. Puentes Jácome, Xuan Cao, Camilla L. Nesbø, Shuiquan Tang, Nadia Morson, Jonas Patron, Line Lomheim, David S. Wishart, and Elizabeth A. Edwards

**Contents**

**Additional data files (separate from this document)**
**File S1.** The phylogenetic, maximum likelihood, nucleotide tree of 551 *rdhA* genes in newick format. Most likely tree of 100 bootstraps.
**File S2.** Perl script which automates searching for gap solutions in scaffolds updated to use linking information from original version found in Tang et al. (2012) PLoS ONE 7(12):e52038
**Table S2. Excel File.** Order of putative reductive dehalogenase gene orthologous groups (OGs) found in genomes of twenty-two different strains of *Dehalococcoides mccartyi*.

## Supplemental Methods

### Metagenomic sequencing and genome assembly

DNA for metagenome sequencing was extracted from larger samples (40-615 mL) taken from the three functionally stable enrichment cultures described above: KB-1/VC-$H_2$ (40 mL culture sample), KB-1/TCE-MeOH (500 mL sample), KB-1/cDCE-MeOH (300 mL culture) and KB-1/1, 2-DCA-MeOH (615 mL sample). Extractions were conducted between February and May 2013. Cultures were filtered using Sterivex™ filters (Millipore 0.2 μm) and the DNA was extracted using the CTAB method (JGI bacterial genomic DNA isolation using CTAB protocol v.3). DNA was sequenced at the Genome Quebec Innovation Sequencing Centre using Illumina HiSeq 2500 technology. Paired-end sequencing with an insert size of ~400 bp and read length of ~150 bp provided roughly 50 million reads per culture. Additional mate-pair sequencing with insert size of ~8000 bp and read length of ~100 bp was conducted for the KB-1/TCE-MeOH and KB-1/1, 2-DCA-MeOH cultures where we had more DNA. In the case of metagenomic sequencing using short-read Next Generation Sequencing (NGS), we have demonstrated the utility of long-insert mate-pair data in resolving challenges in metagenomic assembly, especially those related to repeat elements and strain variation [1]. In this study, we applied a combination of Illumina mate-pair and paired-end sequencing data. Although other long-read sequencing technologies (e.g. PacBio SMRT sequencing, Nanopore sequencing, Illumina Synthetic Long-Read Sequencing Technique and 10x Genomics) are available, Illumina mate-pair sequencing is a cost-effective choice for the goal of obtaining both high sequencing depth and accuracy and

long-distance mate pair links. Raw sequences were trimmed with Trimmomatic [2] to remove bases of low quality and to remove adapters.

The *D. mccartyi* genomes were assembled in six steps as described below and illustrated in Figure S1. In Step 1, we generated ABySS unitigs with Illumina paired-end data using the ABySS assembler [3]. These unitigs were the main building blocks in the assembly of the complete genomes. ABySS assemblies generate both unitigs and contigs. Unlike contigs, unitigs are generated solely by overlapping k-mers and their assembly does not utilize the paired-end constraints. As a result, the maximum overlapping length between unitigs is the length of a k-mer size minus one. When using ABySS to assemble metagenomic data, we used the maximum k-mer size allowed, 96 bp, since the raw read length was 150bp, much longer than 96 bp. When configuring ABySS runs, it was critical to utilize the -c parameter, which specifies a cut-off, the minimum k-mer depth/coverage used in the assembly. Sequences/unitigs with k-mer coverage lower than this cut-off will be ignored in the assembly, which allows users to have good assemblies of high abundance organisms as the interferences caused by low abundance organisms (especially those of close relatives) and sequencing errors are removed. It is important to make sure that the k-mer depth of the sequences of the target genomes is higher than this threshold so that you have all sequences/unitigs you need to close the target genomes. For example, if the average k-mer depth of the target genome is 100, try 20 for the -c cut off. We used a combination of 16S rRNA amplicon sequencing and qPCR to get an idea of the relative abundances of our target organisms in our metagenome prior to attempting different ABySS assemblies.

In Step 2, we generated a genome-wide reference sequence for the target genome, which was subsequently used to guide the scaffolding of unitigs. This reference sequence can be obtained in different ways. If there are long-distance mate-paired data like we had for KB-1/TCE-MeOH and KB-1/1, 2-DCA-MeOH cultures, this reference sequence can be built *de novo*. We used two ways to build it: (1) using a standalone scaffolding program, SSPACE v. 2.0 [4], to generates scaffolds with ABySS contigs/unitigs utilizing the mate-paired constraints, (2) using ALLPATHS-LG [5] to generate the assembly with both paired-end and mate-pair data as inputs. ALLPATHS-LG turned out to be the most effective way in most cases. A publicly available closely related closed genome might also be able to serve as a reference genome to guide scaffolding of unitigs in the next step.

One major challenge in metagenomic assembly is cross-interference between closely related genomes, such as strains of the same species. The sequence similarity/dissimilarity between these closely related genomes tend to break the assembly. If a genome had a closely related genome interfering in its assembly, we attempted to assemble the genome with ALLPATHS-LG using both short-insert paired-end data and long-insert mate-pair data. For genomes that have no closely related genomes, a surprisingly effective way to assemble is to combine Digital Normalization [6] with ALLPATHS-LG. This approach reduces the data redundancy of raw sequences with Digital Normalization by k-mer and then one can assemble the resulting data with ALLPATHS-LG. In our case, we had multiple *D. mccartyi* strains in each metagenome and could not use Digital Normalization. ALLPATHS-LG was able to differentiate our similar strains because their abundances were distinct.

In Step 3, we used the best assembly generated from Step 2 to guide the scaffolding of unitigs generated from Step One. The scaffolding process is based on sequence comparison between the unitigs and the reference assembly by BLAST. After that, unitigs that have a k-mer depth significantly lower or higher than the average k-mer depth of the genome are removed. The basic assumption here is that unitigs with k-mer depth higher than average likely belong to repetitive sequences (such as rRNA gene operons and transposons) and unitigs with a k-mer depth lower than average are more likely to be strain specific. In other words, only unitigs with k-mer depth around average (we used 90%-110% of average) are kept; these unitigs are likely shared by closely related genomes. After that, the gap distance between the neighbouring unitigs is estimated based on the reference assembly. In brief, this process generated a scaffold consisting of unitigs shared by all closely related strains; this will serve as a backbone for subsequent gap resolution.

In Step 4, we identify all potential solutions for all gaps between unitigs in the scaffold. This step is performed by filling the gaps with the remaining unitigs mostly based on sequence overlap between unitigs; we published a similar process previously [1]. In the updated script, we have improved the process by incorporating paired-end and mate-pair link information between unitigs to help guide the searching process. The paired-end and mate-pair links were obtained by mapping raw reads against unitigs. Solutions identified this way fulfill the constraints of sequence overlap, paired-end links and mate-pair links. If there are multiple solutions to a gap and they have k-mer depth lower than the average, this suggests the presence of strain variation.

In the end, this step generates a closed assembly, having some gaps with multiple solutions in cases of strain variation.

In Step 5 we bin these multiple solutions caused by strain variation to different genomes based on sequencing depth or k-mer depth. For example, if there are always two solutions, one of k-mer depth of 60 and the other one of 40, we will assign all solutions with higher depth to one strain and the rest to the other strain. This approach is not feasible if the two strains happen to have similar abundance and similar sequencing depth. Things become more complicated when there are more than two strains; in such case, we only try to resolve the genome of the highest abundance by gathering solutions of highest k-mer depth. The editing of the genome sequences is facilitated by the use of Geneious v. 6.1 [7]. Finally, in Step 6 we polish the assembled genome by mapping raw reads back to the final assembly. SNPs caused by strain variation are identified. If possible, they are resolved based on abundance in the same principle as using k-mer depth to assign alternative solutions in Step 5.

In all cases multiple genomes could be closed from a single enrichment culture because the different populations of *D. mccartyi* were at different abundances (as inferred from read depth) at the time of sampling. Two complete genomes each containing a vinyl chloride reductase gene (*vcrA)* were closed from the KB-1/VC-$H_2$. The naming convention for genomes was based on the culture name (KB-1) electron acceptor (e.g., vinyl chloride or VC) and relative abundance (number 1 for highest abundance and so on). For example, the names of the strains from KB-1/VC-$H_2$ culture are *D. mccartyi* strains KBVC1 and KBVC2. Three genomes each containing vinyl chloride reductase *bvcA* were closed from the KB-1/1, 2-DCA enrichment culture further referred to as strain KBDCA1, KBDCA2 and KBDCA3. Two genomes each containing trichloroethene reductase *tceA* from KB-1/TCE-MeOH culture, strains KBTCE2 and KBTCE3. A *D. mccartyi* complete genome containing a *vcrA* gene was also assembled from KB-1/TCE-MeOH culture, named strain KBTCE1. In all cases low abundance strains of *D. mccartyi* could not be assembled implying that although eight genomes were closed, the total number of KB-1 *D. mccartyi* strains is at least eleven. The genomes were annotated using the RAST [8] and BASyS [9] servers; results were manually inspected and corrected where required. Additional searches for conserved domains were conducted using NCBI conserved domain search (E-value threshold of 0.01). The origin of replication was identified using Oriloc in R [10].

**Quantitative PCR (qPCR) Analysis**

Quantitative polymerase chain reaction (qPCR) was used to estimate the abundance of *rdhA*, and *D. mccartyi* sequences in each of the sequenced cultures. DNA samples were diluted 10, 50 or 100 times with sterile UV treated distilled water (UltraPure), and all subsequent sample manipulations were conducted in a PCR cabinet (ESCO Technologies, Gatboro, PA). Each qPCR reaction was run in duplicate. Four *Dehalococcoides* genes were targeted by qPCR: 1) the phylogenetic 16S rRNA gene for *Dehalococcoides* Dhc1f (5'-GATGAACGCTAGCGGCG-3') and Dhc264r (5'-CCTCTCAGACCAGCTACCGATCGAA-3') [11]; 2) the vinyl chloride reductase gene, *vcrA*, vcrA642f (5'-GAAAGCTCAGCCGATGACTC-3') and vcrA846r (5'-TGGTTGAGGTAGGGTGAAGG-3' ) [12]; 3) *bvcA* dehalogenase, bvcA318f (5'-ATTTAGCGTGGGCAAAACAG-3') and bvcA555r (5'- CCTTCCCACCTTGGGTATTT-3') [12]; and 4) *tceA* dehalogenase: tceA500f (5' TAATATATGCCGCCACGAATGG-3') and tceA795r(5'- AATCGTATACCAAGGCCCGAGG-3') [13]. Samples were also analysed using general bacteria 16S rRNA primers GenBac1055f (5'- ATGGCTGTCGTCAGCT-3') and GenBac1392r (5'- ACGGGCGGTGTGTAC-3') [14]. DNA samples were diluted 10, 50 or 100 times with sterile UltraPure distilled water, and all subsequent sample manipulations were conducted in a PCR cabinet (ESCO Technologies, Gatboro, PA). Each qPCR reaction was run in duplicate. Each qPCR run was calibrated by constructing a standard curve using known concentrations of plasmid DNA containing the gene insert of interest. The standard curve was run with 8 concentrations, ranging from 10 to $10^8$ gene copies/µL. All qPCR analyses were conducted using a CFX96 real-time PCR detection system, with a C1000 thermo cycler (Bio-Rad Laboratories, Hercules, CA). Each 20 µL qPCR reaction was prepared in sterile UltraPure distilled water containing 10 µL of EvaGreen® Supermix (Bio-Rad Laboratories, Hercules, CA), 0.5 µL of each primer (forward and reverse, each from 10 µM stock solutions), and 2 µL of diluted template (DNA extract or standard plasmids). The thermocycling program was as follows: initial denaturation at 95ºC for 2 min, followed by 40 cycles of denaturation at 98ºC for 5s, annealing at 60ºC (for 16S rRNA and *vcrA*, *bvcA* genes, respectively) or 58 ºC for *tceA* or 55 ºC for General Bacteria followed by extension for 10s at 72 ºC. A final melting curve analysis was conducted at the end of the program. $R^2$ values were 0.99 or greater and efficiency values 80-110%.

