 doi:10.1111/j.1462-2920.2004.00665.x

The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments

R. Rabus,[1‡] A. Ruepp,[2†‡] T. Frickey,[3] T. Rattei,[4] B. Fartmann,[5] M. Stark,[2,6] M. Bauer,[1] A. Zibat,[2] T. Lombardot,[1] I. Becker,[2] J. Amann,[1] K. Gellner,[2] H. Teeling,[1] W. D. Leuschner,[2] F.-O. Glöckner,[1] A. N. Lupas,[3] R. Amann[1] and H.-P. Klenk[2,6*]

[1]*Max-Planck-Institute for Marine Microbiology, Celsiusstr. 1, 28359 Bremen, Germany.*

[2§]*EPIDAUROS Biotechnologie AG, Am Neuland 1, 82347 Bernried, Germany.*

[3]*Max-Planck-Institute for Developmental Biology, Spemannstr. 35, 72076 Tübingen, Germany.*

[4]*Technical University Munich, Genome-oriented Bioinformatics, Am Forum 1, 85354 Freising, Germany.*

[5]*MWG Biotech AG, Anzinger Str. 7, 85560 Ebersberg, Germany.*

[6]*e.gene Biotechnologie GmbH, Pöckinger Fußweg 7a, 82340 Feldafing, Germany.*

Summary

Desulfotalea psychrophila **is a marine sulfate-reducing δ-proteobacterium that is able to grow at *in situ* temperatures below 0°C. As abundant members of the microbial community in permanently cold marine sediments, *D. psychrophila*-like bacteria contribute to the global cycles of carbon and sulfur. Here, we describe the genome sequence of *D. psychrophila* strain LSv54, which consists of a 3 523 383 bp circular chromosome with 3118 predicted genes and two plasmids of 121 586 bp and 14 663 bp. Analysis of the genome gave insight into the metabolic properties of the organism, e.g. the presence of TRAP-T systems as a major route for the uptake of C_4-dicarboxylates, the unexpected presence of genes from the TCA cycle, a TAT secretion system, the lack of a β-oxidation complex and typical *Desulfovibrio* cytochromes, such as c_{553}, c_3 and ncc. *D. psychrophila* encodes more than 30 two-component regulatory systems, including a new Ntr subcluster of hybrid kinases, nine putative cold shock proteins and nine potentially cold shock-inducible proteins. A comparison of *D. psychrophila*'s genome features with those of the only other published genome from a sulfate reducer, the hyperthermophilic archaeon *Archaeoglobus fulgidus*, revealed many striking differences, but only a few shared features.**

Received 18 February, 2004; revised 30 April, 2004; accepted 30 April, 2004. *For correspondence. E-mail hans-peter.klenk@egene-biotech.de; Tel. (+49) 8157 9963 164; Fax (+49) 721 151 591 214. †Present address: GSF – Institute for Bioinformatics, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany. ‡These authors contributed equally to this work. §Genomics team disassembled; none of the authors available under this address.

Introduction

Large parts of the Earth's biosphere, in particular marine sediments, are predominantly or even permanently cold. Yet, comparatively little is known about the physiology and genomics of the organisms naturally occurring in these environments. It has long been known that, in temperate marine sediments, up to 50% of carbon remineralization is coupled to sulfate reduction (Jørgensen, 1982). Recent studies indicated that carbon turnover and sulfate reduction in Arctic marine sediments are comparable to those in temperate environments (Sageman *et al*., 1998). Thus, psychrophilic sulfate-reducing bacteria are assumed to contribute significantly to the global carbon and sulfur cycles. *Desulfotalea psychrophila*, the sulfate-reducing δ-proteobacterium, the genome of which we describe here, was isolated from permanently cold Arctic sediments off the coast of Svalbard at 79°N, 11°E (Knoblauch *et al*., 1999). *Desulfotalea* spp. were shown previously to represent abundant members of the microbial communities in Arctic marine sediments (Sahm *et al*., 1999; Ravenschlag *et al*., 2000). *D. psychrophila* grows optimally at 10°C, but is also able to grow at temperatures as low as –1.8°C (Knoblauch *et al*., 1999). Phylogenetically, most sulfate reducers belong to the δ-Proteobacteria; however, the only other sulfate-reducing organism whose genome has been published previously is *Archaeoglobus fulgidus*, an 'oil-well souring' hyperthermophilic archaeon isolated from a markedly different environment (Stetter, 1988; Klenk *et al*., 1997).

Desulfotalea psychrophila strain LSv54 (DSM 12343) is the type strain of the genus *Desulfotalea* ('sulfate-reducing rod'). Members of this genus are Gram-negative, obligatory anaerobic bacteria that use sulfate as

main electron acceptor (Knoblauch *et al*., 1999). Major known carbon or energy sources for members of the genus *Desulfotalea* are lactate, alcohols and hydrogen. Such compounds are generated by fermentative bacteria and are known components of porewater in marine sediments (Sørensen *et al*., 1981; Llobet-Brossa *et al*., 2002). *D. psychrophila* ('cold-loving') has thin, long, rod-shaped cells and an optimal doubling time of 27 h during growth with lactate at 10°C. In addition to sulfate, *D. psychrophila* can also reduce thiosulfate and sulfite to sulfide. *D. psychrophila* uses several organic acids, alcohols and amino acids as growth substrates (Knoblauch *et al*., 1999).

Since the publication of the first genome of an extremophile, *Methanococcus jannaschii* (Bult *et al*., 1996), many efforts have been undertaken to extend our knowledge about genomics of organisms thriving under extreme environmental conditions. With the completion of the *D. psychrophila* genome, we expand the temperature range for organisms with sequenced genomes to a new minimum, although from the environmental perspective, cold environments are common rather than 'extreme' on our planet. Through the analysis of their genomes, extremophilic organisms have provided unique insights into all aspects of their biology. Our academic interest in *D. psychrophila* was to explore the molecular basis of its physiological properties only defined by growth experiments to date (Knoblauch *et al*., 1999). The genome sequence of *D. psychrophila* will also contribute to our general understanding of sulfate-reducing bacteria and allow a comprehensive comparison with the archaeal counterpart, *A. fulgidus*. For growth to occur in low-temperature environments, all cellular processes must be adapted to the cold. This fact, in combination with the diversity of archaea, bacteria and eukaryotes recently isolated from cold environments (Russel and Hamamoto, 1998), highlights the extent of biological products and processes that might soon be exploited for biotechnology. Relative to this undisputed potential, psychrophiles and their products are underused in biotechnology; however, recent advances, particularly with cold-active enzymes, herald rapid growth for this burgeoning field.

Results and discussion

General genome features

The genome of *D. psychrophila* strain LSv54 consists of a 3 523 383 bp circular chromosome and two circular plasmids (121 586 bp and 14 663 bp). Table 1 summarizes the general features of the chromosome and its annotation. A total of 3118 predicted open reading frames (ORFs) with an average length of 968 bp were identified on the chromosome. Seventy-three per cent of the predicted proteins had homologues in public databases; thus, *D. psychrophila* displays a far genetic distance to other organisms with published genomes. Fifty per cent of the predicted proteins could be assigned to a functional role. Six hundred and forty-six of the predicted proteins (21%) could be assigned to protein families. Several of these families (containing a total of 308 members) apparently resulted from recent gene duplications. The members of these latter families are either unique for *D. psychrophila*, or homologues in other organisms are only distantly related to the duplicated *D. psychrophila* proteins. The largest protein families are two-component response regulators (48 members) and ATP-binding subunits of ABC transporters (42 members).

Table 1. General genome features[a].

Genome size (bp)	3 523 383
G+C content (%)	46.8
Coding density[b] (%)	86.6
Predicted protein coding sequences	3 118
with homologues in databases	2 265
with functional assignments	1 545
belong to at least one COG	2 032
contain PFAM domains	1 665
with homologous 3D structure known	752
with ≥2 *trans*-membrane regions	532
with coiled-coil regions	69

a. Chromosome only, without plasmids.
b. 1.02% from 86 genes encoding stable RNAs.

The chromosome contains seven complete 16S–23S–5S operons. Two of them are interrupted between the 16S and 23S rDNAs by genes for $tRNA^{ile}$ and $tRNA^{ala}$ (Fig. 1). Three of the rDNA operons are located in a cluster at 83.3° on the chromosome circle (position 806 225–827 008), which generated a challenge for the assembly of the genome. A total of 64 tRNA genes were identified (Fig. 1). None of them contains an intron, and only 12 are clustered in six tandems of $tRNA^{val}$ and $tRNA^{asp}$ (2 888 212–2 889 554). The gene for tmRNA (Williams and Bartel, 1996) has been identified. In addition to the tRNAs for the 20 regular amino acids, *D. psychrophila* also contains *selC* for the selenocysteine tRNA. The other components necessary for the selenocysteine system, selenocysteine synthetase (*selA*), the specific elongation factor (*selB*) and the selenophosphate synthetase (*selD*), are also present in the genome (Fig. 1). Nine genes with the selenocysteine codon UGA were identified: dTDP-glucose 4,6-dehydratase, ribosomal protein S6 (two UGA codons each), Fe-S oxidoreductase, heterodisulfide reductase subunit A, the δ-subunit of methyl viologen-reducing hydrogenase, as well as four genes for formate dehydrogenase α-subunits (Fig. 1). Two of these formate dehydrogenases are of the membrane-bound *Wolinella* type, whereas the other two enzymes are cytoplasmic with similarities to the enzymes from *Ralstonia* and *Des-*

