

Contribution of mobile genetic elements to *Desulfovibrio vulgaris* genome plasticity

Christopher B. Walker,^{1,2†} Sergey Stolyar,^{1,2}
Dylan Chivian,^{2,3} Nicolas Pinel,¹ Jeffrey A. Gabster,^{1,2}
Paramvir S. Dehal,^{2,3} Zhili He,^{2,4} Zamin Koo Yang,^{2,6}
Huei-Che B. Yen,^{2,5} Jizhong Zhou,^{2,4} Judy D. Wall,^{2,5}
Terry C. Hazen,^{2,3} Adam P. Arkin^{2,3} and
David A. Stahl^{1,2*}

¹Department of Civil and Environmental Engineering,
University of Washington, Seattle, WA, USA.

²Virtual Institute for Microbial Stress and Survival
(vimss.lbl.gov).

³Physical Biosciences Division, Lawrence Berkeley
National Laboratory, Berkeley, CA, USA.

⁴Institute for Environmental Genomics, Oklahoma
University, Norman, OK, USA.

⁵Biochemistry Division, University of Missouri, Columbia,
MO, USA.

⁶Environmental Sciences Division, Oak Ridge National
Laboratory, Oak Ridge, TN, USA.

Summary

The genome of *Desulfovibrio vulgaris* strain DePue, a sulfate-reducing *Deltaproteobacterium* isolated from heavy metal-impacted lake sediment, was completely sequenced and compared with the type strain *D. vulgaris* Hildenborough. The two genomes share a high degree of relatedness and synteny, but harbour distinct prophage and signatures of past phage encounters. In addition to a highly variable phage contribution, the genome of strain DePue contains a cluster of open-reading frames not found in strain Hildenborough coding for the production and export of a capsule exopolysaccharide, possibly of relevance to heavy metal resistance. Comparative whole-genome microarray analysis on four additional *D. vulgaris* strains established greater interstrain variation within regions associated with phage insertion and exopolysaccharide biosynthesis.

Introduction

Bacteriophage and other mobile elements represent a large reservoir of genetic information affecting the structure and evolution of microbial ecosystems. Complete genome sequences for numerous environmentally relevant microbes have highlighted both a relatively conserved core genome and a remarkable contribution of bacteriophage to the pan-genome; the pan-genome composed of the sum of all unique genes distributed in a described species (Tettelin *et al.*, 2005; Coleman *et al.*, 2006; Cuadros-Orellana *et al.*, 2007; Lindell *et al.*, 2007). For example, while comparative genomic analysis of only three *Prochlorococcus* species revealed the core genome, sequencing of 12 isolates suggested a pan-genome at least an order of magnitude larger than any single genome (Lindell *et al.*, 2007). Although anaerobes are as yet less well represented by completed genome sequences, the comparative analyses of closely related *Desulfovibrio* isolates presented here indicates a similar contribution of viruses and other mobile genetic elements to the pan-genome of this representative anaerobe.

Desulfovibrio are sulfate-reducing microorganisms that participate in global sulfur- and carbon-cycling, often forming complex communities with other anaerobes and facilitating the complete decomposition of organic material (for review, see Rabus *et al.*, 2005). The significance of phage to population structure was initially suggested by the isolation and characterization of *D. vulgaris* strain DePue. This sulfate-reducing microorganism, isolated from a heavy metal-impacted lake sediment, lacks prophages present in strain Hildenborough (Walker *et al.*, 2006). The recently completed genome sequence of strain DePue reported here now provides the basis for a more complete census of phage-associated divergence among different described strains of *Desulfovibrio vulgaris* and highlights the role of CRISPR (clusters of regularly interspaced palindromic repeats) immunity among closely related strains.

Previous studies demonstrated that strain DePue was sensitive to lytic infection by viruses carried as prophages in strain Hildenborough (Walker *et al.*, 2006). Complementary whole-genome microarray comparisons with other *D. vulgaris* strains now shows that among this study set, strain DePue is unique in lacking any of the six prophage annotated in the Hildenborough genome. Thus,

Received 9 November, 2008; accepted 30 March, 2009. *For correspondence. E-mail dastahl@u.washington.edu; Tel. (+1) 206 685 3464; Fax (+1) 206 685 9185. †Present address: Geosyntec Consultants, 1370 Stewart Avenue, Suite 205, Seattle, WA 98104, USA.

DEPUE

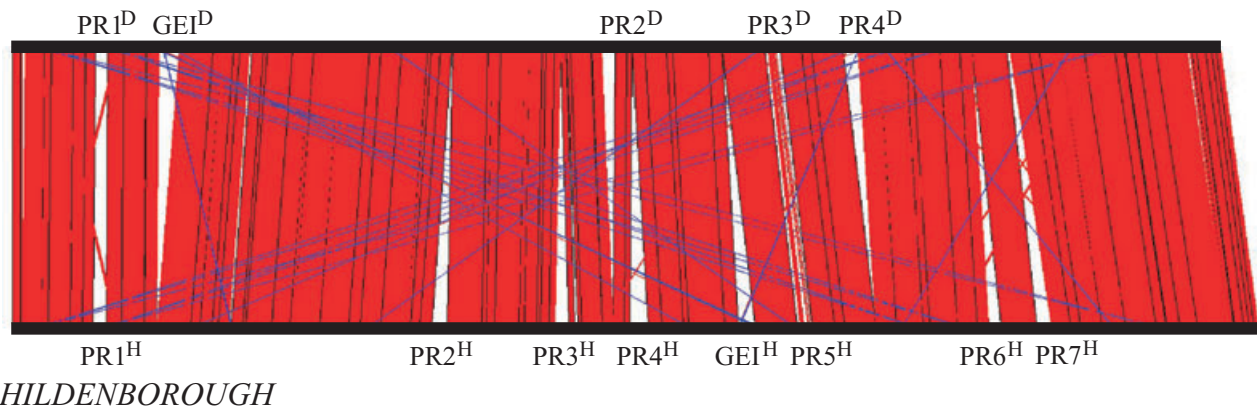


Fig. 1. Genomic comparison of strains DePue and Hildenborough. Regions sharing greater than 90% nucleotide identity between strains are shaded by color: red, direct matches; blue, reverse complement matches; white gaps, unique ORFs; PR#, phage regions for strains DePue and Hildenborough; GEI, genomic islands.