**Amplicon Sequencing and Analysis**

For microbial community analysis, amplicon sequencing was performed on extracted DNA, which was amplified by PCR using general primers for the 16S rRNA gene. The universal primer set, 926f (5'-AAACTYAAAKGAATTGACGG-3') and 1392r (5'-ACGGGCGGTGTGTRC-3'), targeting the V6-V8 variable region of the 16S rRNA gene from bacteria and archaea, as well as the 18S rRNA gene from Eukaryota, was used [15]. The purified PCR products were sent to the McGill University and Genome Quebec Innovation Centre, where they were checked for quality again, pooled and subject to unidirectional sequencing (*i.e.*, Lib-l chemistry) of the 16S gene libraries, using the Roche GS FLX Titanium technology (Roche Diagnostics Corporation, Indianapolis, IN). One to three independent 100 µL PCR amplification reactions were preformed per sample. Each PCR reaction was set up in sterile Ultra-Pure $H_2O$ containing 50uL of PCR mix (Thermo Fisher Scientific, Waltham, MA), 2 µL of each primer (forward and reverse, each from 10 µM stock solutions), and 4 µL of DNA extract. PCR reactions were run on a MJ Research PTC-200 Peltier Thermal Cycler (Bio-Rad Laboratories, Hercules, CA) with the following thermocycling program; 95 °C, 3 min; 25 cycles of 95 °C 30 s, 54 °C 45 s, 72 °C 90 s; 72 °C 10 min; final hold at 4 °C (modified from [16]). The forward and reverse primers included adaptors (926f: CCATCTCATCCCTGCGTGTCTCCGACTCAG and, 1392r: CCTATCCCCTGTGTGCCTTGGCAGTCTCAG), and the reverse primer also included 10bp multiplex identifiers (MID) for distinguishing multiple samples pooled within one sequencing region. The PCR products were verified on a 2% agarose gel and replicates were combined and purified using GeneJET$^{TM}$ PCR Purification Kit (Fermentas, Burlington, ON), according to the manufacturer's instructions. The concentrations of PCR products were determined using a NanoDrop ND-1000 Spectrophotometer at a wavelength of 260 nm (NanoDrop Technologies, Wilmington, DE). The concentrations and qualities of the final PCR products were also evaluated by running them on 2% agarose gels, and comparing band intensities to those from a serial dilution of ladders with known DNA concentrations.

**Taxonomic assignments of 16S rRNA amplicon sequences**

The raw DNA sequences obtained from the sequencing center were processed using the Quantitative Insights Into Microbial Ecology (QIIME v1.5.0) pipeline [17] with default settings, unless stated otherwise. Only sequences of length between 300 and 500 bp, and with homopolymers shorter than 8 bases were processed for downstream analysis. After filtering,

sequences were de-multiplexed into respective samples based on their individual MID. Sequences were further clustered into distinct 16S rRNA gene-based Operational Taxonomic Units (OTUs) using the UCLUST algorithm [18], similarity threshold of 0.97 and the Green Genes database (version 13.5) [19]. Taxonomy was assigned to each OTU by the Ribosomal Database Project (RDP) classifier [20].

**Time since divergence analysis**

The time since divergence was calculated using the number of mutations (identified from phylogenetic alignments and trees) and two different mutation rates: the universal estimate of bacterial mutation in natural environments [21] and a known divergence time of 16 years between the separation of strain *D. mccartyi* 195 [22] as an isolate from its parent culture "DONNA2" as used in McMurdie et al. 2011[23]. A doubling time of 2 was used for *D. mccartyi* calculated from an average of reported doubling times of 2.4 [24] 0.8 [22], 2.5 [25], and 2.2 days [26].

## Figures & Tables

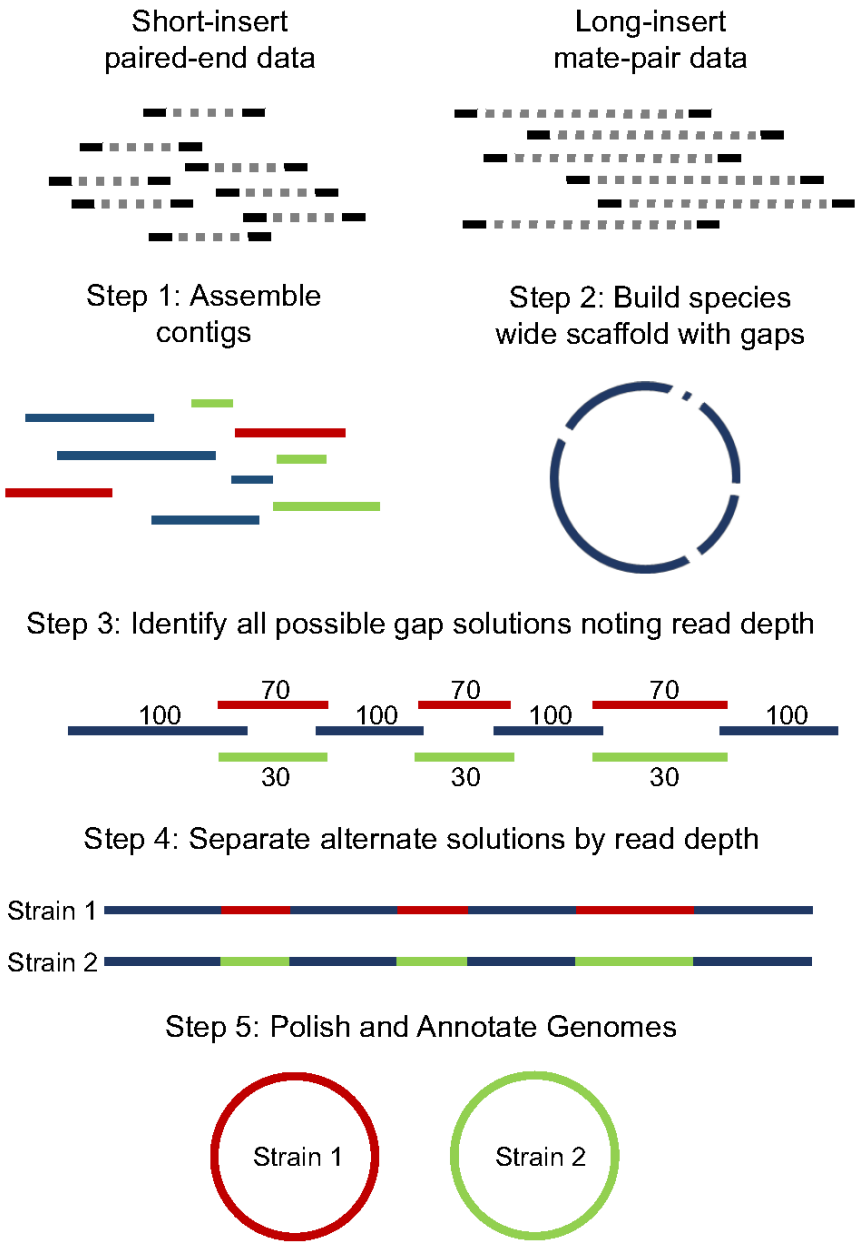


**Figure S1.** Schematic flow chart of workflow used to assemble *Dehalococcoides mccartyi* genomes from KB-1 metagenomes.

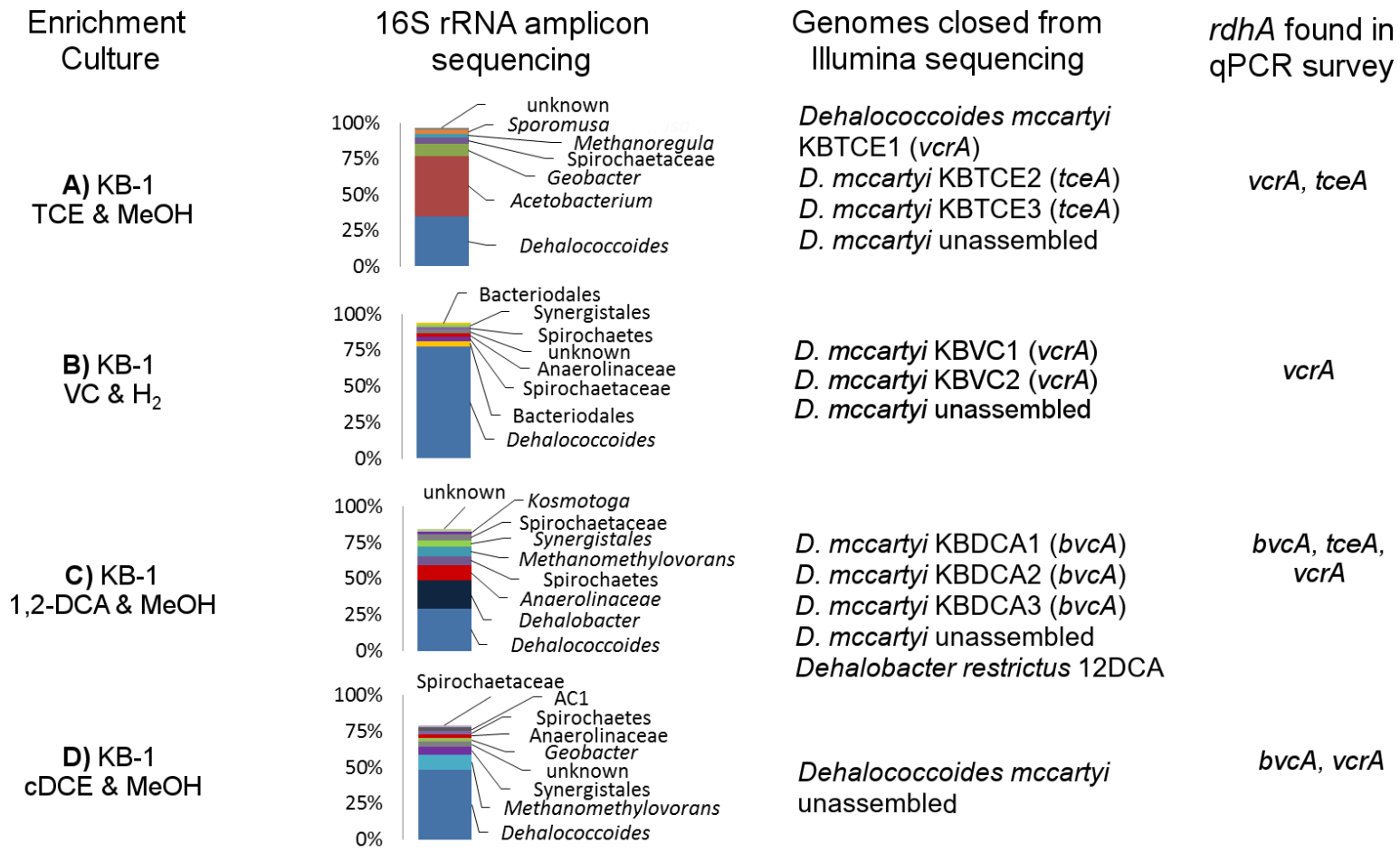


**Figure S2.** Culture composition of KB-1 enrichments and genomes closed from 16S rRNA amplicon sequencing, Illumina metagenomic sequencing and qPCR of *rdhA* genes.