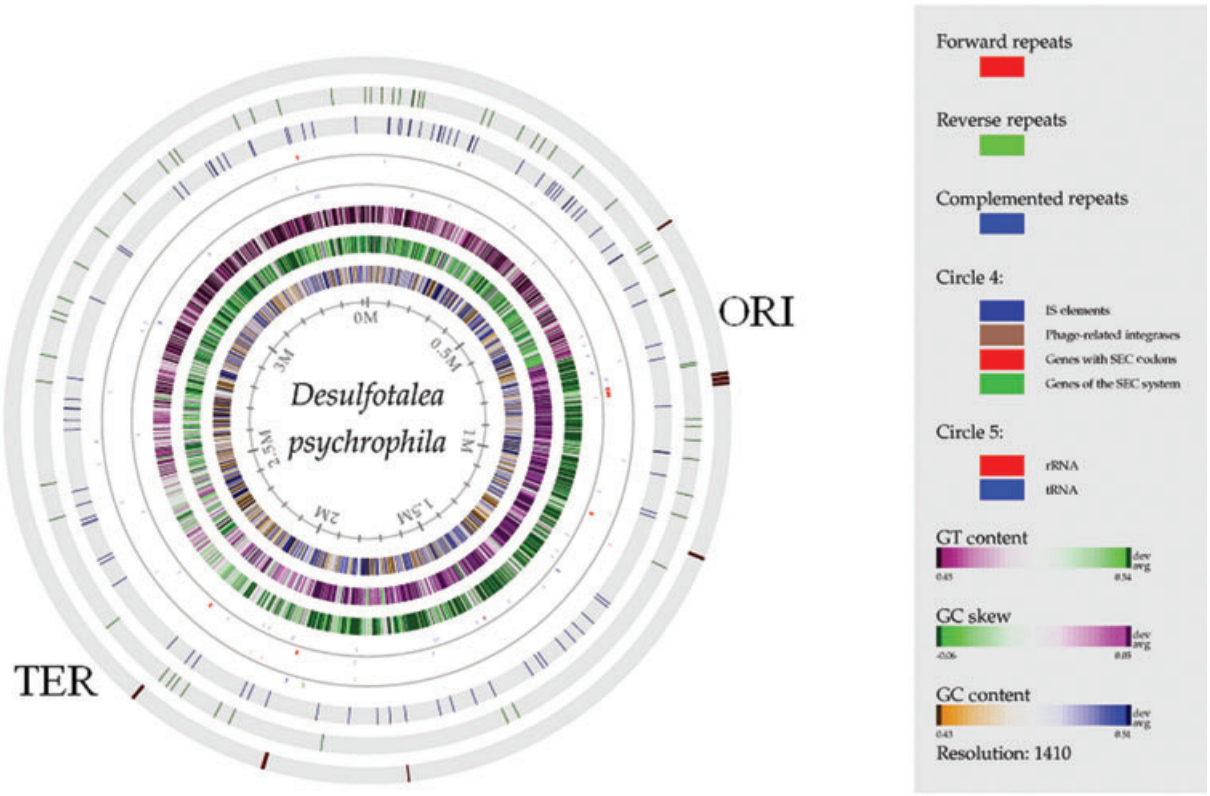

Fig. 1. Structural representation of the chromosome. Circles (from inside outwards): 1, GC skew; 2, G+C content; 3, G+T content; 5, location of rRNA operons (red) and tRNAs (blue); 4, IS elements (dark blue), phage-related integrases (light blue), location of genes with selenocysteine codons (red) and the selenocysteine genes, *sel*ABCD (yellow); 6, complemented repeats; 7, reverse repeats; 8, forward repeats. The origin (ORI) and terminus (TER) of replication are indicated.

ulfovibrio. No inteins have been identified in the proteome of *D. psychrophila*.

Origin and terminus of replication are located at 72.4° and 231.4°, respectively, on the circular chromosome (Fig. 1). Evidence for the postulated *oriC* was derived from the minimum of the cumulative G+C skew and the close proximity of several DnaA binding sites. The genes for DNA polymerase III (*dnaN*) and the subunits of DNA gyrase B (*gyrB*) are located upstream of *oriC* on the leading strand.

The genome contains five regions with low G+C content (<40%). One of these regions (1 271 920–1 290 190) contains a cluster of 32 informational proteins (ribosomal proteins, translation elongation factors and RNA polymerase α-subunit); the other regions (249 000–259 200, 1 271 920–1 290 190, 1 577 790–1 586 810 and 1 293 990–2 303 700) encode conserved hypothetical proteins and predicted proteins without functionally characterized homologues in other genomes. The genome also contains five regions of high G+C content (>53%) encoding genes for cobalamin biosynthesis (229 900–240 770), menaquinone and coenzyme F390 biosynthesis (274 000–285 430), branched-chain amino acid biosynthesis (998 490–1008 150), a large 267 kDa (2630 amino acids) protein with a homologue in *Pseudomonas aeruginosa* (2 406 930–2 417 870) and a functionally diverse ORF cluster (2 810 890–2 823 140) (Fig. 1). The chromosome has only relatively few repeats and mobile elements (Fig. 1). Only five complete IS elements were identified (besides fragments for three partial elements). The amino acid sequence of one of the IS elements (DP0860 + DP0861) is identical to *Escherichia coli insAB*, indicating a recent acquisition of this mobile element. Furthermore, the chromosome harbours four phage-related integrases (Fig. 1).

The G+C content of both plasmids is significantly lower than that of the chromosome (46.8%): large plasmid 43.6%, small plasmid 27.5%. The large plasmid is dominated by genes for conjugal transfer functions, *trbBCEIJLFGM* and *traCGFIJ*, and plasmid-specific functions, e.g. plasmid partition protein ParF or TrfA. The large plasmid also encodes a complete pyruvate dehydrogenase complex and a functionally uncharacterized large protein of 472 793 Da (4749 amino acids). Coding regions for three IS elements were identified; however, all transposase-encoding regions appear to be fragmentary and

are probably non-functional. As on the chromosome, one putative reverse transcriptase/maturase has been identified. In the case of the small plasmid, only three genes with assigned functions were detected: replication protein RepA, a putative primase PrmN1 and a homologue of TrsK/TraK for which involvement in DNA transfer during conjugal transposition has been predicted (Alarcon-Chaidez *et al*., 1997).

Special features of the proteome

Distribution of role categories. Proteins involved in information processing constitute the largest fraction of predicted functions (15%), followed by energy conservation and carbohydrate metabolism/transport (10%). General transport functions (9%) and defence mechanisms (4%) indicate the importance of interaction with the environment for *D. psychrophila*. Seven per cent of the predicted proteins are involved in nucleotide or amino acid metabolism and transport, and more than 2% belong to signal transduction. Twenty-seven per cent of the predicted proteins represent hypothetical proteins.

Sensory functions and regulation of gene expression. Most putative histidine kinases (HKs) from *D. psychrophila* could be assigned to six of the 12 recently defined phylogenetic groups (Koretke *et al*., 2000): Ntr (19 members), Hybrid (five members), Pho (four members), Nar, Cit and Che (one member each). No members of the HK groups Sam, Syn, Anti-Sigma, Mth, Lyt and Arf were found (Fig. 2). In many Bacteria, about two-thirds of all HKs fall into the Hybrid group. In contrast, most HKs from *D. psychrophila* cluster in the Ntr group, where they form a new subcluster of phylogenetically distinct HKs. This subcluster does not include any non-*Desulfotalea* members, even when considering the newest HK entries in the public databases (Fig. 2). Three of the nine members of this new Ntr subcluster show highest sequence similarity among each other and are located directly next to each other on the genome. A common trait of all nine members is the fusion of an HK domain to a response regulator (RR). All defined HK clusters and domain organizations of their respective members are supported by the co-evolution of RR clusters and domains (Fig. 2). Furthermore, *D. psychrophila* has a large number (seven out of 31) of HKs fused to a periplasmic-binding protein. Such fusions have only occasionally been observed in other organisms.

Mobility and chemotaxis. Mobile cells of *D. psychrophila* have only been reported for old cultures (Knoblauch *et al*., 1999). The genes for 22 proteins involved in flagellar organization have been identified, with only a few of the *fli* and *flh* homologues from *E. coli* remaining unaccounted for. In addition, we identified two clusters of *che* genes [*cheABR*(*W*)*Y*], but not *cheCDJLVXZ*. While one *cheABRWY* cluster is 'traditional', the domain architecture of the second *cheABR-Y* cluster is very unconventional for the kinase. The complete mechanism of mobility in *D. psychrophila* remains unresolved at present.

Surface structures. The surface proteins of *D. psychrophila* apparently possess the structural domains known from some pathogens. We have identified a surprisingly large number of autotransporters similar to *Yersinia pestis* YapF. Four very large cell surface proteins were identified, two of them having a complex domain structure with local similarity to a surface protein of the pathogen *Salmonella enterica*, one is similar to RTX exoproteins (primarily known from pathogenic *Vibrio* spp. and enterohaemorrhagic *E. coli*), and one is similar to adhesion proteins known from proteobacterial pathogens. Furthermore, we identified eight TolC-like outer membrane efflux pores and one classical 16-stranded porin, but no classical S-layer protein.

Vitamins and cofactors. Analysis of the genome corroborates the fact that *D. psychrophila* does not require vitamins for growth (Knoblauch *et al*., 1999). We have identified 85 coding regions (2.6% of the predicted proteins) to be involved in biosynthesis of vitamins, cofactors and prosthetic groups, including pathways for the biosynthesis of thiamine, folate, nicotinamide, riboflavin and porphyrin. A stretch of 13 consecutive genes encoding most components of cobalamin biosynthesis and a metal ion transport system has been identified at 23.8° on the circular chromosome. The cobalamin biosynthesis in *D. psychrophila* seems to be most closely related to that of *Pseudomonas aeruginosa* (http://www.genome.ad.jp/kegg/).