sensitivity to infection is correlated with both the absence of closely related prophage and the absence of CRISPR regions complementary to sequences of Hildenborough bacteriophage. Another distinguishing feature of the two genomes is the presence of two unique genomic islands of comparable size (*c.* 47 and 57 kb in strains Hildenborough and DePue respectively), both inserted near a tRNA gene in different regions in the chromosomes. These appear to encode functions that are niche-specific, providing resistance to chemical, oxidative and metal stress.

Results

General features

Strain DePue contains a single circular chromosome and one megaplasmid, both smaller than the strain Hildenborough chromosome and megaplasmid (Heidelberg *et al.*, 2004). Table 1 and Fig. S1 summarize the general features of the chromosome and megaplasmid. Each strain con-

tains five ribosomal RNAs, 68 tRNAs and there is no significant difference in GC content or coding density. Both share a high degree of synteny and gene similarity (Fig. 1 and Fig. S2), with the majority of open-reading frames (ORFs) exhibiting greater than 95% amino acid similarity (Fig. 2 and Table S1). Strain DePue contains 3191 anno-

Table 1. Genome characteristics of *D. vulgaris* strains DePue and Hildenborough.

Feature	DePue	Hildenborough
Chromosome size (nt)	3 462 887	3 570 858
Megaplasmid size (nt)	198 504	202 301
GC content (%)	63.01	63.14
Chromosomal ORFs	3 041	3 480
Megaplasmid ORFs	150	152
Coding density (%)	86.8	86.8
rRNA	5	5
tRNA	68	68
Misc. RNA/pseudogenes	17	20
ORF density (ORF/kb)	0.844	0.936
Average ORF length (kb)	1 028	928
Strain-specific ORFs	277	718

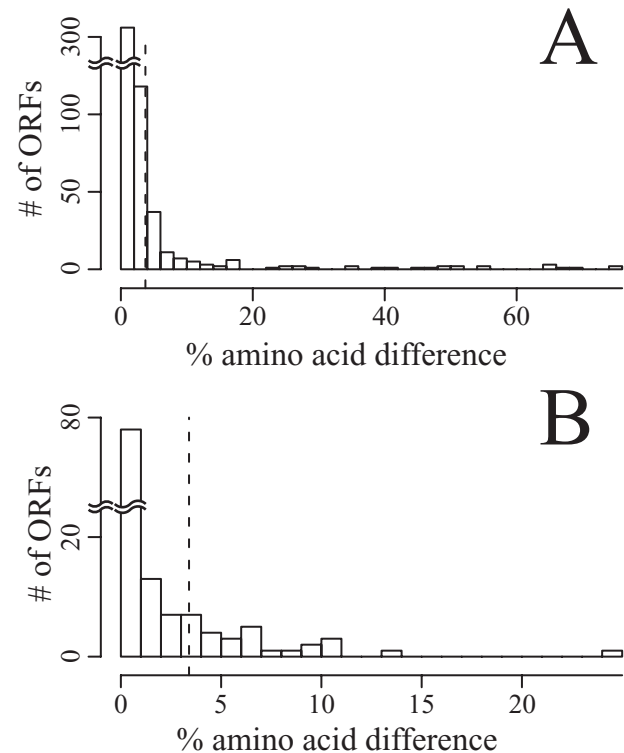


Fig. 2. Histogram of amino acid per cent identity of orthologues in strains Hildenborough and DePue for chromosomally (A) and plasmid (B)-encoded ORFs. The dotted line represents the 80% quantile.

tated protein-coding sequences, 441 fewer than strain Hildenborough. The reduced number of protein-coding sequences appears disproportionate relative to the accompanying reduction in the genome size of strain DePue. Apart from the use of different ORF-determining algorithms, the disproportionately higher number of predicted protein-coding sequences in strain Hildenborough primarily reflects of the larger number of small ORFs assigned in regions of presumptive phage insertion (strain Hildenborough containing seven insertions as opposed to four in strain DePue). Their possible biological significance will likely only be resolved through more extensive comparative analyses of related *Desulfovibrio*.

The two genomes share a total of 2914 genes, representing 91% and 80% of the strain DePue and Hildenborough genomes respectively, with 267 genes unique to strain DePue genes and 718 unique to strain Hildenborough. Assignment of all genes into clusters of orthologous functional groups revealed only minor differences between strains, with the primary divergence in gene content associated with genomic islands in strain DePue (Fig. S3). Small inversions and rearrangements also distinguish the two genomes, most involving tRNAs or mobile genetic elements. Notably, the genome of strain DePue lacks the majority of the insertion sequence elements present in strain Hildenborough.