DNA was extracted from four KB-1 enrichment cultures. The same DNA was split and analysed using 16S rRNA amplicon sequencing, (bar charts), Illumina mate-pair and paired-end assembly and genome closing (genomes listed by rank abundance and strain name if closed) and surveyed for *rdhA* genes using qPCR. Genes found above the detection limit are listed. See supplemental Table S1 for tabular qPCR results

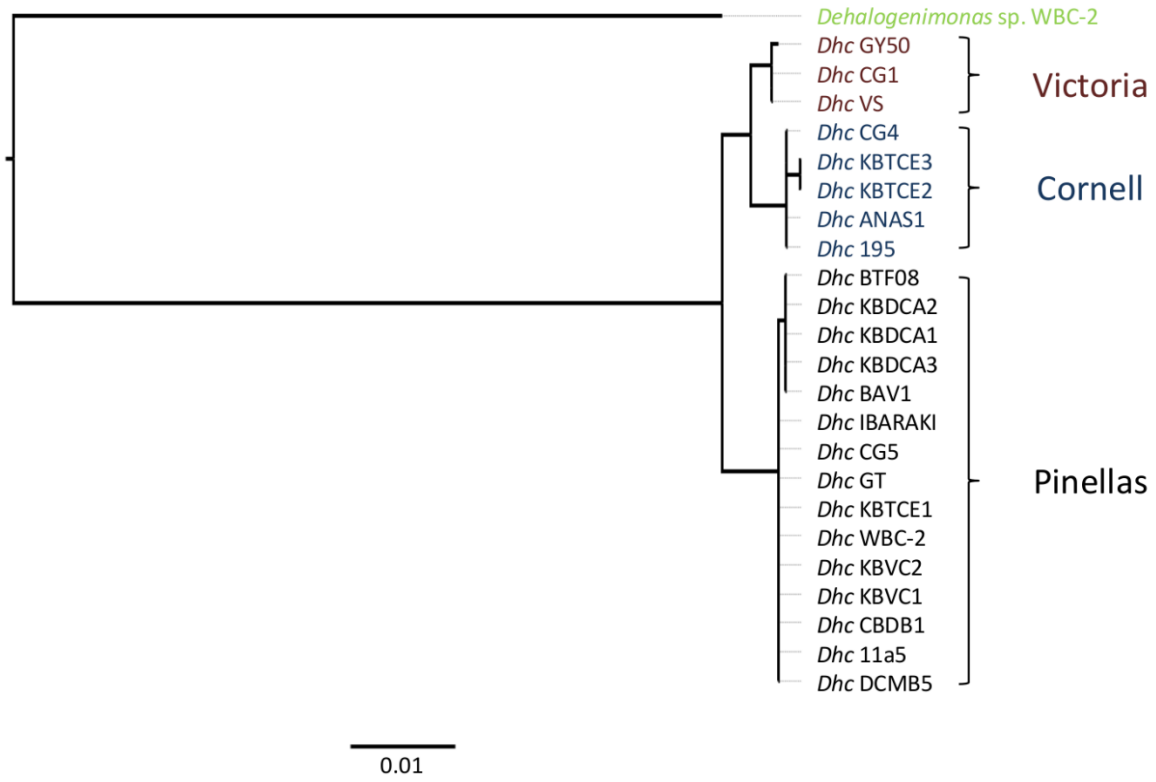


**Figure S3. Maximum likelihood phylogenetic tree (of 100 bootstraps) of nucleotide alignment of 16S rRNA in *Dehalococcoides mccartyi* (*Dhc*) strains.**
*D. mccartyi* genomes available in NCBI and *Dehalogenimonas* sp. WBC-2 used as out-group. Scale indicates number of nucleotide substitutions per site. *Dhc* strains identified by clade.

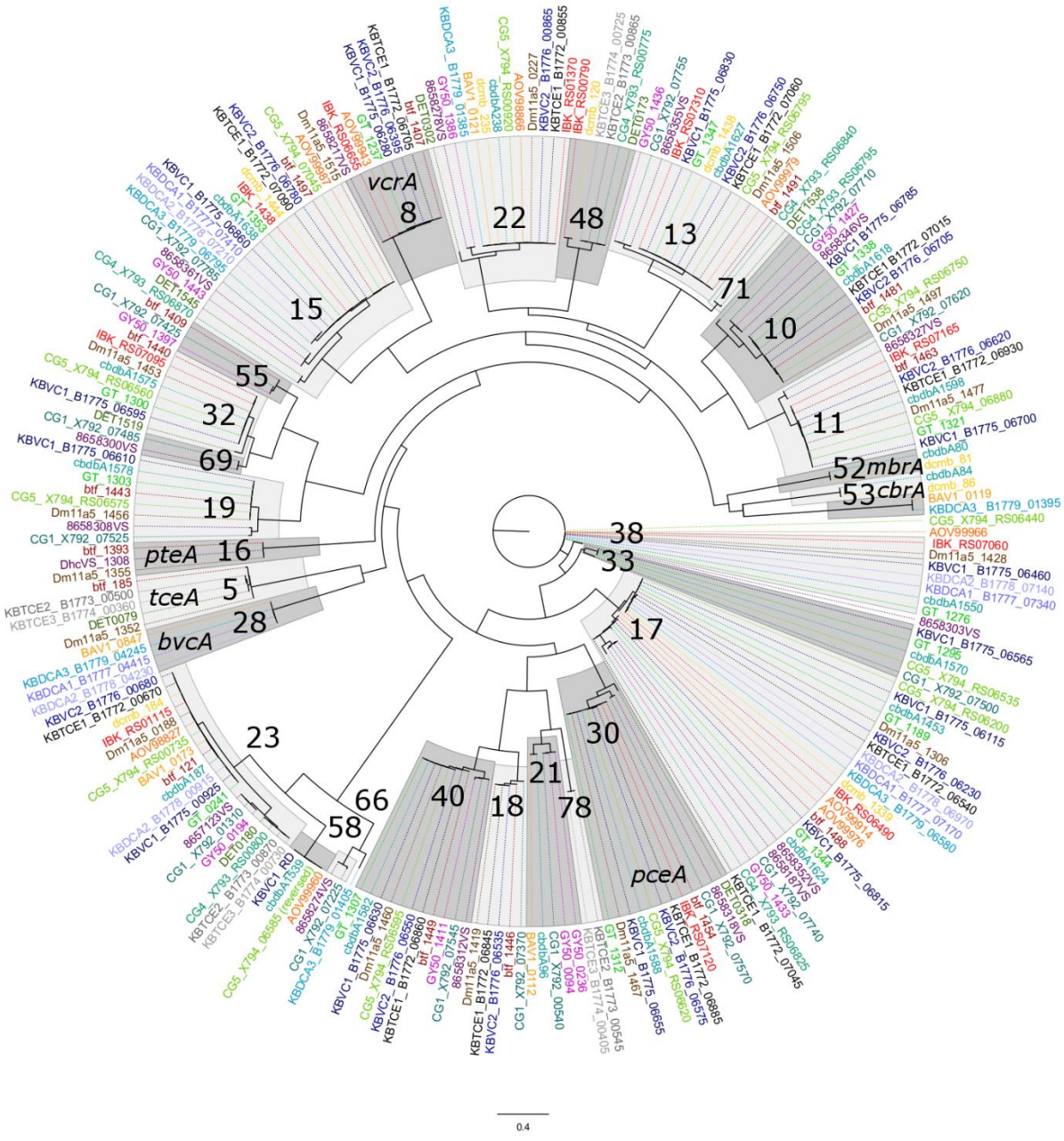


**Figure S4. Phylogenetic nucleotide tree of reductive dehalogenase genes (nucleotide sequences) from *D. mccartyi* closed genomes.**

Most likely tree of 100 bootstraps. Scale indicates number of substitutions per site. Orthologous groups (OGs) of dehalogenases with upwards of 90% amino acid identity are highlighted and identified by number. OGs containing a functionally characterized representative are annotated by dehalogenase gene name. RdhA are coloured by strain of origin. Strain name is listed before locus tag if not included in locus tag.