Secretion of proteins. Export of metal cofactor-containing enzymes such as formate dehydrogenase and hydrogenases to the periplasm is probably mediated by the TAT system, which allows translocation of the mature proteins across the cytoplasmic membrane (Berks *et al*., 2003). This system is composed of a channel-forming subunit (*tatA*) and a receptor complex (*tatBC*) that is involved in signal peptide recognition. All three genes are present in the genome, with three copies for *tatC*. Upstream of the *tatABC* genes, a *tatE* homologue was detected, the function of which is currently unknown. Apparently, the general secretion system (Sec) is also present in *D. psychrophila*, as indicated by many identified signal leader peptides.

Structural and genomic aspects of Psychrophilia

The first whole-genome sequence of a psychrophilic bacterium, as presented here, allows the investigation of

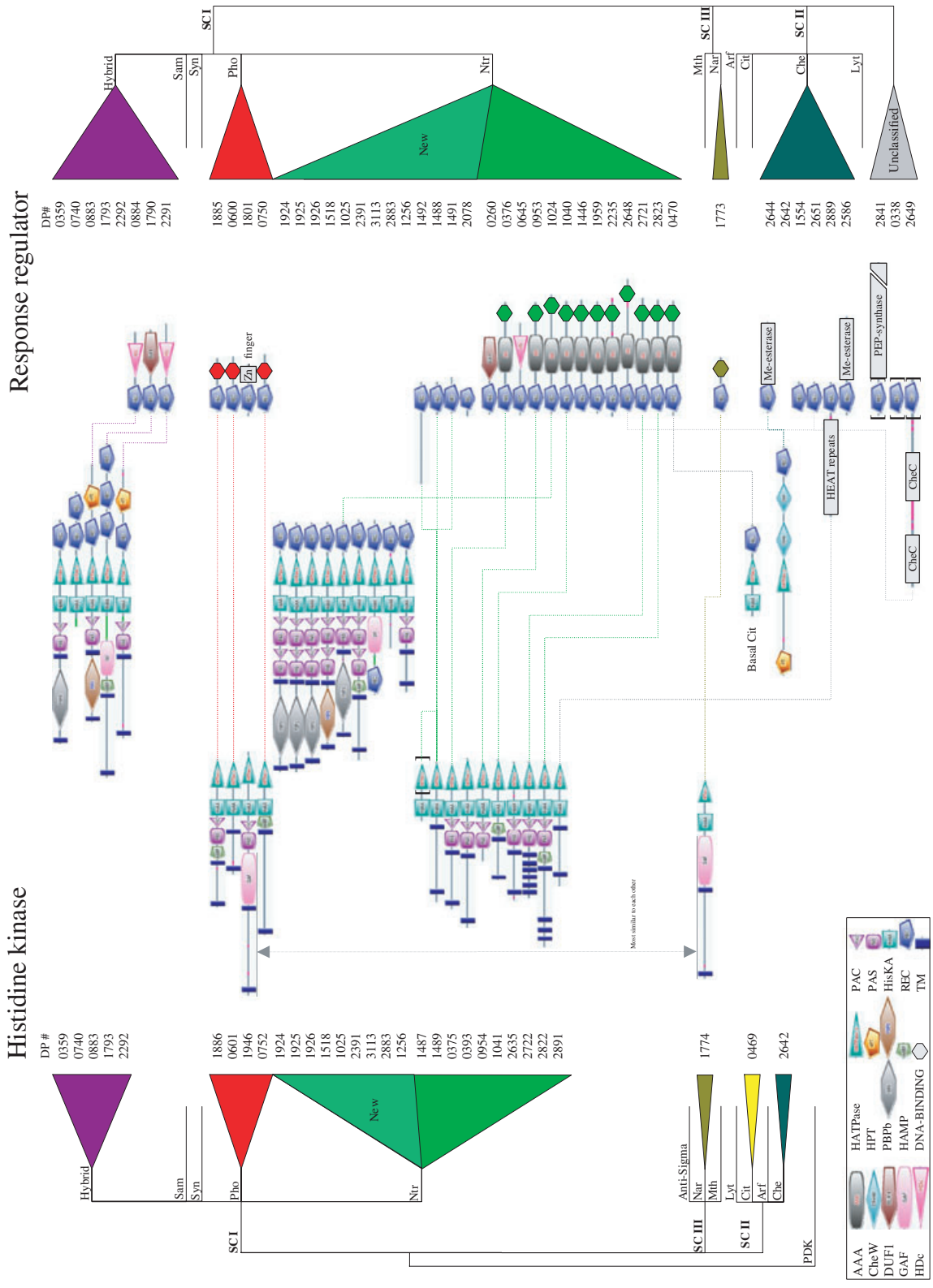

Fig. 2. Phylogenetic map of histidine kinase and response regulator proteins. Phylogenetic clustering is according to Koretke and colleagues (2000). Fused kinase and response regulator hybrids are in the middle. Kinases are to the left and response regulators to the right of the figure. Coloured and grey lines connect proteins in genomic proximity within or across clusters respectively. The domain architecture of the proteins was generated by SMART (Letunic *et al.*, 2002) and refined manually. Square brackets encase degenerate domains. Proteins containing squares with 50% grey hatching indicate potentially disrupted proteins (frameshift or stop), with the hatched square covering one fragment. Domain legend is at the bottom left.

genome-wide characteristics that distinguish psychrophiles from mesophiles and thermophiles. Draft genome sequences of the psychrophilic *Methanogenium frigidum* (Franzmann *et al*., 1997) and the psychrotolerant *Methanococcus burtoni* (Franzmann *et al*., 1992) were generated recently to study adaptation to different temperatures (Saunders *et al*., 2003). Comparative analysis revealed that proteins predicted from the partial genomes of these two archaea are characterized by a higher content of non-charged polar amino acids, particularly glutamine (Gln) and threonine (Thr), and a lower content of hydrophobic amino acids, particularly leucine (Leu). We compared the codon and amino acid composition of *D. psychrophila* with that of a selection of published genomes of organisms with optimal growth temperatures ranging from 27°C (*Streptomyces lividans*) to 100°C (*Pyrobaculum aerophilum*). The results indicated that there is no strict correlation between the optimal growth temperature and the nucleotide composition at any codon position. A noteworthy exception might be that *D. psychrophila* appears to have unusually few guanines (21.6%) and unusually many thymines (30.9%) in the third codon position. Our results in amino acid composition in *D. psychrophila* do not corroborate the observation of Saunders and colleagues (2003) for cold-adapted archaea, namely a higher content of non-charged polar amino acids, particularly Gln and Thr. For Gln, this may be the case relative to the hyperthermophiles (83–100°C, 1.8% Gln), but the coding regions in the genomes of *Pseudomonas putida* (28°C, 4.7%), *E. coli* (37°C, 4.3%) and *Deinococcus radiodurans* (30°C, 4.1%) all contain more Gln codons than *D. psychrophila* (10°C, 3.9%). Also for Thr, *D. psychrophila* (10°C, 5.0%) has a lower fraction of codons compared with some mesophiles such as *S. lividans* (27°C, 6.8%), *D. radiodurans* (30°C, 5.8%) and *E. coli* (37°C, 5.6%). In fact, *D. psychrophila*'s composition of Leu codons and codons for several other amino acids resembles more closely that observed in hyperthermophiles than that in mesophiles. It might be worth noting that *D. psychrophila* contains unusually many codons for cysteines (2.2%), compared with mesophiles (average 0.9%) and hyperthermophiles (average 0.8%), but this might result from the many redox proteins for the sulfur metabolism, an observation corroborated by the so far published genes of *Desulfovibrio* spp. (http://www.kazusa.or.jp/codon).

The low temperatures in *D. psychrophila's* regular habitat supposedly exert no constraints on the stability of its mRNA. Chargaff differences (a measure of purine loading in mRNAs) are greater in thermophiles than in non-thermophiles (Lao and Forsdyke, 2000). The pressure to purine load affects codon choice in thermophiles, indicating that some features of their amino acid composition, e.g. a high level of glutamic acid (Glu; 8.4%), might reflect a purine-loading pressure resulting from constraints on mRNA (Lao and Forsdyke, 2000). The purine-loading index (PLI) of *D. psychrophila* (0.51) is significantly lower than that of the hyperthermophiles (0.55–0.57) and corresponds to the PLI found in many mesophiles (0.50) and even the moderately thermophilic *Thermus thermophilus* (optimal growth temperature 75°C), thus indicating that purine loading is only a relevant feature in hyperthermophiles (optimal growth temperature >85°C). The fraction of Glu codons in the genome of *D. psychrophila* equals that found in most mesophiles (5.6%) and is far below that found in hyperthermophiles (8.4%).

Cold shock proteins. Cold shock proteins (CSPs) are induced upon cold shock and are thought to bind a single-stranded (ss)RNA motif (DB box and RNP1), resulting in reduced secondary structure formation in the RNA and thus increased translation efficiency. Nine *D. psychrophila* homologues of known CSPs were identified. A search pattern [AAxTGGxTx(5–25)TTxGGxT] was generated by looking for similarities between known DB box and RNP1 motifs and the *D. psychrophila* sequences. A new search with this pattern provided an additional nine putative cold shock-inducible proteins, of which five had this pattern shortly after their transcriptional start. These hypothetical and conserved hypothetical proteins provide suitable targets for experimental investigation of their function. In addition to the regular CSPs, we identified three members of group 1 CSPs: PNP, RBFA and NUSA; 12 members of group 2 CSPs, five probable DNA-binding proteins HU, DNA-binding protein H-NS, DNA gyrase subunit A, topoisomerase IV subunit A, RecA, DnaK, trigger factor, translation initiation factor IF-2; and five copies of FABF (group 3 CSPs).