Genomic islands

The 277 ORFs unique to strain DePue genome assemble in five large genomic islands and four smaller clusters (Fig. 1 and Table S2). Hypothetical, conserved hypothetical and bacteriophage-related ORFs comprise most of these unique protein-coding sequences. Four of the large genomic islands (listed in Fig. 1 as Phage Region 1^{DePue}, PR2^D, PR3^D and PR4^D) contain phage or phage-related ORFs, with all but PR1^D located in different chromosomal locations than the prophage of strain Hildenborough. The

fifth genomic island (GEI^D, Fig. 1) contains ORFs coding for exopolysaccharide (EPS) production, modification and transport. PR1^D, PR2^D and PR4^D contain ORFs exclusively found in strain DePue, while approximately half of the ORFs in PR3^D are unique (Fig. S4). The rest of the ORFs of this region demonstrate homology to protein-coding sequences found in strain Hildenborough at approximately the same genomic locus. A tRNA immediately flanks all five genomic islands, a general feature among mobile genetic elements and common to bacteriophage insertion sites (Williams, 2002; Campbell, 2003). The four smaller clusters unique to strain DePue contain ORFs coding for a variety of functions, with annotation suggesting contributions to outer-membrane protein transport/modification, phage resistance and DNA excision/repair.

Exopolysaccharide genomic island

Unlike the other unique genomic islands in strain DePue coding for bacteriophage-related functions, the approximately 57 kb region (GEI^D) located between PR1^D and PR2^D contains 33 ORFs clustered into two distinct functional groups (Fig. 3, Table S3). These ORFs share some homology with genes in previously sequenced genomes; however, no single organism contains more than a few highly similar protein-coding sequences. The first cluster, containing 15 ORFs (Dvul_2564–79) encodes sensor and response modules of unknown function. The cluster is flanked by genes for an integrase and transposase and may be capable of transposition and/or be part of a larger genomic island that includes the adjacent cluster of genes (Dvul_2580–94) encoding for EPS synthesis, export and modification.

The region containing this EPS biosynthetic gene shares features with a genomic island present in strain Hildenborough (GEI^H in Fig. 1) containing 52 annotated genes (DVU2000–DVU2051). The GEIs in strains DePue and Hildenborough are directly inserted in a tRNA-Met gene

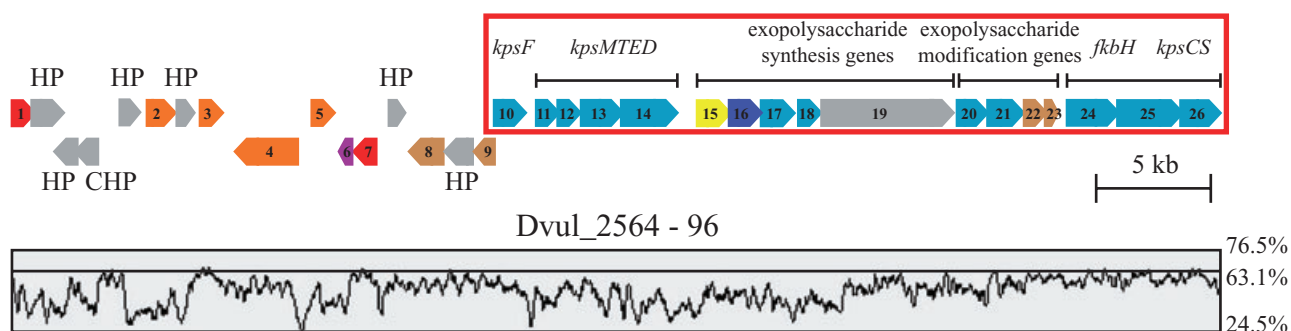


Fig. 3. Unique exopolysaccharide biosynthesis region in the genome of strain DePue. The plot illustrates the GC% of ORFs and intergenic regions. ORF numbering refers to descriptions in Table S3, while color-coding indicates ORFs of similar COGs. The red box highlights the second ORF cluster coding for capsule exopolysaccharide production, modification and export proteins. The line marking 63.1% corresponds to the average genome GC%.