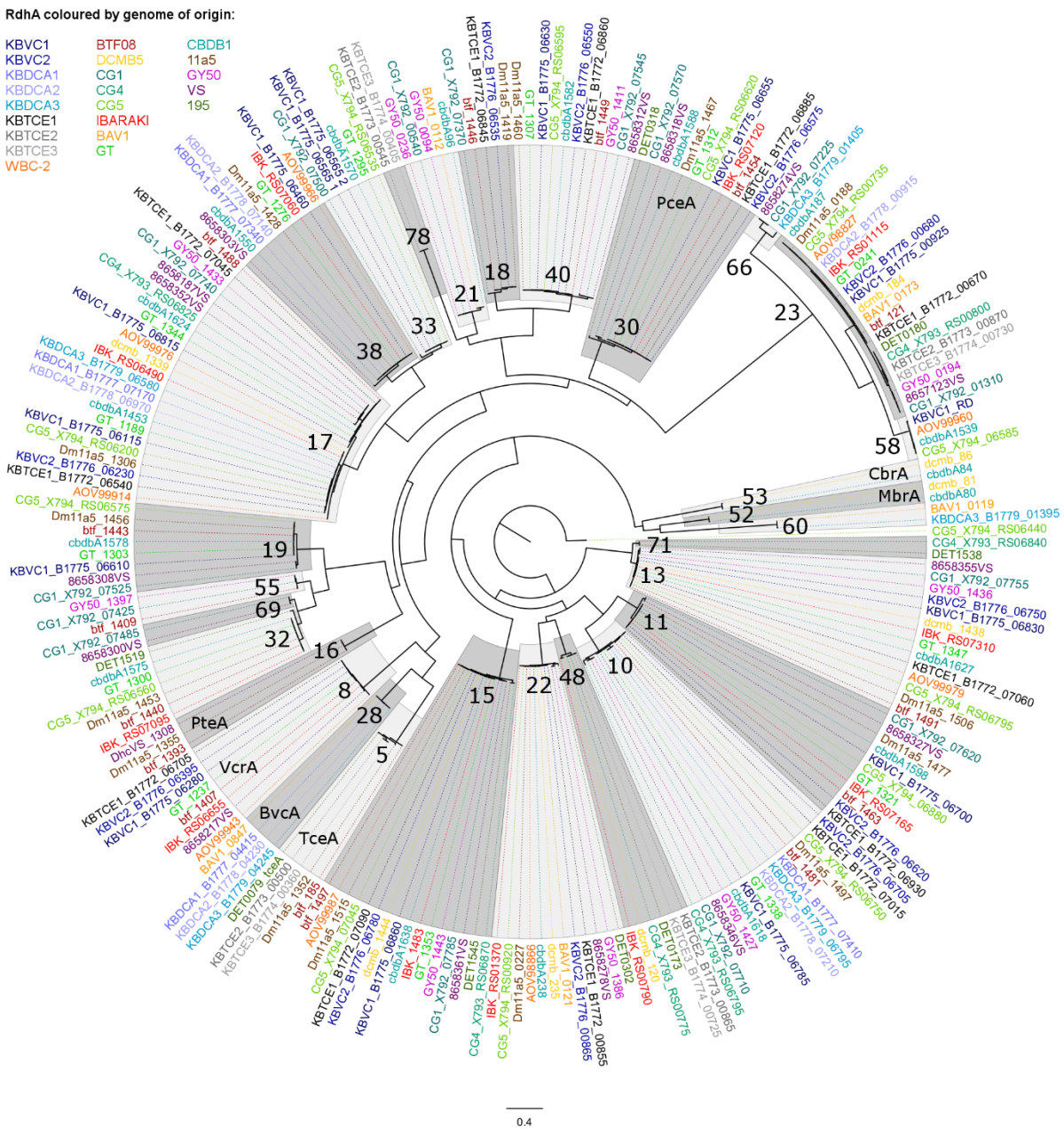


**Figure S5. Phylogenetic amino acid tree of reductive dehalogenases from *D. mccartyi* closed genomes.** Most likely tree of 100 bootstraps. Scale indicates number of substitutions per site. Orthologous groups (OGs) of dehalogenases with upwards of 90% amino acid identity are highlighted and identified by number. OGs containing a functionally characterized representative are annotated by dehalogenase name. RdhA sequences are coloured by genome they originated from. RdhA are named by NCBI locus tag. *D. mccartyi* strain name is indicated before locus tag unless included in locus tag.

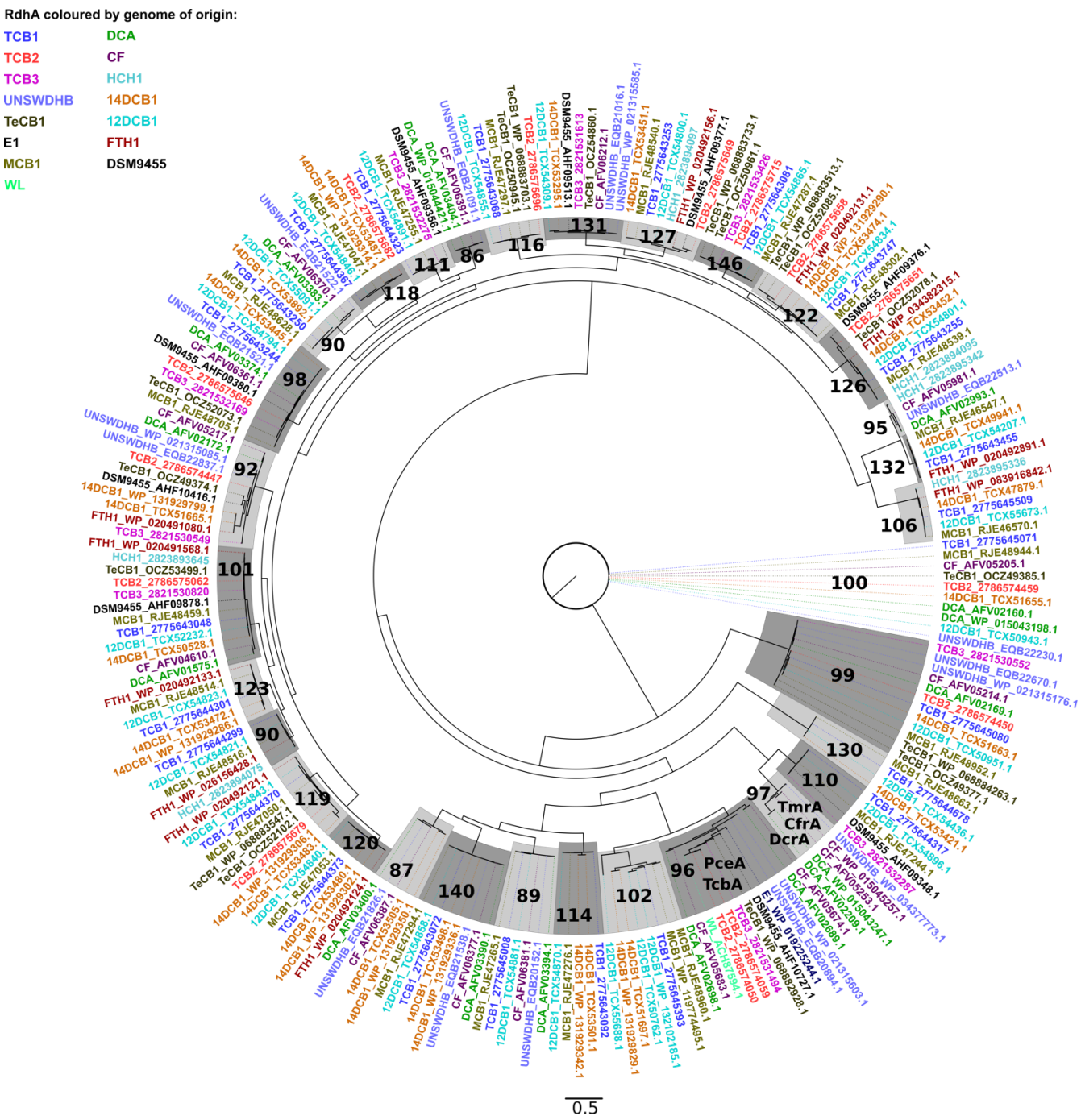


**Figure S6. Phylogenetic amino acid tree of reductive dehalogenases from *Dehalobacter* genomes.** Most likely tree of 100 bootstraps. Scale indicates number of substitutions per site. Orthologous groups (OGs) of dehalogenases with upwards of 90% amino acid identity are highlighted and identified by number. OGs containing a functionally characterized representative are annotated by dehalogenase name. RdhA sequences are coloured by genome of origin. RdhA are named by NCBI protein ID or IMG gene ID.

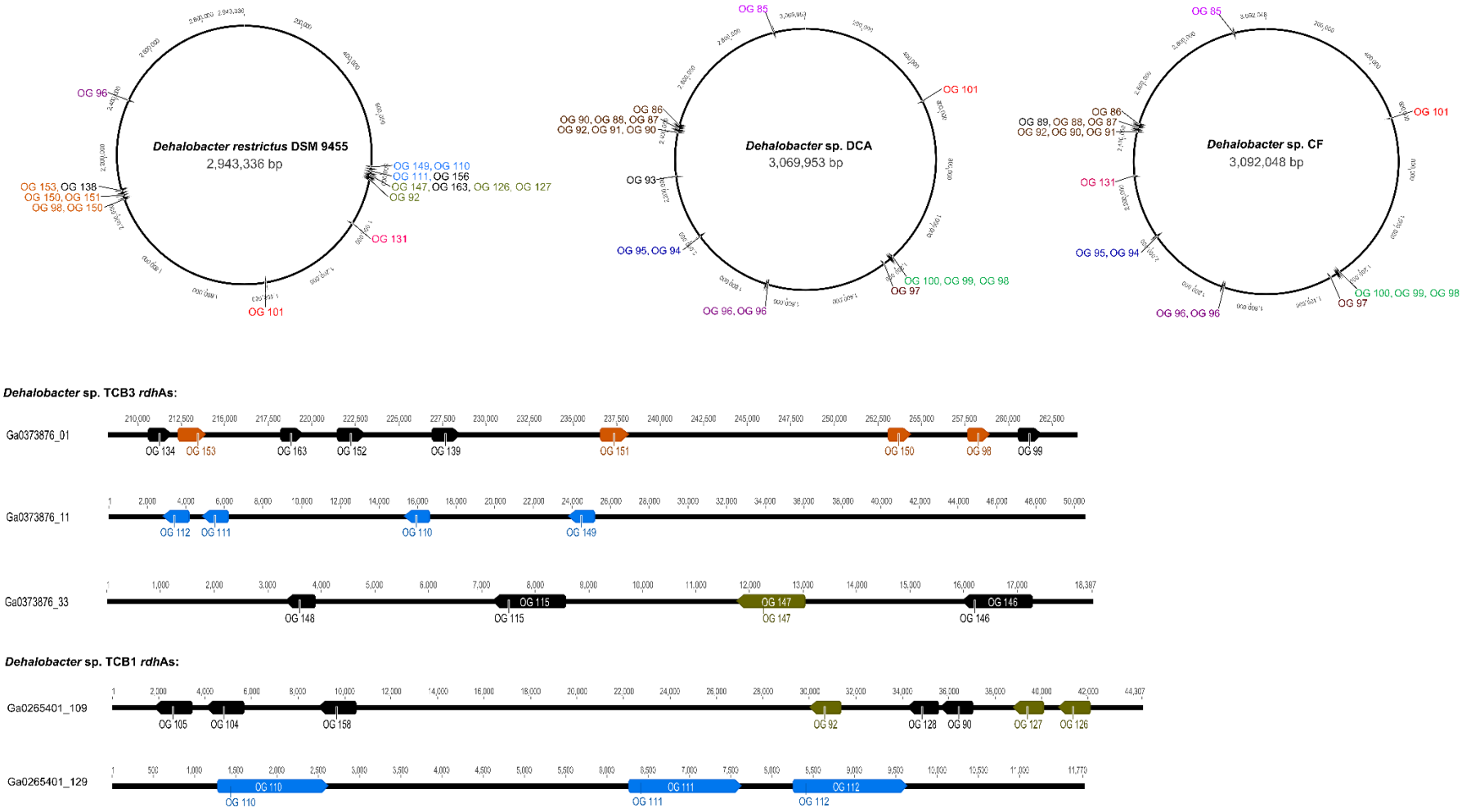


**Figure S7. Evidence of synteny of *rdhA* in *Dehalobacter restrictus* genomes.** In each genome or contig, *rdhA* are labeled by Ortholog Group (OG) number; co-located and shared *rdhA*s (among genomes) are colour-coded.