Growth at low temperature depends on the ability of cells to synthesize proteins at low temperature (Thomas and Cavicchioli, 2003). Analysis of modelled protein structures from cold-adapted archaea showed a strong enrichment of Gln, Thr and hydrophobic residues and fewer charged residues in the solvent-accessible area (Saunders *et al*., 2003). Even though no modelling was performed with the predicted proteins, the overall codon frequency in the genome indicates that this effect might be less pronounced in *D. psychrophila*. A detailed analysis of the transcription elongation factor G from low temperature-adapted bacterium *Arthrobacter globiformis* SI55 (Berchet *et al*., 2000) suggested several structural features that might be important for activity and flexibility at low temperatures: fewer salt bridges in intra- and interdomain positions; increased solvent interactions mediated by greater charge and polarity on domain surfaces; loop insertions; loss of proline residues in loop structures; and an increase in hydrophobicity in core regions. Psychrophilic or psychrotolerant organisms have a superior ability to perform translation at low temperature, and proteins

involved in translation might therefore by crucial for the organisms' ability to grow. *D. psychrophila* maintains genes for three highly conserved but distinct translation elongation factors G that might contribute to this process.

The mean G+C content of the tRNAs from *D. psychrophila* (60.4%) is significantly higher than that in the protein coding regions (46.2%) and corresponds to the mean G+C contents of tRNAs from mesophiles. Thus, there seems to be no specific effect of low temperature habitats on the G+C content of tRNAs of the organisms living therein. Previous studies on tRNAs from psychrophilic bacteria indicated that the maintenance of flexibility at low temperature is important for tRNA function, and that this is mainly achieved through post-transcriptional incorporation of dihydrouridine (Dalluge *et al*., 1997), whereas in hyperthermophiles, a variety of modified nucleotides is incorporated (McCloskey *et al*., 2001). *D. psychrophila* encodes a putative dihydrouridine synthase (DP2602) in its genome, in addition to a dozen other tRNA-modifying enzymes. As this is in contrast to cold-adapted archaea, which belong to those organisms with the lowest level of tRNA modification (Noon *et al*., 2003), it might be of interest to study the tRNA modification in *D. psychrophila* experimentally. In addition, *D. psychrophila* might use a new mechanism of transcriptional regulation involving a response regulator fused to HEAT repeats; this type of fusion protein has no homologues in the current databases. *D. psychrophila* also has DEAD box RNA helicases, which are involved in essential cellular processes, e.g. translation initiation and ribosome biogenesis. Such helicases have recently been implicated in enabling bacteria to survive cold shock and grow at low temperature (Lim *et al*., 2000).

Lipids. Cold adaptation has long been known to influence the properties of the cytoplasmic membrane. A decrease in the phase transition temperature is achieved by an increased ratio of unsaturated fatty acids (Gounot and Russell, 1999). In *D. psychrophila*, unsaturated fatty acids can account for about 80% of total fatty acids (Knoblauch *et al*., 1999; Könneke and Widdel, 2003). Correspondingly, *D. psychrophila* possesses, in addition to the enzymes of regular fatty acid biosynthesis, two copies of β-keto-acyl-CoA synthase I, which elongates unsaturated fatty acids.

Metabolic capacities

The prominent role of sulfate-reducing bacteria during carbon remineralization in marine sediments prompted us to reconstruct the metabolic capacities of *D. psychrophila* previously defined with growth experiments (Knoblauch *et al*., 1999) at the genomic level. Integration of identified genes/pathways (Figs 3 and 4) will contribute to our general understanding of the molecular physiology of sulfate-reducing bacteria and allow selection of key target genes for *in situ* expression analysis (Neretin *et al*., 2003).

C_4-dicarboxylates. Uptake of malate, fumarate and succinate probably occurs preferentially via TRAP transporters, as the genome contains genes for eight complete DctPQM sets, two individual binding proteins (DctP) and one individual permease (DctM). Interestingly, TRAP transporters are apparently not encoded in the genome of *A. fulgidus* (Klenk *et al*., 1997). The recently recognized TRAP-T family (Shaw and Kelly, 1991; Rabus *et al*., 1999a) is unique in combining a high-affinity periplasmic substrate-binding protein (DctP) with H^+-driven permeases (DctMQ). In addition, four H^+ or Na^+/C_4-dicarboxylate symporters (Dcu or DAACS families) were detected. After uptake, C_4-dicarboxylates can be metabolized via reactions of the TCA cycle. The presence of malic enzyme allows conversion to pyruvate/acetyl-CoA for anabolic purposes.

Alcohols. Ethanol most probably enters the cells by simple diffusion across the cytoplasmic membrane. Alcohol- and acetaldehyde dehydrogenases (*adh*) oxidize ethanol to acetyl-CoA. A total of nine alcohol and four aldehyde dehydrogenases were predicted.

Amino acids. Among the 54 proteins involved in amino acid transport (about nine ABC transport systems), six Na^+- or H^+-dependent symporters possibly function in the uptake of alanine, glycine and serine. Deamination of alanine to pyruvate is catalysed by alanine dehydrogenase (*ald*; three paralogues). Glycine is converted to serine by the glycine cleavage system (*gcs*) and glycine hydroxymethyltransferase (*glyA*). Serine dehydratase (*sdh*) finally allows the formation of pyruvate from serine.

Lactate. Sulfate-reducing bacteria are most often cultivated with lactate. In *D. psychrophila*, uptake of lactate probably proceeds via an H^+-dependent permease (*lctP*). Oxidation of lactate to pyruvate is catalysed by lactate dehydrogenase (*ldh*). The divergently orientated *lctP* and *ldh* genes are closely linked on the genome sequence.

Pyruvate. Conversion of pyruvate to the central intermediate acetyl-CoA is probably catalysed by pyruvate:ferredoxin oxidoreductase (*por*, *pfor*, *ofor*). *PorABC* gene products are closely related to archaeal counterparts, whereas the *pfor* gene product is homologous to the enzyme from *Desulfovibrio africanus*. Recently, the structure of PFOR from *D. africanus* was resolved, demonstrating the involvement of TPP and Fe/S clusters during catalysis (Chabriere *et al*., 1999). Evidence for only one subunit of pyruvate dehydrogenase typical of aerobic

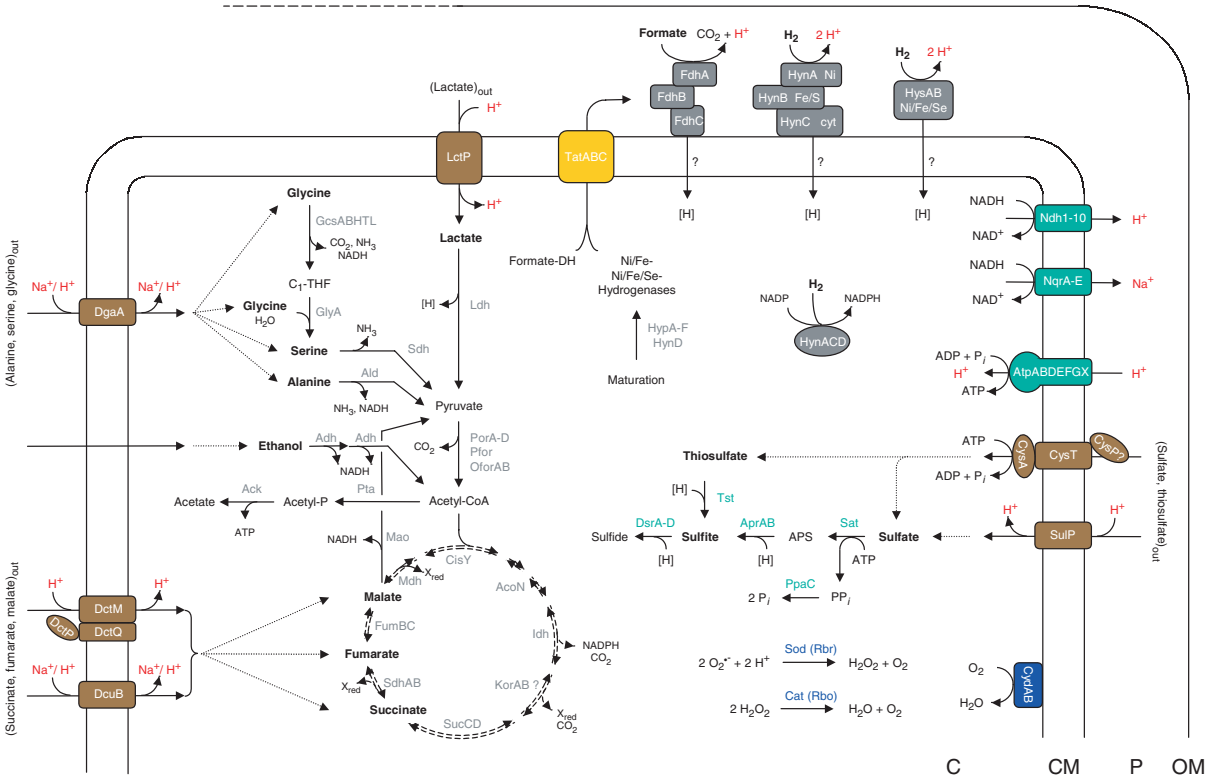

Fig. 3. Integration of predicted proteins involved in uptake and catabolism of known growth substrates and in energy generation. Transport systems (brown) allow uptake of organic substrates and electron acceptors. Catabolic enzymes (grey) are mostly localized in the cytoplasm. Periplasmic hydrogenases and formate dehydrogenase are exported in the folded state via the TAT system (yellow). Proteins involved in respiratory energy generation are indicated in green. Several proteins (blue) allow detoxification of oxygen. C, cytoplasm; CM, cytoplasmic membrane; P, periplasm; OM, outer membrane.

organisms was obtained. It should be noted that *D. psychrophila* also encodes predicted proteins similar to pyruvate formate lyase. The presence of phosphotransacetylase (*pta*) and acetate kinase (*act*) principally allows energy generation during acetate formation.