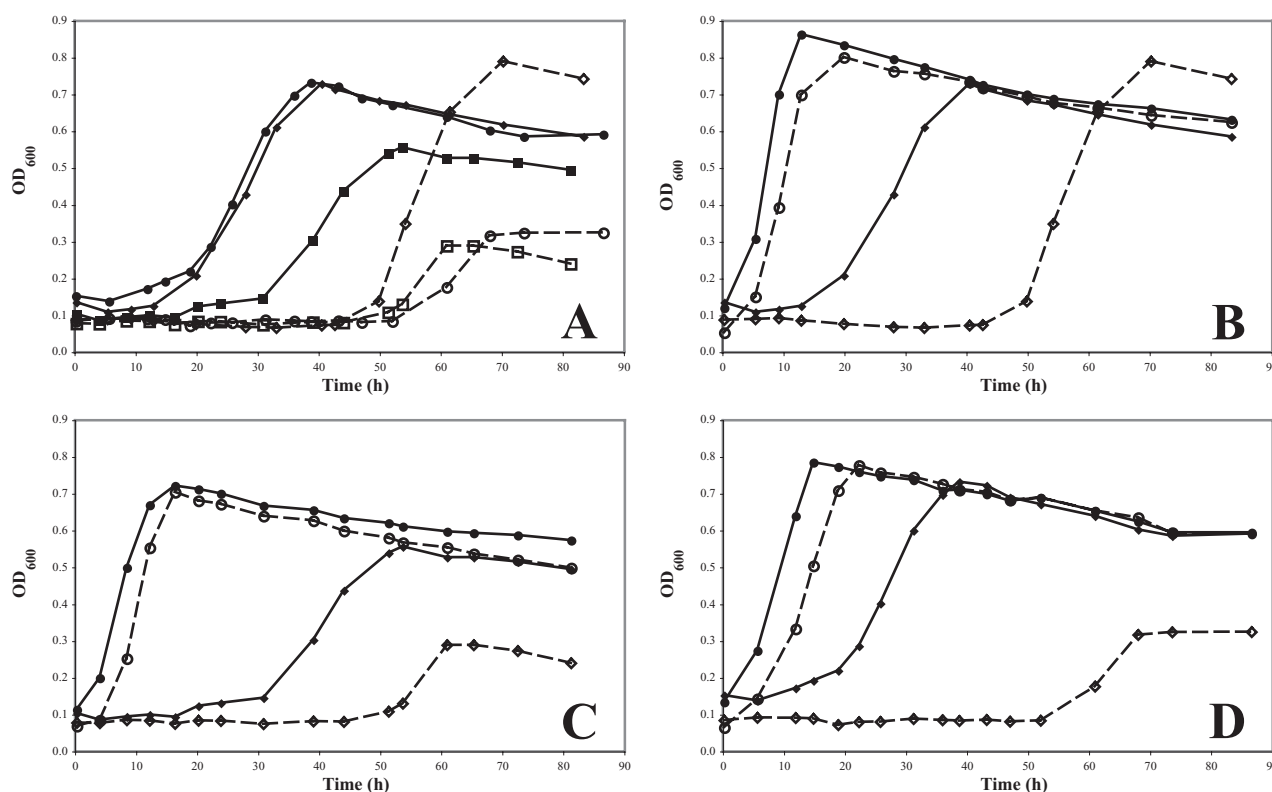


Fig. 4. Effect of 50 μM Cr(VI) on growth of *D. vulgaris* strains. Closed symbols represent 20% (v/v) inoculum size. Open symbols represent 10% inoculum size. No growth was observed for any strain with less than 10% inoculum. A. Growth of strain DePue (diamonds) and two variants of strain Hildenborough, one possessing a megaplasmid (squares) and one lacking a megaplasmid (circles). B. Growth of strain DePue in the absence (circles) and presence (diamonds) of chromate. C. Growth of strain Hildenborough possessing the megaplasmid in the absence (circles) and presence (diamonds) of chromate. D. Growth of strain Hildenborough lacking the megaplasmid in the absence (circles) and presence (diamonds) of chromate.

and flanked by identical 49 bp direct repeat sequences. Both GEIs have GC skews well below the genome average and contain genes for site-specific recombinases and transposases (Fig. 3) (Johnston *et al.*, 2008). This sequence-based suggestion of acquisition by lateral transfer is now supported by recent studies demonstrating that this region in strain Hildenborough is spontaneously excised in a small fraction (*c.* 3%) of cells maintained under standard laboratory growth conditions (Johnston *et al.*, 2008). Strains lacking this GEI demonstrated faster growth with lactate and sulfate under anoxic conditions than those carrying it (Johnston *et al.*, 2008). However, the presence of this GEI is strongly implicated in contributing to greater fitness under environmental conditions of chemical and oxidative stress (Johnston *et al.*, 2008).

The structurally similar GEI in strain DePue likely also confers selective advantage under certain environmental conditions. For example, key enzymes in the pathway for sialic acid biosynthesis (highlighted in red on Table S3) are among the multiple genes encoding for EPS biosynthesis in this GEI. Past studies have shown the capacity of *D. vulgaris* to reduce Cr(VI) to the less toxic Cr(III), and to

accumulate precipitates of the less toxic trivalent chromium on their cell surface, inner and outer membranes (Lloyd and Lovley, 2001; Chardin *et al.*, 2003; Goulhen *et al.*, 2006). Independent characterization of the role of organic ligands in stabilizing chromium in metal-polluted soils implicated the acidic sugars of bacterial EPS in both stabilizing and immobilizing trivalent chromium (Kantar *et al.*, 2008). Although sialic acid surface modification of pathogenic bacteria is better recognized for its role in avoidance of immune surveillance, these acidic groups could also serve to increase the tolerance of strain DePue to chromium and other heavy metals, such as those contaminating the sediments from which this strain was isolated (Walker *et al.*, 2006). A possible function of this capsule in heavy metal resistance is supported by the greater tolerance of strain DePue to chromate relative to strain Hildenborough (Fig. 4).

Megaplasmids

Both strains carry very similar megaplasmids (202 and 198.5 kb in strain Hildenborough and DePue respec-