**Table S1. qPCR survey of *Dehalococcoides mccartyi* 16S rRNA gene (*Dhc*), General Bacteria 16S rRNA, (GenBac) and reductive dehalogenases: *vcrA*, *bvcA* and *tceA* of four different KB-1 enrichment cultures. ND - not detected NT- not tested**

| | | | | Raw data - SQ (starting quantity) (copies/uL) | | | | | Concentration in culture (copies/mL) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Culture** | **Amount filtered (mL)** | **Dilution factor for sample** | **Elution volume after DNA extraction (uL)** | ***vcrA* SQ** | ***bvcA* SQ** | ***tceA* SQ** | ***Dhc* SQ** | **GenBac SQ** | ***vcrA*** | ***bvcA*** | ***tceA*** | ***Dhc*** | **GenBac** |
| KB-1/1,2-DCA-MeOH sample 11:10 | 40 | 10 | 50 | 6761 | 150007 | 29615 | 45702 | 626868 | 8.45E+04 | 1.88E+06 | 3.70E+05 | 5.71E+05 | 7.84E+06 |
| KB-1/1,2-DCA-MeOH sample 1 1:10 | 40 | 10 | 50 | 6621 | 133025 | 29201 | 74777 | 545799 | 8.28E+04 | 1.66E+06 | 3.65E+05 | 9.35E+05 | 6.82E+06 |
| KB-1/1,2-DCA-MeOH sample 1 1:50 | 40 | 50 | 50 | 1524 | 21897 | 3977 | 11583 | 127699 | 9.53E+04 | 1.37E+06 | 2.49E+05 | 7.24E+05 | 7.98E+06 |
| KB-1/1,2-DCA-MeOH sample 1 1:50 | 40 | 50 | 50 | 1711 | 21125 | nt | 12343 | 127051 | 1.07E+05 | 1.32E+06 | | 7.71E+05 | 7.94E+06 |
| KB-1/1,2-DCA-MeOH sample 2 1:10 | 40 | 10 | 50 | 170565 | 2073701 | 887853 | 1024617 | 3996939 | 2.13E+06 | 2.59E+07 | 1.11E+07 | 1.28E+07 | 5.00E+07 |
| KB-1/1,2-DCA-MeOH sample 2 1:10 | 40 | 10 | 50 | 170764 | 2255491 | 828537 | 1022573 | 3916842 | 2.13E+06 | 2.82E+07 | 1.04E+07 | 1.28E+07 | 4.90E+07 |
| KB-1/1,2-DCA-MeOH sample 2 1:50 | 40 | 50 | 50 | 45550 | 574361 | 120877 | 265142 | 981316 | 2.85E+06 | 3.59E+07 | 7.55E+06 | 1.66E+07 | 6.13E+07 |
| KB-1/1,2-DCA-MeOH sample 2 1:50 | 40 | 50 | 50 | 47177 | 585780 | 125475 | 278127 | 1051573 | 2.95E+06 | 3.66E+07 | 7.84E+06 | 1.74E+07 | 6.57E+07 |
| KB-1/cDCE-MeOH 1:10 | 40 | 10 | 50 | 85917 | 306010 | ND | 235024 | 630396 | 1.07E+06 | 3.83E+06 | ND | 2.94E+06 | 7.88E+06 |
| KB-1/cDCE-MeOH 1:10 | 40 | 10 | 50 | 85548 | 327799 | ND | 286083 | 655866 | 1.07E+06 | 4.10E+06 | ND | 3.58E+06 | 8.20E+06 |
| KB-1/cDCE-MeOH 1:50 | 40 | 50 | 50 | 21990 | 81368 | ND | 44930 | 149034 | 1.37E+06 | 5.09E+06 | ND | 2.81E+06 | 9.31E+06 |
| KB-1/cDCE-MeOH 1:50 | 40 | 50 | 50 | 23477 | 82195 | ND | 46705 | 171444 | 1.47E+06 | 5.14E+06 | ND | 2.92E+06 | 1.07E+07 |
| KB-1/VC-$H_2$ 1:10 | 40 | 10 | 50 | 7015011 | ND | ND | 3423443 | 4154129 | 8.77E+07 | ND | ND | 4.28E+07 | 5.19E+07 |
| KB-1/VC-$H_2$ 1:10 | 40 | 10 | 50 | 6988920 | ND | ND | 3380919 | 4210962 | 8.74E+07 | ND | ND | 4.23E+07 | 5.26E+07 |
| KB-1/VC-$H_2$ 1:50 | 40 | 50 | 50 | 1886096 | ND | ND | 923854 | 1072650 | 1.18E+08 | ND | ND | 5.77E+07 | 6.70E+07 |
| KB-1/VC-$H_2$ 1:50 | 40 | 50 | 50 | 1975811 | ND | nt | 995994 | 1075217 | 1.23E+08 | ND | | 6.22E+07 | 6.72E+07 |
| KB-1/TCE-MeOH 1:50 | 400 | 50 | 50 | 7006510 | ND | 235535 | 9655782 | 15516397 | 4.38E+07 | ND | 1.47E+06 | 6.03E+07 | 9.70E+07 |
| KB-1/TCE-MeOH 1:50 | 400 | 50 | 50 | 6723999 | ND | 230317 | 9821546 | 13875805 | 4.20E+07 | ND | 1.44E+06 | 6.14E+07 | 8.67E+07 |
| KB-1/TCE-MeOH 1:100 | 400 | 100 | 50 | 6359129 | ND | 236495 | nt | nt | 7.95E+07 | ND | 2.96E+06 | nt | nt |
| KB-1/TCE-MeOH 1:100 | 400 | 100 | 50 | 3704187 | ND | | nt | nt | 4.63E+07 | ND | | nt | nt |

**Table S2 is in accompanying Excel file**

**Table S3. Number of mutations incurred since Most Recent Common Ancestor (MRCA) of *D. mccartyi* and select reductive dehalogenase genes.** myr- million years
Sample calculation in Table S3.

| Divergence | Method and Reference | # mutations | Time (myr) |
|---|---|---|---|
| Dehalococcoidia MRCA | Core genes tree (Fig. 4) | $8.6e^{05}$ | 0.28-3.21 |
| *D. mccartyi* MRCA (clades split) | Core gene tree (Fig. 4) | $8.4e^{04}$ | 0.03-0.31 |
| OG 13 and OG 71 MRCA | *rdhA* gene tree | $1.5e^{05}$ | 0.05-0.58 |
| OG 15 MRCA | *rdhA* gene tree | $9.6e^{04}$ | 0.03-0.36 |
| OG 34 (duplicate) MRCA | *rdhA* gene tree | $7.3e^{04}$ | 0.02-0.27 |
| OG 5 (*tceA*) MRCA | *rdhA* gene tree | $5.9e^{03}$ | 0.00-0.22 |
| OG 8 (*vcrA*) MRCA | *rdhA* gene tree [23] | $5.0e^{03}$ | 0.00-0.19 |
| OG 28 (*bvcA*) MRCA | *rdhA* gene tree | $2.8e^{03}$ | 0.00-0.10 |
| OG 52 (*mbrA*) MRCA | *rdhA* gene tree | $3.5e^{03}$ | 0.00-0.13 |
| OG 53 (*cbrA*) MRCA | *rdhA* gene tree | $2.0e^{03}$ | 0.00-0.07 |
| OG 16 (*pteA*) MRCA | *rdhA* gene tree | $1.0e^{03}$ | 0.00-0.04 |

**Table S4. Parameters used to estimate divergence age (4-part Table).** Data shown in Tables S4-1 to S4-3 were used to calculate final divergence times in Table S4-4. Table S4-1: Mutation rates used in divergence calculations (1) estimated universal bacterial error rate (2) empirically determined *Escherichia coli* mutation rate and (3) empirically determined mutation rate between *D. mccartyi* DONNA2 and 195 over 16 years. Table S4-2: Calculation for average Dehalococcoidia genome size, Table S4-3: Calculation for *D. mccartyi* average doubling time. Table S4-4: Example calculated divergence times in years for three different mutation rates (found in Table S2-1). *Dhc – Dehalococcoides, Dhg – Dehalogenimonas.*

**Table S4-1** **Mutation Rates**

| | **bacteria** | ***E. coli*** | ***D. mccartyi* DONNA2/strain 195** | unit |
|---|---|---|---|---|
| Reference | Ochman et al 1999 | Drake et al 1998 | McMurdie et al 2011 | |
| Overall error rate nt/genome/generation | 1.00E-09 | 5.40E-09 | 2.08405E-06 | $bp^{-1}$ |
| # bp per mutation | 1.00E+09 | 185185185.2 | 479835.6062 | bp |
| Average Dehalococcoidia genome size | 1439363 | | | bp |
| Replications per single fixed neutral mutation per genome | 694.8 | 128.7 | 0.3 | replications |
| *D. mccartyi* doubling time (days) | 2 | | | days |
| **Time estimated for single mutation (days)** | 1.37E+03 | 2.54E+02 | 6.58E-01 | days |
| years | 3.76 | 0.70 | 0.00 | years |

**Table S4-2**

| Species | Genome size (bp) |
|---|---|
| *D. mccartyi* GT | 1360154 |
| *D. mccartyi* CBDB1 | 1395502 |
| *D. mccartyi* 195 | 1469720 |
| *D. mccartyi* BAV1 | 1341892 |
| *D. mccartyi* CG5 | 1362151 |
| *D. mccartyi* WBC-2 | 1374583 |
| *D. mccartyi* VS | 1413462 |
| *D. mccartyi* CG4 | 1382308 |
| *D. mccartyi* GY50 | 1407418 |
| *D. mccartyi* DCMB5 | 1431902 |
| *D. mccartyi* IBARAKI | 1451902 |
| *D. mccartyi* BTF08 | 1451056 |
| *D. mccartyi* 11a5 | 1461973 |
| *D. mccartyi* CG1 | 1486678 |
| *D. mccartyi* CG3 | 1521286 |
| *D. mccartyi* KBDCA1 | 1428463 |
| *D. mccartyi* KBDCA2 | 1394319 |
| *D. mccartyi* KBDCA3 | 1337486 |
| *D. mccartyi* KBTCE1 | 1388914 |
| *D. mccartyi* KBTCE2 | 1329198 |
| *D. mccartyi* KBTCE3 | 1271604 |
| *D. mccartyi* KBVC1 | 1359904 |
| *D. mccartyi* KBVC2 | 1337731 |
| *Dehalogenimonas* sp. WBC-2 | 1725730 |
| *Dehalogenimonas lykanthroporepellens* | 1686510 |
| *Dehalogenimonas alkenigignens* | 1851580 |
| **Average *Dhc* genome** | **1398244** |
| **Average *Dhg* genome** | **1754607** |
| **Weighted Average Genome Size** | **1439363** |

**Table S4-3**

| ***D. mccartyi* strain** | ***D. mccartyi* doubling times (days)** | **Reference** |
|---|---|---|
| FL2 | 2.4 | He *et al.* EM, 2005 |
| DE195 | 0.8 | Maymo-Gatell *et al.* Science, 1997 |
| VS | 2.5 | Cupples *et al.* AEM 2003 |
| BAV1 | 2.2 | He *et al.* Science, 2003 |
| **AVERAGE doubling time** | **2** | **days** |