TCA cycle. Many genes from the TCA cycle are present in the genome of *D. psychrophila.* The implicated possibility of complete oxidation would be unexpected, as *D. psychrophila* was originally described as an incomplete oxidizer (Knoblauch *et al.*, 1999). In *Desulfobacter* sp., the TCA cycle is known as the route for complete oxidation of acetyl-CoA to CO_2, whereas other completely oxidizing sulfate-reducing bacteria use the C_1/CO-dehydrogenase pathway (Rabus *et al.*, 1999b). No evidence was obtained for this pathway to be functional in *D. psychrophila*, but it is present in *A. fulgidus* (Klenk *et al.*, 1997).

Previous studies (Knoblauch *et al.*, 1999) indicated that *D. psychrophila* LSv54 cannot grow with propionate or higher fatty acids. Consistent with these findings, neither

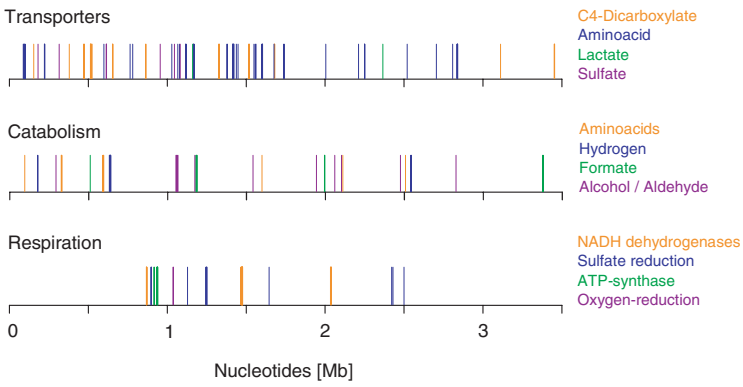

Fig. 4. Genomic organization of genes involved in transport, catabolism and respiration. Displayed genes correspond to those used for metabolic reconstruction (Fig. 3).

genes for methyl-malonyl CoA mutase required for propionate utilization nor key components (FadAB) of the β-oxidation complex could be detected in the genome. In contrast, more than 50 proteins involved in β-oxidation were predicted for *A. fulgidus* (Klenk *et al*., 1997).

Formate. Utilization of formate could possibly occur in the periplasm and cytoplasm. Two putative operons each coding for a periplasmic formate dehydrogenase (*fdhABC*) were detected in the genome. They are homologous to the enzyme from *Wolinella succinogenes* (Bokranz *et al*., 1991). The α-subunits (*fdhA*) carry twin-arginine motif-containing leader peptides, implying export of the mature protein via the TAT export system. In addition, most of the *hyc* genes coding for cytoplasmic formate hydrogenlyase in *E. coli* (Sauter *et al*., 1992) were found.

Hydrogen. Several hydrogenases are present in *D. psychrophila*. The *hynBAC* and the *hysAB* genes code for periplasmic Ni/Fe and Ni/Fe/Se hydrogenase respectively. The HynB and HysB subunits carry the twin-arginine motif-containing leader peptides, directing export via the TAT export system. The *hynBAC* genes are located in a large putative operon structure together with the *hypFC-DEhynDhypAB* genes, which are involved in hydrogenase maturation. In addition, *hnd* genes were detected, which probably code for an NADP-reducing hydrogenase reported previously for *Desulfovibrio fructosovorans* (Malki *et al*., 1995).

Regeneration of NAD. NADH dehydrogenases are usually required for regeneration of NAD and channelling of electrons into the respiratory chain. In *D. psychrophila*, we detected the *ndh1–10* genes coding for H^+-translocating NADH dehydrogenase complex. In addition, the *nrqA–E* genes encode a Na^+-translocating NADH-quinone oxidoreductase, possibly allowing the generation of a sodium gradient. The latter could energize Na^+-dependent symporters of organic substrates.

Sulfate reduction. The main mode of energy generation in *D. psychrophila* is sulfate reduction. Sulfate uptake was previously suggested to be driven by an ion gradient (Cypionka, 1989). Accordingly, an H^+-dependent symporter (*sulP*) for sulfate was predicted. Interestingly, evidence was also obtained for the presence of an ABC transporter (*cysPTA*) for sulfate/thiosulfate, which was also found in *A. fulgidus* (Klenk *et al*., 1997). Activation of the chemically sluggish sulfate to adenosine-5′-phosphosulfate (APS) is performed by ATP-sulfurylase (*sat*; two paralogues). This reaction is facilitated by pyrophosphatase (*ppaC*)-catalysed hydrolysis of PP_i. APS is reduced to sulfite by APS reductase (*aprAB*). The final step probably involves a six electron reduction of sulfite to sulfide, carried out by dissimilatory sulfite reductase (*dsrABCD*). Consistent with the ability of *D. psychrophila* to use thiosulfate as an electron acceptor, a thiosulfate sulfurtransferase (*tst*) was identified, which forms sulfite from thiosulfate. The mode of electron transfer from e.g. the quinone pool to the enzymes of sulfate reduction remains elusive so far.

ATPase. Proton gradients generated by periplasmic formate dehydrogenase, hydrogenases or H^+-translocating NADH-dehydrogenase complex are exploited by an F1/F0-type ATPase (*atpABDEFX*) to form ATP.

Cytochromes. Redox proteins have been intensively investigated in sulfate-reducing bacteria over recent years (Rabus *et al*., 1999b). In the well-studied *Desulfovibrio* species, many cytochromes have been identified and biochemically characterized. They were shown to be involved in electron transfer, e.g. from the periplasmic hydrogenases. Thus, it was a major surprise that *D. psychrophila* did not contain any homologue of cytochromes c_{553}, c_3 or ncc. With respect to transmembrane-spanning Hmc, only partial sequences could be detected. This finding was supported by the absence of absorption bands typically known for c-type cytochromes (Fig. 5). Analogously, such cytochromes appear to be absent in sulfur-reducing *Desulfurella acetivorans* (Bonch-Osmolovskaya *et al*., 1990), whereas they are present in sulfur-reducing *Desulfuromonas acetoxidans* (Pfennig and Biebl, 1976). The genome of the sulfur- and metal-reducing *Geobacter sulfurreducens* even contained 111 cytochrome-encoding genes (Methe *et al*., 2003). In *A. fulgidus* also, only a few cytochrome-encoding genes were detected (Klenk *et al*., 1997). Considering the unfavourable redox potential of cytochromes relative to the terminal electron acceptor sulfate, some sulfate reducers may use alternative electron carriers. In fact, we detected a flavodoxin (DP0917), which could function in periplasmic electron transfer, because it carries a leader peptide (twin-arginine containing). It will be a promising future goal to define the alternative electron carriers used by *D. psychrophila* and the extent to which cytochromes are present in genomes of other sulfur- and sulfate-reducing bacteria. The lack of c-type cytochromes may not contradict the suggested Fe(III)-reducing capacity of *D. psychrophila* (Knoblauch *et al*., 1999), as their role in Fe(III) reduction is disputed, and alternative electron shuttles between the cell and Fe(III) should also be considered (Straub and Schink, 2003).

Oxygen. The relation of sulfate-reducing bacteria to oxygen has long been disputed. Over the last several years, evidence has accumulated that *Desulfovibrio* species pos-

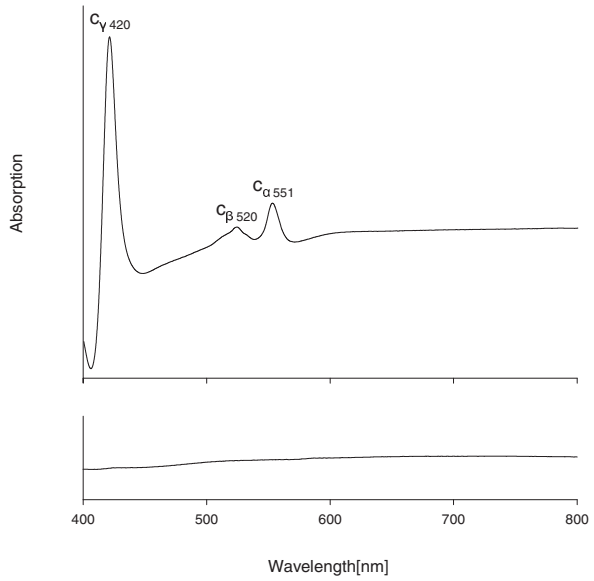

Fig. 5. Difference spectra (reduced minus oxidized) of cell-free extracts of *D. psychrophila* (lower line) and *D. vulgaris* (upper line). For *D. vulgaris*, c-type cytochromes could be detected. Numbers shown indicate the experimentally determined wavelengths of absorption (nm) for specific peaks. They fall into the range of α, β and γ absorption bands of c-type cytochromes as described by Poole (1994). For *D. psychrophila*, no indication of the presence of cytochromes was observed.