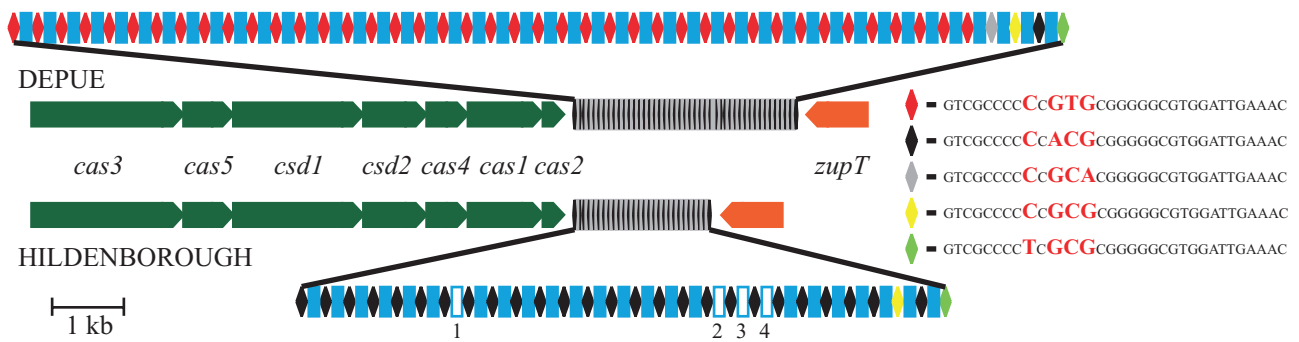


Fig. 5. CRISPR and *cas* gene region in megaplasms of strains DePue and Hildenborough. A zinc transporter (*zupT*) and CRISPR-associated (*cas*) genes flank the CRISPR region. Rectangles indicate spacer regions and diamonds indicate repeat structures. Color-coding represents identical repeat sequences. The variable repeat sequence nucleotides are highlighted in red. The numbered open rectangles represent spacer sequences with identical matches in the strain DePue genome: 1, intergenic region between Dvul_1048–9; 2, hypothetical protein (Dvul_1039); 3, NAD-dependent DNA ligase (Dvul_1045); 4, peptidase (Dvul_1049). See Table S4 for spacer sequences.

tively). As the plasmid is not required for growth of strain Hildenborough under laboratory conditions, being frequently lost during routine culture maintenance (J.D. Wall, unpublished), it is notable that strains isolated from different continents and decades apart carry near identical plasmids. Differences are relatively minor – the strain DePue megaplasmid lacks 21 ORFs present in strain Hildenborough and contains 19 unique genes (Dvul_3099–101, Fig. S5). With the exception of ORFs previously assigned roles in chromate transport/resistance in strain Hildenborough, most of those absent in DePue are of unknown function (Klonowska *et al.*, 2007). No mobile genetic elements or repetitive signatures flank these genes in strain Hildenborough.

The shared presence and conservation of the megaplasms in the two isolates are suggestive of strong selective pressure under normal environmental conditions. Both megaplasms code for nitrogen fixation (Dvul_3089–98), EPS biosynthesis (Dvul_3055–72), possible defence against grazers via secretion system III (Dvul_2984–3010), surface lipopolysaccharide modification proteins (Dvul_3028–39), a protein/peptide transport system for potential antibacterial compounds (Dvul_3083–85) and phage immunity systems involving an abortive infection protein and CRISPR defence mechanisms (Dvul_2973–78, Fig. S4 and Table S2). All could serve functions in the environment that would have less adaptive significance in pure culture.

The most striking difference between these closely related strains (sharing greater than 99.8% nucleotide identity of 16S rRNA genes) is carriage of distinct prophage and short sequence motifs associated with acquired phage immunity. The megaplasms of both strains contain unique CRISPR loci and highly similar CRISPR-associated (*cas*) genes (Fig. 5), generally recognized as providing bacteriophage immunity (Mojica *et al.*, 2005; Barrangou *et al.*, 2007; Brouns *et al.*, 2008). Both strains

contain a zinc transporter gene (*zupT*, 97% amino acid similarity) following the distal repeat sequence and a highly similar leader sequence (87% nucleotide identity) between the *cas2* gene (97% amino acid similarity) and the proximal repeat sequence. Despite the high amino acid similarities in flanking genes, each CRISPR region is unique in sequence and length (strain DePue contains 45 repeats versus 28 for strain Hildenborough, Table S4). Both strains contain nearly identical 32 bp repeat sequences, with strain-specific variations occurring at common variable regions (Kunin *et al.*, 2007). The same three unique repeat sequences occur in both strains at the distal end of the CRISPR region (Fig. 5, green, black and yellow diamonds), with strain DePue containing a fourth unique repeat sequence (Fig. 5, grey diamond). One of these distal repeat sequences (Fig. 5, black diamond) shares 100% nucleotide identity with the majority of the strain Hildenborough repeat sequences. Small intrastrain variations in CRISPR repeat sequences are common in genome sequences available for other microbes; however, the conserved variations in the distal repeat sequences appear unique for strains DePue and Hildenborough.

Intra- and interstrain comparisons indicate no redundant spacer sequences with lengths varying from 29 to 36 nucleotides (Table S4). Four strain Hildenborough spacer sequences (Fig. 5, open blue boxes) are identical to sequences within PR3^D of strain DePue. Our previous observation that two morphologically distinct bacteriophage carried by Hildenborough can infect strain DePue (Walker *et al.*, 2006) is consistent with the absence of spacer sequences in strain DePue having significant nucleotide matches in strain Hildenborough. A search of the Global Ocean Sampling metavirome and available viral genome sequences revealed highly similar, but non-identical matches to strain DePue and Hildenborough spacer sequences (Rusch *et al.*, 2007; Williamson *et al.*, 2008).