**Table S4-4**

| Divergence | method | branch lengths | # mutations | Time since divergence (years) from three different mutation rates | | |
|---|---|---|---|---|---|---|
| | | | | **bacteria** | ***E. coli*** | ***D. mccartyi* DONNA2/strain 195** |
| MRCA *Dhg* | concat tree (ML) | 0.2589 | 372651 | 1,395,876 | 258,496 | 124,229 |
| MRCA *Dhc* | concat tree (ML) | 0.2832 | 407627 | 1,526,891 | 282,758 | 135,889 |
| *Dehalococcoidia* | concat tree (ML) | 0.5936 | 854406 | 3,200,433 | 592,673 | 284,830 |
| *D. mccartyi* clades (Pinellas, Cornell, Victoria) | concat tree (ML) | 0.0582 | 81378 | 304,825 | 56,449 | 27,926 |
| KBTCE2 MRCA (w/KBTCE3) | concat tree (ML) | 1.80E-05 | 25 | 94 | 17 | 9 |
| KBTCE3 MRCA (w/KBTCE2) | concat tree (ML) | 1.14E-04 | 159 | 597 | 111 | 55 |
| KBTCE1 MRCA (w/KBVC2) | concat tree (ML) | 8.00E-06 | 11 | 42 | 8 | 4 |
| KBVC2 MRCA (w/KBTCE1) | concat tree (ML) | 3.30E-05 | 46 | 173 | 32 | 16 |
| GT MRCA (w/KBVC1) | concat tree (ML) | 5.75E-04 | 804 | 3,012 | 558 | 276 |
| KBVC1 MRCA (w/GT) | concat tree (ML) | 9.69E-04 | 1355 | 5,075 | 940 | 465 |
| KBDCA1 MRCA (w/KBDCA2) | concat tree (ML) | 8.00E-06 | 11 | 42 | 8 | 4 |
| KBDCA2 MRCA (w/KBDCA1) | concat tree (ML) | 4.10E-05 | 57 | 215 | 40 | 20 |

**Table S5. Reductive dehalogenases whose function has been inferred or biochemically characterized to any extent with ortholog group listed. Listed by primary substrate studied. S – single sequence, no other members currently in ortholog group.**

| Organism | Characterized RDase | main reaction catalyzed | Ortholog Group Number | Ref |
|---|---|---|---|---|
| *Dehalobacter restrictus* PERK23 | PceA | tetrachloroethene to *cis*-dichloroethene | 96 | [27] |
| *Dehalobacter* sp. TeCB1 | TcbA | 1,2,4,5-tetrachlorobenzene to 1,2,4-TCB and 1,2,4-TCB to 1,3- and 1,4-dichlorobenzene | 96 | [28] |
| *Dehalobacter* sp. TCB3 | TcbA | 1,2,4-TCB to 1,3- and 1,4-dichlorobenzene | 96 | [29] |
| *Dehalobacter* sp. CF | CfrA | chloroform to dichloromethane and 1,1,1-trichloromethane to dichloroethane | 97 | [30] |
| *Dehalobacter* sp. DCA | DcrA | 1,1-dichloroethane to monochloroethane | 97 | [30] |
| *Dehalobacter* sp. UNSWDHB | TmrA | chloroform to dichloromethane, 1,1,1-trichloroethane to 1,1-dichloroethane, and bromoform to dibromomethane | 97 | [31] |
| *Dehalobacter* sp. THM1 | ThmA | chloroform to dichloromethane | 97 | [32] |
| *Desulfitobacterium* sp. PR | CtrA | 1,1,2-trichloroethane to 1,2-dichloroethane | 97 | [33] |
| *Desulfitobacterium hafinense* PCP-1 | PentaCPh-CprA3 | Pentachlorophenol | 4 | [34] |
| *Desulfomonile tiedjei* | 3-CBA RDase | 3-chlorobenzoate to benzoate | S-46 | [35] |
| *Dehalobacter* sp. HCH1 | HchA | gamma-hexachlorocyclohexane to monochlorobenzene | 145 | [29] |
| *Dehalobacter* sp. TCP1 | DebcprA | 2,4,6-trichlorophenol to 4-chlorophenol | 91 | [36] |
| *Dehalogenimonas* sp. WBC-2 | TdrA | 1,2-*trans*-dichloroethene to vinyl chloride | S-263 | [37] |
| *Dehalogenimonas lykanthroporepellens* | DcpA | 1,2-dichloropropane to propene | 179 | [38] |
| *Dehalogenimonas* sp. GP | CerA | vinyl chloride to ethene | S-280 | [39] |
| *Dehalococcoides mccartyi* 195 | TceA | trichloroethene to *cis*-dichloroethene | 5 | [40] |
| *Dehalococcoides mccartyi* KB-1 | TceA | trichloroethene to *cis*-dichloroethene | 5 | [41] |
| *Dehalococcoides mccartyi* 195 | PceA | perchloroethene to trichloroethene | 30 | [40] |
| *Dehalococcoides mccartyi* CBDB1 | PceA | 2,3-dichlorophenol to monochlorophenol | 30 | [13] |
| *Dehalococcoides mccartyi* VS | VcrA | vinyl chloride to ethene | 8 | [42] |
| *Dehalococcoides mccartyi* KB-1 | VcrA | vinyl chloride to ethene | 8 | [41] |
| *Dehalococcoides mccartyi* WBC-2 | VcrA | vinyl chloride to ethene | 8 | [37, 43] |
| *Dehalococcoides mccartyi* BAV1 | BvcA | *cis*-dichloroethene to vinyl chloride to ethene | 28 | [44] |
| *Dehalococcoides mccartyi* KB-1 | BvcA | *cis*-dichloroethene to vinyl chloride to ethene | 28 | [41] |
| *Dehalococcoides mccartyi* CBDB1 | CbrA | 1,2,3,4-tetrachlorobenzene to 1,2,4-TCB to 1,4-DCB and 1,2,3-TCB to 1,3-DCB | 53 | [45] |
| *Dehalococcoides mccartyi* MB | MbrA | perchloroethene to *trans*-dichloroethene | 52 | [46] |
| *Dehalococcoides mccartyi* 11a | PteA | trichloroethene to *cis*-dichloroethene | 16 | [47] |
| *Dehalococcoides mccartyi* CG1, CG4 and CG5 | PcbA | dechlorination of PCB Aroclor 1260 | 35 | [48] |
| *Desulfitobacterium* sp. KBC1 | CprA | 3-chloro-4-hydroxyphenylacetate to 4-hydroxyphenylacetate | 2 | [49] |

| | | | | |
|---|---|---|---|---|
| *Desulfitobacterium* sp. PCE-1 | CprA | 3-chloro-4-hydroxyphenylacetate to 4-hydroxyphenylacetate and 2,4-dichlorophenol and 2,4,6-trichlorophenol | 2 | 50 |
| *Desulfitobacterium* sp. PCE-1 | PceA | tetrachloroethene to trichloroethene | 2 | 50 |
| *Desulfitobacterium* sp. Viet-1 | CprA | 2,4-dichlorophenol to 4-chlorophenol | 2 | 51 |
| *Desulfitobacterium hafniense* DCB-2 | CprA | 3-chloro-4-hydroxyphenylacetate to 4-hydroxyphenylacetate | 2 | 52-54 |
| *Sulfurospirillum multivorans* DSM 12446 | PceA | tetrachloroethene to trichloroethene to *cis*-dichloroethene. 1,2-dibromoethene and tribromoethene to vinyl bromide. Also 2,3-dichloropropene and 2,3-dibromopropene | 1 | 55-57 |
| *Sulfurospirillum* mixed culture SL2 | PceA-TCE | tetrachloroethene to trichlorothene | 1 | 58 |
| *Sulfurospirillum* mixed culture SL2 | PceA-DCE | tetrachlorothene to trichloroethene to *cis*-dichlorothene | 1 | 58 |
| *Desulfitobacterium hafniense* TCE1 | PceA | tetrachloroethene to *cis*-dichloroethene | 6 | 59 |
| *Desulfitobacterium dichloroeliminans* LMG P-21439 | DcaA | 1,2-dichloroethane to ethene | 6 | 60 |
| *Desulfitobacterium* sp. PCE-S | PceA | tetrachloroethene to trichloroethene | 6 | 56, 61 |
| *Desulfitobacterium hafniense* Y51 | PceA | tetrachloroethene to trichloroethene to *cis*-dichloroethene | 6 | 62 |
| *Desulfitobacterium hafniense* strain PCP-1 | CprA | 3-chloro-4-hydroxyphenylacetate to 4-hydroxyphenylacetate | 9 | 34 |
| *Desulfitobacterium sp.* KBC1 | PdrA | tetrachloroethene to trichloroethene | S-152 | 49 |
| *Desulfitobacterium chlororespirans* Co23 | CprA | 3-chloro-4-hydroxyphenylacetate to 4-hydroxyphenylacetate. di- tri- and tetra-chlorophenols. | S-153 | 63 |
| *Comamonas* sp. 7D-2 | BhbA | 3,5-dibromo-4-hydroxybenzonitrile | S-154 | 64 |
| *Desulfitobacterium hafniense* PCP-1 | CrdA | 2,4,6-trichlorophenol to 2,4-dichlorophenol. tetra,tri and di-chlorophenols | S-155 | 65 |
| *Nitratireductor pacificus* pht-3B | NpRdhA | 3,5-dibromo-4-hydroxybenzoic acid | S-156 | 66 |
| *Shewanella sediminis* HAW-EB3 | PceA | tetrachloroethene to trichlorothene | S-157 | 67 |