sess oxygen-reducing systems (Cypionka, 2000). Interestingly, *D. psychrophila* contains genes (*cydAB*) possibly encoding a cytochrome D ubiquinol oxidase, which is involved in the terminal step of oxygen respiration at low oxygen concentrations (Green *et al*., 1988). Such a cytochrome was recently purified from *Desulfovibrio gigas* and shown to reduce O_2 in the presence of NADH (Lemos *et al*., 2001). Some *Desulfovibrio* sp. were also shown to reduce O_2 in the periplasm (Baumgarten *et al*., 2001). Superoxide dismutase (*sod*) and catalase (*cat*) are also present in *D. psychrophila*, allowing oxygen detoxification. Both enzymes have long been known to occur in sulfate-reducing bacteria (Hatchikian *et al*., 1977). Rubrerythrin and rubredoxin oxidoreductases were recently described in *Desulfovibrio vulgaris* as an alternative oxygen protection system functioning similarly to superoxide dismutase/catalase (Lumppio *et al*., 2001). The coding *rbr* and *rob* genes are also present in *D. psychrophila*. To date, sulfate-reducing bacteria have not been demonstrated to grow aerobically (Cypionka**,** 2000), but rather to be growth inhibited by O_2 (Fareleira *et al*., 2003). This suggests that the above-mentioned systems function primarily in oxygen detoxification. As marine sediments generally contain dynamic oxic–anoxic interfaces with varying oxygen penetration depth, such systems are likely to be essential for survival. In addition, sulfate-reducing bacteria may actually respond to exposure to O_2 by directed migration to more anoxic zones, as indicated by band formation in gradient cultures of *Desulfovibrio* (Eschemann *et al*., 1999). Such behaviour could be also expected for *D. psychrophila*, as it possesses a chemotaxis system (see above).

Genomic organization. The organization of the genes described above is displayed in Fig. 4. Although genes involved in transport and catabolism are basically distributed across the entire genome, those coding for respiratory functions appear to be clustered in the middle section of the genome. Overall, the first half of the genome contains the majority of the respective genes.

Comparative genomics

The genetic distance of *D. psychrophila* from all other organisms with published genomes becomes obvious from the list of top hits generated by FASTA comparison with a database containing all presently published genomes. The largest fraction of 138 top hits (4.4% of the searched predicted proteins) is related to *Magnetococcus* mc1, with the next most frequently hit genome being that of *Vibrio cholerae* (102 hits, 3.3%), *Desulfitobacterium hafniense* (86 hits, 2.8%) and *Thermoanaerobacter tengcongensis* (82 hits, 2.6%). Number 19 on this ranking list is *E. coli* (26 hits, 0.8%), followed by the only other published genome of a sulfate reducer, *A. fulgidus* (24 hits, 0.8%). The largest fraction of the derived proteins, 839 (27.0%), have no detectable homologues outside the *D. psychrophila* genome.

Archaeoglobus fulgidus. Only 396 of the predicted proteins in *D. psychrophila* (12.8%) appear to have homologues in the genome of *A. fulgidus* on the basis of a ratio >0.2 between the heterologous score and the selfscore of the *D. psychrophila* protein. Presumably, most of the genes behind these ORFs belong to the universal gene set shared by all organisms. For only 24 of these 396 proteins is the respective *A. fulgidus* homologue the most similar protein. Besides nine conserved hypotheticals, these proteins function in energy metabolism and transport. Five proteins are exclusively shared between the two sulfate reducers and are not found in any of the other genomes: four conserved hypotheticals and a probable heterodisulfide reductase A subunit. There are only seven clusters with more than four *A. fulgidus* homologues in a close neighbourhood on the *D. psychrophila* genome. These clusters contain genes for (i) the biosynthesis of lysine and arginine; (ii) probable Ni/Fe hydrogenase subunits and accessory proteins; (iii) a sugar ABC transporter; (iv) probable methyl viologen-reducing hydrogenase subunits; (v) heterodisulfide reductase subunits; (vi) indolepyruvate:ferredoxin oxidoreductase subunits and a probable high-affinity branched-chain amino

acid ABC transporter; and (vii) one functionally diverse cluster.

With 3.66 Mbp (including the two plasmids), the size of the *D. psychrophila* genome surpasses that of *A. fulgidus* (2.18 Mbp, no plasmids) by more than two-thirds. The predicted proteins of *D. psychrophila* outnumber those of *A. fulgidus* by almost 700. With an average of 968 bp, the protein coding regions of *D. psychrophila* appear to be significantly longer than those of *A. fulgidus* (822 bp), although the coding density appears to be lower in *D. psychrophila* (86.6%, Table 1) than in *A. fulgidus* (92.2%). It is evident that these numbers depend at least in part on the applied annotation strategies, e.g. annotation/deletion of small hypothetical coding regions without any similarity to known proteins. Both organisms are relatively rich in gene families, with 21% of the predicted *D. psychrophila* proteins and 30% of the predicted *A. fulgidus* proteins belonging to such families. While the ATP-binding subunits of ABC transporters (about 40 members in each organism) and the histidine kinases/response regulators belong to the largest protein families identified in both organisms, only *A. fulgidus* contains a large family of acyl-CoA ligases. Inteins appear to be absent in both organisms. The G+C content of DNA in both organisms is rather similar: 46.8% in *D. psychrophila* versus 48.5% in *A. fulgidus*. This is unexpected when considering that their optimal growth temperatures differ by 73°C and their extreme growth temperatures by 97°C. Only for the first codon position is the G+C content of both organisms similar (*D. psychrophila* 52%, *A. fulgidus* 53%), whereas *D. psychrophila* has a higher G+C content at the second codon position (42% versus 37%) and a significantly lower G+C content at the third position (45% versus 58%).

Both genomes contain in one of their high G+C content (>53%) regions a cluster of genes that are involved in coenzyme biosynthesis (*D. psychrophila*, cobalamin; *A. fulgidus*, haem). In contrast, a cluster of ribosomal protein-encoding genes is present in one of two high-G+C islands in *A. fulgidus*, whereas it is localized in one of five low-G+C islands in *D. psychrophila*. The genome of *A. fulgidus* contains three copies of the typical archaeal SRSR element (short randomly spaced repeats), which is absent in *D. psychrophila*. Although *A. fulgidus* encodes only one unlinked rDNA operon, *D. psychrophila* has seven highly conserved and linked rDNA operons (Fig. 1). The gene for 7S RNA is present only in the *A. fulgidus* genome. *D. psychrophila* encodes more tRNA genes (64) than *A. fulgidus* (46). In *D. psychrophila*, none of these tRNAs contains an intron, whereas five are found in *A. fulgidus* tRNAs. Neither the genes for the selenocysteine system (*selABCD*) nor any putative genes with selenocysteine codons are present in *A. fulgidus*; however, this system appears to play a role in *D. psychrophila*. In contrast to *A. fulgidus*, genes coding for all 20 tRNA-synthetases, including glutamine and asparagine synthetase, are present in *D. psychrophila*. However, like *A. fulgidus*, it also contains the *gatABC* genes (encoding Glu-tRNA amidotransferase) for *in situ* transamidation of Glu to Gln. In addition to the class I Lys aminoacyl-tRNA synthetase found in *A. fulgidus*, *D. psychrophila* also possesses a class II enzyme.

Genes of numerous transporters for physiologically important ions, amino acids and peptides have been identified in both organisms. Transporters for the uptake of carbohydrates and lipids as an energy source are rather specific for *A. fulgidus*. Both organisms harbour a significant number of genes for two-component regulatory systems (*D. psychrophila* even more than *A. fulgidus*), which enable them to maintain complex regulatory and sensory systems, including that for chemotaxis (Che). The genes for the structural basis of motility (flagella) are also present in both genomes, as well as the suite of genes necessary for the synthesis of most vitamins and cofactors. The defence mechanism against alien DNA depends in both organisms on one single type I restriction–modification system; genes for type II restriction–modification systems have not been identified. As expected from the strikingly different habitats of the two organisms, *D. psychrophila* harbours a suite of cold shock proteins, which are missing in the genome of *A. fulgidus*.

Other genomes and gene families. More than one-third of the *D. psychrophila* genes for which we could not detect any homologue in the other genomes are located within eight stretches of 33–74 ORFs (Fig. 6). These stretches are allowed to contain a limited number of ORFs with homologues. The G+C content in four of these stretches is decreased by 2.2% to 3.8% compared with the surrounding area of the genome; however, no indications for specific functions in these regions can be described.

A preliminary comparison with data from the genome of *G. sulfurreducens* (Methe *et al.*, 2003) identified more than 100 clusters of two to 61 homologues with identical gene order. Most frequently, these clusters contain genes encoding transporters, two-component response regulators, ribosomal proteins, hydrogenases/dehydrogenases and proteins involved in flagellar, cell wall and lipopolysaccharide biosynthesis.