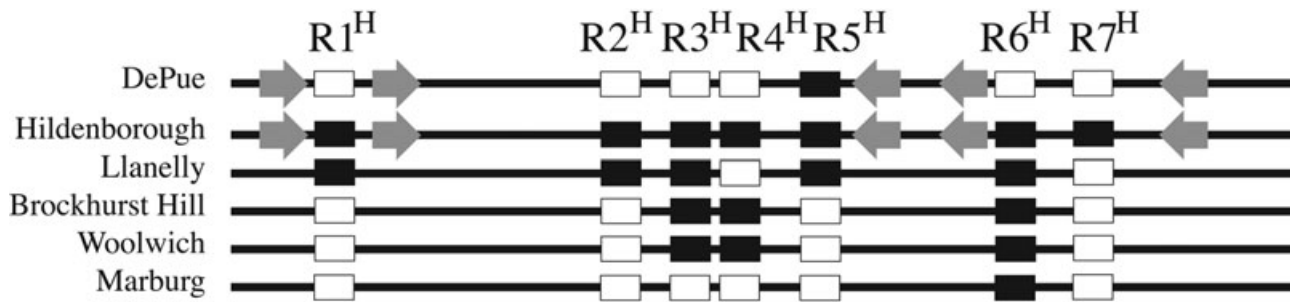


Fig. 6. Presence/absence distribution mapping of strain Hildenborough bacteriophage among *D. vulgaris* strains. The presence (black rectangles) and absence (white rectangles) of bacteriophage are mapped against the genomes of strain DePue and Hildenborough. The genomic location of bacteriophage present in strains Llanelly (NCIMB 8446), Brockhurst Hill (NCIMB 8306), Woolwich (NCIMB 8457) and Marburg (ATCC 35115) may differ. Grey arrows indicate ribosomal RNAs.

Hybridization array comparisons

Comparative analysis also revealed a correlation between this highly elaborated system of phage immunity and an unusually large variety of prophage harboured by the two strains of *D. vulgaris*. Together the two strains contain 11 unique phage-specific genomic islands (7 in strain Hildenborough and 4 in strain DePue, Fig. S4). This density (prophage/Mb of genome, Table S5) is among the highest in available genome sequences and suggests a high impact of viral infection on the genetic content of *D. vulgaris*. Based on this observation, we examined the distribution of Hildenborough bacteriophage among four other *D. vulgaris* strains (Llanelly, Brockhurst Hill, Woolwich and Marburg) by hybridizing genomic DNA from each strain to an array built from the sequence of strain Hildenborough. Each strain hybridized to at least 90% of the genes represented on strain Hildenborough array [genomic DNA (gDNA) sample strain/gDNA strain Hildenborough > 0.70], with the majority of missing ORFs (defined by absence of hybridization) annotated as phage or phage-related (Fig. 6 and Table S6). Although all strains have highly similar (> 99%) 16S rRNA gene nucleotide identity (Table S7), little correlation exists between relatedness and carriage of specific phage. All members of this study set, except strain DePue, hybridized to array elements targeting ORFs encoding PR6^H, a lambdoid-like phage. Conversely, none of the strains hybridized to probes targeting PR7^H, a mu-like phage. With the exception of strain Llanelly, all did not hybridize with probes for another mu-like phage located in PR1^H. The most divergent strains (Brockhurst Hill and Woolwich) contain the same set of Hildenborough viral elements, hybridizing to probes for PR3^H, PR4^H and PR6^H.

Physiological comparisons

As yet no marked physiological differences have been observed between the two strains. Previous characterization revealed little difference in growth temperature

optimum or substrate utilization (Walker *et al.*, 2006). Given the observation that genes associated with chromate resistance are missing from the strain DePue megaplasmid and in consideration that this strain was isolated from a heavy metal-impacted lake, the response to chromate stress was examined. Although both strains exhibit considerable chromate tolerance, strain DePue demonstrated greater resistance to chromate at low inoculum transfer than either the plasmid-plus or plasmid-minus variant of strain Hildenborough (Fig. 4). Thus, chromate resistance does not appear to be primarily determined by plasmid-encoded traits, including a gene annotated in chromate efflux, but may be associated with the EPS encoded by the strain-specific genomic island.

Discussion

The concept of a microbial species has yet to accommodate the remarkable diversity revealed by ongoing genome sequencing projects. Although adaptive features such as those exhibited in *Prochlorococcus* species likely exist in anaerobic populations, the reduced number of sequenced anaerobic organisms limits detailed genomic comparisons. The characterization described here of two closely related *D. vulgaris* strains provides initial comparative perspective on the physiological and genomic plasticity within a common species of sulfate reducers.

Both *D. vulgaris* strains share a high percentage of genes (80–91%); however, approximately 1000 are strain-specific. As is typical for other reported strain divergence, variation occurs primarily within 'islands' (frequently encoding surface-associated features, mobile genetic elements and prophage) flanked by more highly conserved regions. The greatest source of sequence divergence between these two strains derives from major differences in type and number prophage, and in two unique genomic islands. Complementary microarray studies further documented the presence of a diverse and active phage population carried among closely related

strains of *D. vulgaris*. Although variable prophage distribution and associated immunity systems are expected to play important roles in determining population's structure, the unique GEI elements can be more directly associated with niche specialization. Comparison of two variants of strain Hildenborough, containing or missing the GEI unique to this strain, suggested that this element could confer significant adaptive advantage in environments experiencing fluctuations in chemical/oxidative stress (Johnston *et al.*, 2008). Similarly, modification of surface features attributed to the GEI present in strain DePue would be expected to also confer selective advantage in certain environments – possibly contributing to greater chromate tolerance, resistance to protozoal grazing or phage infection by masking receptors. However, the influence of lysogeny on the fitness of *Desulfovibrio* and the cost of maintaining specific phage resistance mechanisms remain questions for future study.

Apart from a major contribution of prophage to the genomic structure of the two strains, the presence of very similar megaplasmids is also notable. As this plasmid does not appear to alter the growth of strain Hildenborough under laboratory conditions, it likely confers significant adaptive advantage in non-laboratory environments (Johnston *et al.*, 2008). As briefly presented in the results, plasmid-encoded functions that could contribute to strong environmental selection include systems for nitrogen fixation, and resistance to viral and protozoal attack. In addition to a recognized contribution of the plasmid to biofilm formation (Klonowska *et al.*, 2007), well known to suppress grazing, both megaplasmids encode a protein secretion system III usually associated with plant and animal pathogenesis (Hueck, 1998).