**Table S6. List of genomes (closed and contigs) analyzed in this paper. CG-complete genome.**

| Organism Name | Accession (NCBI or IMG) | Type of Assembly |
|---|---|---|
| *Dehalococcoides mccartyi* 195 | NC_002936.3/CP000027.1 | NCBI – CG |
| *Dehalococcoides mccartyi* CG5 | NZ_CP006951.1/CP006951.1 | NCBI – CG |
| *Dehalococcoides mccartyi* CG3 | NZ_CP013074.1/CP013074.1 | NCBI – CG |
| *Dehalococcoides mccartyi* CG1 | NZ_CP006949.1/CP006949.1 | NCBI – CG |
| *Dehalococcoides mccartyi* 11a5 | NZ_CP011127.1/CP011127.1 | NCBI – CG |
| *Dehalococcoides mccartyi* BTF08 | NC_020387.1/CP004080.1 | NCBI – CG |
| *Dehalococcoides mccartyi* IBARAKI | NZ_AP014563.1/AP014563.1 | NCBI – CG |
| *Dehalococcoides mccartyi* DCMB5 | NC_020386.1/CP004079.1 | NCBI – CG |
| *Dehalococcoides mccartyi* KBDCA1 | NZ_CP019867.1/CP019867.1 | NCBI – CG |
| *Dehalococcoides mccartyi* FL2 | NZ_CP038470.1/CP038470.1 | NCBI – CG |
| *Dehalococcoides mccartyi* VS | NC_013552.1/CP001827.1 | NCBI – CG |
| *Dehalococcoides mccartyi* GY50 | NC_022964.1/CP006730.1 | NCBI – CG |
| *Dehalococcoides mccartyi* CBDB1 | NC_007356.1/AJ965256.1 | NCBI – CG |
| *Dehalococcoides mccartyi* KBDCA2 | NZ_CP019868.1/CP019868.1 | NCBI – CG |
| *Dehalococcoides mccartyi* KBTCE1 | NZ_CP019999.1/CP019999.1 | NCBI – CG |
| *Dehalococcoides mccartyi* UCH-ATV1 | NZ_AP017649.1/AP017649.1 | NCBI – CG |
| *Dehalococcoides mccartyi* CG4 | NZ_CP006950.1/CP006950.1 | NCBI – CG |
| *Dehalococcoides mccartyi* GT | NC_013890.1/CP001924.1 | NCBI – CG |
| *Dehalococcoides mccartyi* KBVC1 | NZ_CP019968.1/CP019968.1 | NCBI – CG |
| *Dehalococcoides mccartyi* BAV1 | NC_009455.1/CP000688.1 | NCBI – CG |
| *Dehalococcoides mccartyi* KBVC2 | NZ_CP019969.1/CP019969.1 | NCBI – CG |
| *Dehalococcoides mccartyi* KBDCA3 | NZ_CP019946.1/CP019946.1 | NCBI – CG |
| *Dehalococcoides mccartyi* KBTCE2 | NZ_CP019865.1/CP019865.1 | NCBI – CG |
| *Dehalogenimonas lykanthroporepellens* BL-DC-9 | NC_014314.1/CP002084.1 | NCBI – CG |
| *Dehalogenimonas alkenigignens* IP3-3 | LFDV01 | NCBI – contigs |
| *Dehalogenimonas alkenigignens* BRE15M | QEFQ01 | NCBI – contigs |
| *Dehalogenimonas formicexedens* NSZ-14 | NZ_CP018285.1/ CP018285.1 | NCBI – CG |
| *Dehalogenimonas* sp. WBC-2 | CP011392.1 | NCBI – CG |
| *Dehalogenimonas* sp. GP | JQAN02 | NCBI – contigs |
| *Dehalobacter* sp. CF | NC_018867.1/CP003870.1 | NCBI – CG |
| *Dehalobacter* sp. DCA | NC_018866.1/CP003869.1 | NCBI – CG |
| *Dehalobacter* sp. FTH1 | AQYY01 | NCBI – contigs |
| *Dehalobacter* sp. 14DCB1 | PNXX01 | NCBI – contigs |
| *Dehalobacter* sp. 12DCB1 | POSF01 | NCBI – contigs |
| *Dehalobacter* sp. MCB1 | LXPC01 | NCBI – contigs |
| *Dehalobacter* sp. TeCB1 | MCHF01 | NCBI – contigs |
| *Dehalobacter* sp. E1 | CANE01 | NCBI – contigs |
| *Dehalobacter* sp. UNSWDHB | AUUR01 | NCBI – contigs |
| *Dehalobacter restrictus* DSM 9455 | NZ_CP007033.1/CP007033.1 | NCBI – CG |
| *Dehalobacter restrictus* 12DCA | CP046996.1 | NCBI – CG |
| *Dehalobacter* sp. TCB1 (referred to as KB1_124TCB1 in ref 29) | 2775506732 | IMG/JGI - contigs |
| *Dehalobacter* sp. TCB2 (referred to as KB1_124TCB2 in ref 29) | 2786546105 | IMG/JGI - contigs |
| *Dehalobacter* sp. TCB3 (referred to as KB1_124TCB3 in ref 29) | 2821530269 | IMG/JGI - contigs |
| *Dehalobacter* sp. HCH1 (referred to as GT1 in ref 29) | 2823892816 | IMG/JGI - contigs |

# REFERENCES


1. S. Tang, Y. Gong and E. A. Edwards, Semi-automatic in silico gap closure enabled de novo assembly of two *Dehalobacter* genomes from metagenomic data, *PLoS One*, 2012, **7**.
2. A. M. Bolger, M. Lohse and B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*, 2014, **30**, 2114-2120.
3. J. T. Simpson, K. Wong, S. D. Jackman, J. E. Schein, S. J. Jones and I. Birol, ABySS: a parallel assembler for short read sequence data, *Genome Res.*, 2009, **19**, 1117-1123.
4. M. Boetzer, C. V. Henkel, H. J. Jansen, D. Butler and W. Pirovano, Scaffolding pre-assembled contigs using SSPACE, *Bioinformatics*, 2011, **27**, 578-579.
5. S. Gnerre, I. Maccallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T. Sharpe, G. Hall, T. P. Shea, S. Sykes, A. M. Berlin, D. Aird, M. Costello, R. Daza, L. Williams, R. Nicol, A. Gnirke, C. Nusbaum, E. S. Lander and D. B. Jaffe, High-quality draft assemblies of mammalian genomes from massively parallel sequence data, *Proc Natl Acad Sci U S A*, 2011, **108**, 1513-1518.
6. C. T. Brown, A. Howe, Q. Zhang, A. B. Pyrokosz and T. H. Brom, A reference-free algorithm for computational normalization of shotgun sequencing data, *Quantitative Biology*, 2012.
7. M. Kearse, R. Moir, A. Wilson, S. Stones-Havas,M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes and A. Drummond, Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, *Bioinformatics*, 2012, **28**, 1647-1649.
8. R. Aziz, D. Bartels, A. D. Best, M, T. Disz, R. Edwards, K. Forsma, S. G. Gerdes, EM, M. Kubal, F. Meyer, G. Olsen, R. Olson, A. Ostermn, R. Overbeek, L. McNeil, D. Paarmann, T. Paczian, B. Parello, G. Push, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke and O. Zagnitko, The RAST server: rapid annotations using subsystems technology, *BMC Genomics*, 2008, **9**.
9. G. H. Van Domselaar, P. Stothard, S. Shrivastava, J. A. Cruz, A. Guo, X. Dong, P. Lu, D. Szafron, R. Greiner and D. S. Wishart, BASys: a web server for automated bacterial genome annotation, *Nucleic Acids Res.*, 2005, **33**, W455-W459.
10. A. Frank and J. Lobry, Oriloc: prediction of replication boundaries in unannotated bacterial chromosomes, *Bioinformatics*, 2000, **16**, 566-567.
11. A. Grostern and E. A. Edwards, Characterization of a *Dehalobacter* coculture that dechlorinates 1,2-dichloroethane to ethene and identification of the putative reductive dehalogenase gene, *Appl. Environ. Microbiol.*, 2009, **75**, 2684-2693.
12. A. S. Waller, R. Krajmalnik-Brown, F. E. Löffler and E. A. Edwards, Multiple reductive-dehalogenase-homologous genes are simultaneously transcribed during dechlorination by *Dehalococcoides*-containing cultures, *Appl. Environ. Microbiol.*, 2005, **71**, 8257-8264.
13. J. M. Fung, R. M. Morris, L. Adrian and S. H. Zinder, Expression of reductive dehalogenase genes in *Dehalococcoides ethenogenes* strain 195 growing on tetrachloroethene, trichloroethene, or 2,3-dichlorophenol, *Appl. Environ. Microbiol.*, 2007, **73**, 4439-4445.
14. K. M. Ritalahti, B. K. Amos, Y. Sung, Q. Wu, S. S. Koenigsberg and F. E. Löffler, Quantitative PCR targeting 16S rRNA and multiple reductive dehalogenase genes simultaneously monitors multiple *Dehalococcoides* strains, *Appl. Environ. Microbiol.*, 2006, **72**, 2765–2774.
15. A. Engelbrektson, V. Kunin, K. C. Wrighton, N. Zvenigorodsky, F. Chen, H. Ochman and P. Hugenholtz, Experimental factors affecting PCR-based estimates of microbial species richness and evenness, *ISME J*, 2010, **4**, 642-647.
16. E. Ramos-Padron, S. Bordenave, S. Lin, I. M. Bhaskar, X. Dong, C. W. Sensen, J. Fournier, G. Voordouw and L. M. Gieg, Carbon and sulfur cycling by microbial communities in a gypsum-treated oil sands tailings pond, *Environ Sci Technol*, 2011, **45**, 439-446.
17. J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R.

Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld and R. Knight, QIIME allows analysis of high-throughput community sequencing data, *Nat. Methods*, 2010, **7**.
18. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST, *Bioinformatics*, 2010, **26**, 2460-2461.
19. T. Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu and G. L. Andersen, Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB, *Appl. Environ. Microbiol.*, 2006, **72**, 5069-5072.
20. Q. Wang, G. M. Garrity, J. M. Tiedje and J. R. Cole, Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy, *Appl. Environ. Microbiol.*, 2007, **73**, 5261-5267.
21. H. Ochman, S. Elwyn and N. A. Moran, Calibrating bacterial evolution, *Proc Natl Acad Sci U S A*, 1999, **96**, 12638-12643.
22. X. Maymo-Gatell, Y. Chien, J. M. Gossett and S. H. Zinder, Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene, *Science*, 1997, **276**, 1568-1571.
23. P. J. McMurdie, L. A. Hug, E. A. Edwards, S. Holmes, A. A. Spormann and Site-specific mobilization of vinyl-chloride respiration islands by a mechanism common in *Dehalococcoides*, *BMC Genomics*, 2011, **12**, 287-302.
24. J. He, Y. Sung, R. Krajmalnik-Brown, K. M. Ritalahti and F. E. Löffler, Isolation and characterization of *Dehalococcoides* sp. strain FL2, a trichloroethene (TCE)- and 1,2-dichloroethene-respiring anaerobe, *Environ. Microbiol.*, 2005, **7**, 1442-1450.
25. A. M. Cupples, A. M. Spormann and P. L. McCarty, Growth of a Dehalococcoides-like microorganism on vinyl chloride and cis-dichloroethene as electron acceptors as determined by competitive PCR, *Appl. Environ. Microbiol.*, 2003, **69**, 953-959.
26. J. He, K. M. Ritalahti, K. L. Yang, S. S. Koenigsberg and F. E. Löffler, Detoxification of vinyl chloride to ethene coupled to growth of an anaerobic bacterium, *Nature*, 2003, **424**, 62-65.
27. J. Maillard, W. Schumacher, F. Vazquez, C. Regeard, W. R. Hagen and C. Holliger, Characterization of the corrinoid iron-sulfur protein tetrachloroethene reductive dehalogenase of *Dehalobacter restrictus*, *Appl. Environ. Microbiol.*, 2003, **69**, 4628-4638.
28. R. Alfán-Guzmán, H. Ertan, M. Manefield and M. Lee, Isolation and characterization of *Dehalobacter* sp. strain TeCB1 Including identification of TcbA: A novel tetra- and trichlorobenzene reductive dehalogenase, *Front Microbiol*, 2017, **8**, 558-558.
29. L. A. P. Jacome, Ph.D. Thesis, University of Toronto, 2019.
30. S. Tang and E. A. Edwards, Identification of *Dehalobacter* dehalogenases that catalyse dechlorination of chloroform, 1,1,1-trichloroethane and 1,1-dichloroethane, *Philos. Trans. R. Soc. Biol. Sci.*, 2013, **368**.
31. B.-E. Jugder, S. Bohl, H. Lebhar, R. D. Healey, M. Manefield, C. P. Marquis and M. Lee, A bacterial chloroform reductive dehalogenase: purification and biochemical characterization, *Microb Biotechnol*, 2017, **10**, 1640-1648.
32. S. Zhao, M. J. Rogers and J. He, Microbial reductive dehalogenation of trihalomethanes by a *Dehalobacter*-containing co-culture, *Appl. Microbiol. Biotechnol.*, 2017, **101**, 5481-5492.
33. S. Zhao, C. Ding and J. He, Detoxification of 1,1,2-trichloroethane to ethene by *Desulfitobacterium* and identification of its functional reductase gene, *PLoS One*, 2015, **10**, e0119507.
34. A. Bisaillon, R. Beaudet, F. Lépine, E. Déziel and R. Villemur, Identification and characterization of a novel CprA reductive dehalogenase specific to highly chlorinated phenols from *Desulfitobacterium hafniense* strain PCP-1, *Appl. Environ. Microbiol.*, 2010, **76**, 7536-7540.
35. S. Ni, J. K. Fredrickson and L. Xun, Purification and characterization of a novel 3-chlorobenzoate-reductive dehalogenase from the cytoplasmic membrane of *Desulfomonile tiedjei* DCB-1, *J. Bacteriol.*, 1995, **177**, 5135-5139.