Conclusion

Coastal sediments off Svalbard represent the typical natural habitat of *D. psychrophila*. Nutrient and redox gradients established by microbial activities in such sediments can be disturbed by tidal activity, fluctuating organic input and bioturbation. In particular, bioturbation by worms and by walruses feeding on mussels causes intensive mixing

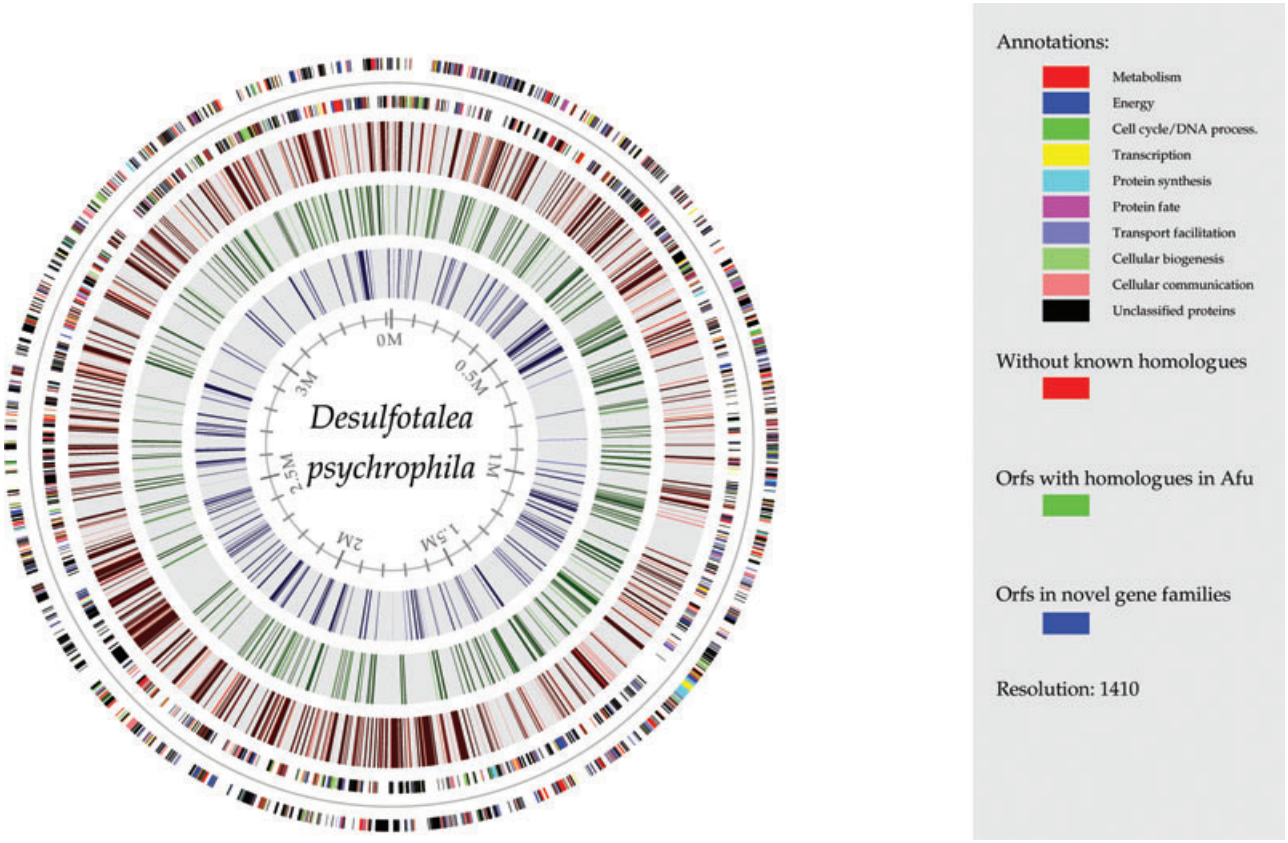

Fig. 6. Functional and comparative representation of the chromosome. Outer circle, predicted coding regions on the plus strand are colour coded by role categories. Second circle, predicted coding regions on the minus strand are colour coded by role categories (see above for plus strand). Third circle, hypothetical proteins without known homologues. Fourth circle, predicted proteins with a homologue in *A. fulgidus*. Fifth circle, members of novel gene families.

of the Svalbard sediment and oxygen penetration down to about 20 cm depth (Hop *et al*., 2002; Finke, 2003). To survive sudden oxygen exposure until aerobic members of the microbial community have consumed the oxygen, *D. psychrophila* is likely to depend on its own oxygen detoxification systems. The suite of identified regulatory proteins should enable *D. psychrophila* to adapt its metabolism effectively to changing environmental conditions, e.g. to temporary availability of amino acids.

Comparison of the genomes of *D. psychrophila* and *A. fulgidus* reveals only a few specifically shared features, in particular the genes involved in sulfate reduction and several islands of common hypothetical proteins. Instead, plenty of striking differences were recognized, e.g. in their metabolism. The differences caused by *D. psychrophila* being a bacterium and *A. fulgidus* being an archaeon, as well as the regions that are unique in both genomes, clearly outnumber the shared features and indicate that the fraction of the genomic inventory that is necessary for sulfate reduction is rather small. Considering the lack of c-type cytochrome-encoding genes, which are abundant in *Desulfovibrio* spp., the question of a general genome evolution among bacterial sulfate reducers arises. To refine the minimal genomic basis that is generally common to sulfate reducers, a comprehensive comparison considering the genomes from several diverse species is needed. In this context, the current sequencing projects on *D. vulgaris* (http://www.tigr.org), *Desulfovibrio desulfuricans* and *D. acetoxidans* (both http://www.jgi.doe.gov) will provide a promising opportunity.

Experimental procedures

Cultivation of cells and preparation of genomic DNA

Desulfotalea psychrophila strain LSv54 (DSM 12343) and *D. vulgaris* Hildenborough (DSM 644) were grown in bicarbonate-buffered, sulfide-reduced mineral media with lactate (20 mM) as the only source of organic carbon, as described previously (Widdel and Bak, 1992; Knoblauch *et al*., 1999). *D. psychrophila* was cultivated at 10°C and *D. vulgaris* at 37°C. Cells were harvested in the exponential growth phase, frozen in liquid nitrogen and stored at −80°C. Genomic DNA from *D. psychrophila* was prepared from about 3 g of cells (wet weight), essentially according to the method described by Rabus and colleagues (2002).

Genome sequencing, assembly and gap closure

Genomic DNA from *D. psychrophila* was sequenced using a conventional whole-genome shotgun strategy. Fragments of 1–2.5 kbp and 2.5–6 kbp were isolated after mechanical shearing, end-repaired and cloned via A-T cloning in *pGEM-T easy* vector (Promega). More than 20 000 plasmid DNAs were isolated on *BioRobots* 9600 (Qiagen). About 40 000 sequence reactions were generated from the plasmid templates using primers *pGEM1-abi* (forward) and *pGEM2-abi* (reverse) and either ABI BigDye™ Terminator chemistry (Applied Biosystems) or ET-Terminator chemistry (Amersham Biosciences). Sequence reactions were analysed on automated sequencers, ABI PRISM instrument models 377-96, 3100 and 3700, providing average useable read lengths of 677 nt. Sequence traces were first processed with PHRED for base calling (Ewing and Green, 1998) and data quality assessment (Ewing *et al*., 1998), then assembled with PHRAP (P. Green, University of Washington) and visualized and edited with CONSED (Gordon *et al*., 1998). In this process, we also used the program GAP4 from the Staden package (Staden *et al*., 2000). Primer walking on templates from the two plasmid libraries closed all but 233 gaps, for which no gap-spanning clones could be identified. These physical gaps were closed by direct sequencing on PCR fragments generated with the long-distance sequencer method described by Hagiwara and Harris (1996). After sequence finishing and polishing, the genome had a 6.4-fold sequence redundancy and an average PHRED score above 40. The final chromosome sequence was assembled from 34 078 sequence reads. The two plasmids were assembled from 799 and 77 sequence reads respectively.

Prediction and annotation of ORFs

Open reading frames were predicted by three independent ORF-finding programs: GLIMMER (Delcher *et al*., 1999), CRITICA (Badger and Olsen, 1999) and ORPHEUS (Frishman *et al*., 1998). A non-redundant list of ORFs was generated from the three sets of ORFs by parsing the results with PERL scripts as described by Glöckner and colleagues (2003). This set of non-redundant ORFs was the basis for the annotation performed with PEDANT-PRO (BioMax) (Frishman *et al*., 2001), which in addition to BLAST searches against a non-redundant database automatically collected results from analyses performed with BLOCKS, PROSITE, PFAM, 3D, SCOP and COG. All automatically generated assignments were reviewed and curated in two manual annotation rounds. Frameshifts were detected and corrected after the first round of manual annotation. Remaining frameshifts are considered to be authentic. ORFs overlapping with validated ORFs were deleted from the final ORF list when no functional assignment could be justified. The N-termini of the predicted proteins were adjusted to the best database match only in cases where sufficient reason for shortening of the N-terminus was available. Gene designations were taken from SWISSPROT (Boeckmann *et al*., 2003) or from respective references. Functional assignment of identified transporters was carried out using the transporter classification system (Saier, 2000; Busch and Saier, 2003).

Gene family identification and comparative genomics

All predicted proteins from the *D. psychrophila* genome as well as from all other published genomes in the SIMAP database (Mewes *et al*., 2004) were compared using FASTA (Pearson and Lipman, 1988) and the BLOSUM50 similarity matrix (Henikoff and Henikoff, 1993). *D. psychrophila* families as well as homologues from other genomes were identified by screening the SIMAP output with custom-made PERL scripts. The ratio between the optimal score obtained from a *D. psychrophila* protein and its homologue in the SIMAP database and the selfscore of the *D. psychrophila* protein was the guiding selection criterion. Ratios below 0.20 were ignored; in some (denoted) analyses, cut-off values greater than 0.20 were necessary to increase selectivity.