Experimental procedures

High-molecular-weight genomic DNA preparation

Strain DePue was grown as previously described in 500 ml of media in 1 l pyrex bottles fitted with black butyl rubber stoppers and screw-top caps (Walker *et al.*, 2006). Cells were harvested in mid-exponential phase into 500 ml polycarbonate wide-mount centrifuge bottle (Nalgene Labware, Rochester, NY) previously stored in an anaerobic chamber for 48 h. Cells were centrifuged for 30 min at 6000 r.p.m. (6084 *g*) and 4°C in a Sorvall RC-5B refrigerated centrifuge equipped with a GS-3 rotor. After centrifugation, supernatant was poured off and the pellets frozen at –80°C until processing. High-molecular-weight DNA was isolated as previously described using phenol : chloroform : isoamyl (Heidelberg *et al.*, 2004).

Genome sequencing

A completely sequenced genome of *D. vulgaris* strain DePue was obtained through collaboration with the Joint Genome Institute. Whole-genome shotgun sequencing of 3, 8 and 40 kb DNA libraries produced at least 8× coverage of the

entire genome. Specifics of clone library generation, sequencing and assembly strategies may be found at the DOE JGI website (<http://www.jgi.doe.gov/sequencing/index.html>).

Genome analysis

Analysis of the completely sequence *D. vulgaris* DePue genome was completed using MicrobesOnline (<http://www.microbesonline.org/>), the DOE JGI web-browsing software (<http://genome.ornl.gov/microbial/dvul/03oct06/>) and the TIGR Manatee program (<http://stahl.ce.washington.edu/genomes/>).

Direct comparison of the *D. vulgaris* DePue and Hildenborough genomes was performed using both MicrobesOnline and the Artemis Comparison Tool (Carver *et al.*, 2005) with a comparison library generated through WebACT (<http://www.webact.org/WebACT/home>). A list of orthologues shared by the two genomes was generated using a reciprocal best-hit approach.

Comparisons between the *D. vulgaris* DePue and the Global Ocean Sampling and ALOHA/HOTS metagenomic data sets were performed using CAMERA (<http://camera.calit2.net/index.php>).

CRISPR analysis

The CRISPR regions were determined using the Java-based CRISPR recognition tool and FASTA files of both the chromosome and megaplasmid from strains DePue and Hildenborough (Bland *et al.*, 2007). The minimum number of repeats was set to 3, the minimum repeat length set to 20 and the maximum repeat length set to 45. Search windows of 6 and 9 were used separately and compared, although no difference was noted. The minimum spacer length was set at 15 nt and the maximum spacer length set at 50 nt. All spacer and repeat sequences were compared against the genomes of each strain, the general NCBI database and the CAMERA metagenomic database using BLAST.

DNA extraction from *D. vulgaris* strains

Desulfovibrio vulgaris strains DePue, Hildenborough, Brockhurst Hill (NCIMB 8306), Woolwich (NCIMB 8457), Llanelly (NCIMB 8446) and Marburg (ATCC 35115) were grown and genomic DNA extracted as previously described (Walker *et al.*, 2006). In order to verify the identity of the *D. vulgaris* strains, PCR amplification and sequencing of most of the 16S rRNA, the complete 16S–23S internal transcribed spacer and a small portion of the 23S rRNA gene were performed as previously described (Walker *et al.*, 2006). Phylogenetic relationships were screened using BLAST (Altschul *et al.*, 1997).

Genomic DNA microarray comparison

The gDNA microarray hybridizations and imaging of *D. vulgaris* strains were performed as previously described (Walker *et al.*, 2006). Normalized signal intensities for each ORF were compared manually against the *D. vulgaris* Hildenborough control gDNA and signal intensity ratios calculated (sample

strain/strain Hildenborough). Signal intensity ratios less than 0.5 were considered as gene absent from genomes of comparison strains.

Acknowledgements

The authors wish to thank the Marcin Joachimiak (Lawrence Berkeley National Laboratory) for assistance with MicrobesOnline and David Bruce, Paul Richardson and Lynne Goodwin (Department of Energy Joint Genome Institute) for management of the genome sequencing.