36. S. Wang, W. Zhang, K.-L. Yang and J. He, Isolation and characterization of a novel *Dehalobacter* species strain TCP1 that reductively dechlorinates 2,4,6-trichlorophenol, *Biodegradation*, 2014, **25**, 313-323.
37. O. Molenda, A. T. Quaile and E. A. Edwards, *Dehalogenimonas* sp. strain WBC-2 genome and identification of its *trans*-dichloroethene reductive dehalogenase, TdrA, *Appl. Environ. Microbiol.*, 2016, **82**, 40-50.
38. E. Padilla-Crespo, J. Yan, C. Swift, D. D. Wagner, K. Chourey, R. L. Hettich, K. M. Ritalahti and F. E. Löffler, Identification and environmental distribution of *dcpA*, which encodes the reductive dehalogenase catalyzing the dichloroelimination of 1,2-dichloropropane to propene in organo-halide-respiring chloroflexi, *Appl. Environ. Microbiol.*, 2014, **80**, 808-818.
39. Y. Yang, S. A. Higgins, J. Yan, B. Simsir, K. Chourey, R. Iyer, R. L. Hettich, B. Baldwin, D. M. Ogles and F. E. Löffler, Grape pomace compost harbors organohalide-respiring *Dehalogenimonas* species with novel reductive dehalogenase genes, *ISME J*, 2017, **11**, 2767-2780.
40. J. K. Magnuson, M. F. Romine, D. R. Burris and M. T. Kingsley, Trichloroethene reductive dehalogenase from *Dehalococcoides ethenogenes:* sequence of *tceA* and substrate range characterization, *Appl. Environ. Microbiol.*, 2000, **66**, 5141-5147.
41. X. Liang, O. Molenda, S. Tang and E. A. Edwards, Identity and substrate-specificity of reductive dehalogenases expressed in *Dehalococcoides*-containing enrichment cultures maintained on different chlorinated ethenes, *Appl. Environ. Microbiol.*, 2015, **81**, 4626-4633.
42. J. A. Müller, B. M. Rosner, G. Von Abendroth, G. Meshulam-Simon, P. L. McCarty and A. M. Spormann, Molecular identification of the catabolic vinyl chloride reductase from *Dehalococcodies* sp. strain VS and its environmental distribution., *Appl. Environ. Microbiol.*, 2004, **70**, 4880-4888.
43. O. Molenda, S. Tang and E. A. Edwards, Complete genome sequence of *Dehalococcoides mccartyi* strain WBC-2, capable of anaerobic reductive dechlorination of vinyl chloride, *Genome Announc.*, 2016, **4**, e01375-01316.
44. R. Krajmalnik-Brown, T. Holscher, I. N. Thomson, F. M. Saunders, K. M. Ritalahti and F. E. Löffler, Genetic identification of a putative vinyl chloride reductase in *Dehalococcoides* sp. strain BAV1, *Appl. Environ. Microbiol.*, 2004, **70**, 6347-6351.
45. L. Adrian, J. Rahnenführer, J. Gobom and T. Hölscher, Identification of a chlorobenzene reductive dehalogenase in *Dehalococcoides* sp. strain CBDB1, *Appl. Environ. Microbiol.*, 2007, **73**, 7717-7724.
46. D. Cheng and J. He, Isolation and characterization of "*Dehalococcoides"* sp. strain MB which dechlorinates tetrachloroethene to *trans*-1,2-dichloroethene, *Appl. Environ. Microbiol.*, 2009, **75**, 5910-5918.
47. S. Zhao, C. Ding and J. He, Genomic characterization of *Dehalococcoides mccartyi* strain 11a5 reveals a circular extrachromosomal genetic element and a new tetrachloroethene reductive dehalogenase gene, *FEMS Microbiol. Ecol.*, 2017, **93**, fiw235.
48. S. Wang, K. R. Chng, A. Wilm, S. Zhao, K.-L. Yang, N. Nagarajan and J. He, Genomic characterization of three unique *Dehalococcoides* that respire on persistent polychlorinated biphenyls, *PNAS*, 2014, **111**, 12103-12108.
49. N. Tsukagoshi, S. Ezaki, T. Uenaka, N. Suzuki and R. Kurane, Isolation and transcriptional analysis of novel tetrachloroethene reductive dehalogenase gene from *Desulfitobacterium* sp. strain KBC1, *Appl. Microbiol. Biotechnol.*, 2006, **69**, 543-553.
50. J. Gerritse, V. Renard, T. M. Pedro Gomes, P. A. Lawson, M. D. Collins and J. C. Gottschal, *Desulfitobacterium* sp. strain PCE1, an anaerobic bacterium that can grow by reductive dechlorination of tetrachloroethene or ortho-chlorinated phenols, *Arch. Microbiol.*, 1996, **165**, 132-140.

51. J. M. Tront, B. K. Amos, F. E. Löffler and F. M. Saunders, Activity of *Desulfitobacterium* sp. strain Viet1 demonstrates bioavailability of 2,4-dichlorophenol previously sequestered by the aquatic plant Lemna minor, *Environ. Sci. Technol.*, 2006, **40**, 529-535.
52. S. H. Kim, C. Harzman, J. K. Davis, R. Hutcheson, J. B. Broderick, T. L. Marsh and J. M. Tiedje, Genome sequence of *Desulfitobacterium hafniense* DCB-2, a Gram-positive anaerobe capable of dehalogenation and metal reduction, *BMC Microbiol.*, 2012, **12**, 21.
53. N. Christiansen, B. K. Ahring, G. Wohlfarth and G. Diekert, Purification and characterization of the 3-chloro-4-hydroxy-phenylacetate reductive dehalogenase of *Desulfitobacterium hafniense*, *FEBS Lett.*, 1998, **436**, 159-162.
54. A. MacNelly, Kai, M., Svatos, M., Diekert, G.,& Schubert, T., Functional heterologous production of reductive dehalogenases from *Desulfitobacterium hafniense* strains, *App. Envron. Microbiol.*, 2014, 80(14), 4313-4322.
55. A. Neumann, G. Wohlfarth and G. Diekert, Purification and characterization of tetrachloroethene reductive dehalogenase from *Dehalospirillum multivorans*, *J. Biol. Chem.*, 1996, **271**, 16515-16519.
56. L. Ye, A. Schilhabel, S. Bartram, W. Boland and G. Diekert, Reductive dehalogenation of brominated ethenes by *Sulfurospirillum multivorans* and *Desulfitobacterium hafniense* PCE-S, *Environ. Microbiol.*, 2010, **12**, 501-509.
57. R. P. Schmitz, J. Wolf, A. Habel, A. Neumann, K. Ploss, A. Svatos, W. Boland and G. Diekert, Evidence for a radical mechanism of the dechlorination of chlorinated propenes mediated by the tetrachloroethene reductive dehalogenase of *Sulfurospirillum muftivorans*, *Environ. Sci. Technol.*, 2007, **41**, 7370-7375.
58. G. F. Buttet, C. Holliger and J. Maillard, Functional genotyping of *Sulfurospirillum* spp. in mixed cultures allowed the identification of a new tetrachloroethene reductive dehalogenase, *Appl. Environ. Microbiol.*, 2013, **79**, 6941-6947.
59. J. Maillard, C. Regeard and C. Holliger, Isolation and characterization of Tn-Dha1, a transposon containing the tetrachloroethene reductive dehalogenase of *Desulfitobacterium hafniense* strain TCE1, *Environ. Microbiol.*, 2005, **7**, 107-117.
60. M. Marzorati, F. de Ferra, H. Van Raemdonck, S. Borin, E. Allifranchini, G. Carpani, L. Serbolisca, W. Verstraete, N. Boon and D. Daffonchio, A novel reductive dehalogenase, identified in a contaminated groundwater enrichment culture and in *Desulfitobacterium dichloroeliminan*s strain DCA1, is linked to dehalogenation of 1,2-dichloroethane, *Appl. Environ. Microbiol.*, 2007, **73**, 2990-2999.
61. E. Miller, G. Wohlfarth and G. Diekert, Purification and characterization of the tetrachloroethene reductive dehalogenase of strain PCE-S, *Arch. Microbiol.*, 1998, **169**, 497-502.
62. A. Suyama, M. Yamashita, S. Yoshino and K. Furukawa, Molecular characterization of the PceA reductive dehalogenase of *Desulfitobacterium* sp. strain Y51, *J. Bacteriol.*, 2002, **184**, 3419-3425.
63. J. Krasotkina, T. Walters, K. A. Maruya and S. W. Ragsdale, Characterization of the B12- and iron-sulfur-containing reductive dehalogenase from *Desulfitobacterium chlororespirans*, *J. Biol. Chem.*, 2001, **276**, 40991-40997.
64. K. Chen, L. Huang, C. Xu, X. Liu, J. He, S. H. Zinder, S. Li and J. Jiang, Molecular characterization of the enzymes involved in degradation of a brominated aromatic herbicide, *Mol. Microbiol.*, 2013, **89**, 1121-1139.
65. A. Boyer, R. Pagé-BéLanger, M. Saucier, R. Villemur, F. Lépine, P. Juteau and R. Beaudet, Purification, cloning and sequencing of an enzyme mediating the reductive dechlorination of 2,4,6-trichlorophenol from *Desulfitobacterium frappieri* PCP-1, *Biochem. J.*, 2003, **373**, 297-303.
66. K. A. Payne, C. P. Quezada, K. Fisher, M. S. Dunstan, F. A. Collins, H. Sjuts, C. Levy, S. Hay, S. E. Rigby and D. Leys, Reductive dehalogenase structure suggests a mechanism for B12-dependent dehalogenation, *Nature*, 2015, **517**, 513-516.

67. S. T. Lohner and A. M. Spormann, Identification of a reductive tetrachloroethene dehalogenase in *Shewanella sediminis*, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 2013, **368**, 20120326.