References

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709–1712.
- Bland, C., Ramsey, T.L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N.C., and Hugenholtz, P. (2007) CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**: 209.
- Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuys, R.J., Snijders, A.P., *et al.* (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960–964.
- Campbell, A. (2003) Prophage insertion sites. *Res Microbiol* **154**: 277–282.
- Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., and Parkhill, J. (2005) ACT: the Artemis comparison tool. *Bioinformatics* **21**: 3422–3423.
- Chardin, B., Giudici-Orticoni, M.T., De Luca, G., Guigliarelli, B., and Bruschi, M. (2003) Hydrogenases in sulfate-reducing bacteria function as chromium reductase. *Appl Microbiol Biotechnol* **63**: 315–321.
- Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., DeLong, E.F., and Chisholm, S.W. (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**: 1768–1770.
- Cuadros-Orellana, S., Martin-Cuadrado, A.B., Legault, B., D'Auria, G., Zhaxybayeva, O., Papke, R.T., and Rodriguez-Valera, F. (2007) Genomic plasticity in prokaryotes: the case of the square haloarchaeon. *ISME J* **1**: 235–245.
- Goulhen, F., Gloter, A., Guyot, F., and Bruschi, M. (2006) Cr(VI) detoxification by *Desulfovibrio vulgaris* strain Hildenborough: microbe–metal interactions studies. *Appl Microbiol Biotechnol* **71**: 892–897.
- Heidelberg, J.F., Seshadri, R., Haveman, S.A., Hemme, C.L., Paulsen, I.T., Kolonay, J.F., *et al.* (2004) The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. *Nat Biotechnol* **22**: 554–559.
- Hueck, C.J. (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* **62**: 379–433.
- Johnston, S., Lin, S., Lee, P., Caffrey, S.M., Wildschut, J., Voordouw, J.K., *et al.* (2008) A genomic island of the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough promotes survival under stress conditions while decreasing the efficiency of anaerobic growth. *Environ Microbiol* **11**: 981–991.
- Kantar, C., Cetin, Z., and Demiray, H. (2008) *In situ* stabilization of chromium(VI) in polluted soils using organic ligands: the role of galacturonic, glucuronic and alginic acids. *J Hazardous Materials* **159**: 287–293.
- Klonowska, A., He, Z., He, Q., Clark, M.E., Theiman, S.B., Hazen, T.C., *et al.* (2007) Global transcriptomic analysis of chromium(VI) export of *Desulfovibrio vulgaris* Hildenborough under sulfate-reducing conditions. In: *American Society for Microbiology 107th General Meeting*. Toronto, Canada: American Society for Microbiology Press.
- Kunin, V., Sorek, R., and Hugenholtz, P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* **8**: R61.
- Lindell, D., Jaffe, J.D., Coleman, M.L., Futschik, M.E., Axmann, I.M., Rector, T., *et al.* (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**: 83–86.
- Lloyd, J.R., and Lovley, D.R. (2001) Microbial detoxification of metals and radionuclides. *Curr Opin Biotechnol* **12**: 248–253.
- Mojica, F.J.M., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**: 174–182.
- Rabus, R., Hansen, T., and Widdel, F. (2005) Gram-negative sulfate-reducing bacteria. In *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community*. Ballows, A., Trueper, H.G., Dworkin, M., Harder, W., and Schleifer, K.H. (eds). New York, USA: Springer-Verlag.
- Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., *et al.* (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5**: 398–431.
- Tettelin, H., Massignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., *et al.* (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial ‘pan-genome’. *Proc Natl Acad Sci U S A* **102**: 13950–13955.
- Walker, C.B., Stolyar, S.S., Pintel, N., Yen, H.C.B., He, Z.L., Zhou, J.Z., *et al.* (2006) Recovery of temperate *Desulfovibrio vulgaris* bacteriophage using a novel host strain. *Environ Microbiol* **8**: 1950–1959.
- Williams, K.P. (2002) Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res* **30**: 866–875.
- Williamson, S.J., Rusch, D.B., Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass, J.I., *et al.* (2008) The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS ONE* **3**: e1456.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Chromosome genomic plot of *D. vulgaris* DePue. Center ring indicates GC skew $(G - C)/(G + C)$ with 1 kb steps and 10 kb windows; 2nd ring is deviation from average G + C (higher A + T in red, higher G + C in blue); 3rd ring is rRNA loci; 4th ring represents tRNA loci; outer ring represent ORFs color-coded according to clusters of orthologous group (COG) categories. The position markers represent 100 kb windows.

Fig. S2. Synteny plots for the chromosome and megaplasmid of strains DePue and Hildenborough.

Fig. S3. Clusters of orthologous groups (COG) for strains DePue and Hildenborough. Categories are amino acid transport (E), carbohydrate transport and metabolism (G), cell division and chromosome partitioning (D), cell envelope biogenesis (M), cell motility and secretion (N), chromatin structure and dynamics (B), coenzyme metabolism (H), DNA replication, recombination and repair (L), defence mechanisms (V), energy production and conservation (C), function unknown (S), general function prediction only (R), inorganic ion transport and metabolism (P), intracellular trafficking and secretion (U), lipid metabolism (I), nucleotide transport and metabolism (F), post-translational modification, protein turnover, chaperones (O), secondary metabolites biosynthesis, transport and catabolism (Q), signal transduction mechanisms (T), transcription (K) and translation, ribosomal structures and biogenesis (J).

Fig. S4. Regions of strain DePue and strain Hildenborough genomes containing genes of primarily associated with phage function.

Fig. S5. Comparison of strains DePue and Hildenborough plasmid gene content. The outer ring corresponds to ORFs

identified in the DePue plasmid. The adjacent ring displays homologues in strain Hildenborough shaded in correspondence similarity to the strain DePue counterparts. The next interior rings display ORFs unique to strain Hildenborough. The interior and adjacent rings display, respectively, G + C skew and deviation from average %GC. Strain DePue ORFs discussed in the text are labelled by locus number.

Table S1. Reciprocal amino acid identities of orthologues for *D. vulgaris* strains DePue and Hildenborough.

Table S2. ORFs present in strain DePue and not strain Hildenborough. HP, hypothetical protein; CHP, conserved hypothetical protein.

Table S3. ORFs within the strain DePue unique exopolysaccharide coding region.

Table S4. CRISPR spacer sequences for strains DePue and Hildenborough.

Table S5. Prophage density in select microbial genome sequences.

Table S6. Signal intensity ratios of gDNA from *Desulfovibrio vulgaris* strains hybridized against a strain Hildenborough complete genome microarray.

Table S7. 16S rRNA nucleotide identity of *D. vulgaris* strains.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.