Histidine kinases were found with PSI-BLAST (Altschul *et al*., 1997) for the NtrB kinase domain against the non-redundant database. Using the last PSI-BLAST checkpoint file, a search was performed on the *D. psychrophila* proteome database and all hits ($e < 1$) were extracted. Those hits were then checked manually for histidine kinase motifs (H-N-D-F-G boxes). The genes remaining after manual inspection were then included in the alignment used for generating the phylogenetic tree in Fig. 2 (from alignment of only the histidine kinase domains). Response regulators were found by the same procedure as histidine kinases, using CheY as PSI-BLAST seed. A few proteins could not be classified (DP0338, DP2649 and DP2841) because of degenerate receiver domains. DNA repeats in the genome were analysed with REPUTER (Kurtz *et al*., 2001), and GENEWIZ (Pedersen *et al*., 2000) was used for whole-genome visualization.

Preparation of cell-free extracts and analysis of cytochromes

Frozen cell pellets were resuspended in 0.1 M Tris-HCl buffer (pH 7.5) to a concentration of $\approx$ 200 mg ml^{-1} (wet weight). Cell breakage was performed with the PlusOne sample grinding kit (Amersham Biosciences), according to the manufacturer's instructions. Cell-free extracts were obtained by centrifugation (12 000 r.p.m., Biofuge 13; Heraeus Instruments). The analysis of cytochromes was essentially carried out as described by others (Masau *et al*., 2001). Absorption of cytochromes in cell-free extracts was monitored at room temperature with a UV/Vis spectrometer (Lambda 20) equipped with a 1 cm light path. Extracts were diluted 10 times in 0.1 M Tris-HCl buffer (pH 7.5). When recording the absorption spectra, 0.1 M Tris-HCl buffer was used as reference. Extracts were oxidized by adding ammonium persulfate and, after measurement of absorption, they were completely reduced by adding Na-dithionite. Then the reduced spectrum was measured with the same extract. Difference spectra were obtained by manually subtracting the spectra of the oxidized extract from the spectra of the reduced extract.

Supporting information and database submission

The complete annotation data are available on the home page of the REGX Project (Real Environmental GenomiX),

http://www.regx.de. A BLAST server is available for public use. The nucleotide sequence of the chromosome and the two plasmids was submitted to EMBL under accession numbers CR522870 (chromosome) CR522871 (large plasmid) and CR522872 (small plasmid).

Acknowledgements

This work was supported by the German Federal Ministry of Education and Research, grant no. 03F0279A and E, and the Max Planck Society. The responsibility for the content of the publication is solely with the authors. We are indebted to F. Widdel and B. B. Jørgensen for general support of genomic work in Bremen and for inspiring discussions. We thank W. Ludwig for bioinformatics support. We acknowledge the technical assistance of D. Lange in cultivation of bacteria, J. Wulf in preparation of genomic DNA, and R. Müller in DNA sequencing and gap closure.

Glossary of expressions and abbreviations not explained elsewhere in the manuscript

Chargaff difference, deviation from Chargaff's second parity rule (balance of purines and pyrimidines); hypothetical proteins, predicted proteins without homologues in public databases; NUSA, transcription termination factor; PNP, polyribonucleotide nucleotidyltransferase; RBFA, ribosome-binding factor A; TCA cycle, tricarboxylic acid cycle; tmRNA, 10Sa RNA or SsrA that tags unfinished protein for proteolysis; *TPP*, thiamine pyrophosphate; TRAP-T system, Tripartite ATP-independent Periplasmic Transporter.

Note added in proof

A third genome of a (mesophilic) sulfate reducer has meanwhile been published by Heidelberg and colleagues (2004), *Nat Biotechnol* **5:** 554–559.

References

Alarcon-Chaidez, F., Sampath, J., Srinivas, P., and Vijayakumar, M.N. (1997) TN5252: a model for complex streptococcal conjugative transposons. *Adv Exp Med Biol* **418:** 1029–1032.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., 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.

Badger, J.H., and Olsen, G.J. (1999) CRITICA: coding region identification tool invoking comparative analysis. *Mol Biol Evol* **16:** 512–524.

Baumgarten, A., Redenius, I., Kranczoch, J., and Cypionka, H. (2001) Periplasmic oxygen reduction by *Desulfovibrio* species. *Arch Microbiol* **176:** 306–309.

Berchet, V., Thomas, T., Cavicchioli, R., Russell, N.J., and Gounot, A.M. (2000) Structural analysis of the elongation factor G protein from the low-temperature-adapted bacterium *Arthrobacter globiformis* SI55. *Extremophiles* **4:** 123–130.

Berks, B.C., Palmer, T., and Sargent, F. (2003) The TAT protein translocation pathway and its role in microbial physiology. *Adv Microb Physiol* **47:** 187–254.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL. *Nucleic Acids Res* **31:** 365–370.

Bokranz, M., Gutmann, M., Kortner, C., Kojro, E., Fahrenholz, F., Lauterbach, F., and Kröger, A. (1991) Cloning and nucleotide sequence of the structural genes encoding the formate dehydrogenase of *Wolinella succinogenes*. *Arch Microbiol* **156:** 119–128.

Bonch-Osmolovskaya, E.A., Sokolova, T.G., Kostrikina, N.A., and Zavarzin, G.A. (1990) *Desulfurella acetivorans* gen. nov. and sp. nov. – a new thermophilic sulfur-reducing eubacterium. *Arch Microbiol* **153:** 151–155.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., *et al.* (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273:** 1058–1073.

Busch, W., and Saier, M.H., Jr (2003) The IUBMB-endorsed transporter classification system. *Methods Mol Biol* **227:** 21–36.

Chabriere, E., Charon, M.H., Volbeda, A., Pieulle, L., Hatchikian, E.C., and Fontecilla-Camps, J.C. (1999) Crystal structure of the key anaerobic enzyme pyruvate: ferredoxin oxidoreductase, free and in complex with pyruvate. *Nature Struct Biol* **6:** 182–190.

Cypionka, H. (1989) Characterization of sulfate transport in *Desulfovibrio desulfuricans*. *Arch Microbiol* **152:** 237–243.

Cypionka, H. (2000) Oxygen respiration by *Desulfovibrio* species. *Annu Rev Microbiol* **54:** 827–848.

Dalluge, J.J., Hamamoto, T., Horikoshi, K., Morita, R.Y., Stetter, K.O., and McCloskey, J.A. (1997) Posttranscriptional modification of tRNA in psychrophilic bacteria. *J Bacteriol* **179:** 1918–1923.

Delcher, A.L., Harmon, D., Kasif, S., White, O., and Salzberg, S.L. (1999) Improved microbial gene identification with Glimmer. *Nucleic Acids Res* **27:** 4636–4641.

Eschemann, A., Kuhl, M., and Cypionka, H. (1999) Aerotaxis in *Desulfovibrio*. *Environ Microbiol* **1:** 489–494.

Ewing, B., and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8:** 186–194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998) Base calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8:** 175–185.

Fareleira, P., Santos, B.S., Antonio, C., Moradas-Ferreira, P., LeGall, J., Xavier, A.V., and Santos, H. (2003) Response of a strict anaerobe to oxygen: survival strategies in *Desulfovibrio gigas*. *Microbiology* **149:** 1513–1522.

Finke, N. (2003) The role of volatile fatty acids and hydrogen in the degradation of organic matter in marine sediments. PhD Thesis, University of Bremen, Bremen, Germany.

Franzmann, P.D., Springer, N., Ludwig, W., Conway de Macario, E., and Rohde, M. (1992) A methanogenic archaeon from ace lake, Antarctica: *Methanococcoides burtonii* sp. nov. *Syst Appl Microbiol* **15:** 573–581.

Franzmann, P.D., Liu, Y., Balkwill, D.L., Aldrich, H.C., Conway de Macario, E., and Boone, D.R. (1997) *Methanogenium frigidum* sp. nov., a psychrophilic, H_2-using methanogen from ace lake, Antarctica. *Int J Syst Bacteriol* **47:** 1068–1072.

Frishman, D., Mironov, A., Mewes, H.W., and Gelfand, M. (1998) Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. *Nucleic Acids Res* **26:** 2941–2947.

Frishman, D., Albermann, K., Hani, J., Heumann, K., Metanomski, A., Zollner, A., and Mewes, H.W. (2001) Functional and structural genomics using PEDANT. *Bioinformatics* **17:** 44–57.

Glöckner, F.O., Kube, M., Bauer, M., Telling, H., Lombardot, T., Ludwig, W., *et al.* (2003) Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc Natl Acad Sci USA* **100:** 8298–8303.

Gordon, D., Abajian, C., and Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res* **8:** 195–202.

Gounot, A.M., and Russell, N.J. (1999) Physiology of cold-adapted microorganisms. In *Cold-Adapted Organisms*. Margesin, R., and Schinner, F. (eds). New York, USA: Springer Verlag, pp. 33–51.

Green, G.N., Fang, H., Lin, R.J., Newton, G., Mather, M., Georgiou, C.D., and Gennis, R.B. (1988) The nucleotide sequence of the *cyd* locus encoding the two subunits of the cytochrome *d* terminal oxidase complex of *Escherichia coli*. *J Biol Chem* **263:** 13138–13143.

Hagiwara, K., and Harris, C.C. (1996) 'Long distance sequencer' method; a novel strategy for large DNA sequencing projects. *Nucleic Acids Res* **24:** 2460–2461.

Hatchikian, E.C., LeGall, J., and Bell, G.R. (1977) Significance of superoxide dismutase and catalase activities in the strict anaerobes, sulphate-reducing bacteria. In *Superoxide and Superoxide Dismutases*. Michelson, A.M., McCord, J.M., and Fridovitch, I. (eds). New York, USA: Academic Press, pp. 159–172.

Henikoff, S., and Henikoff, J.G. (1993) Performance evaluation of amino acid substitution matrices. *Proteins* **17:** 49–61.

Hop, H., Pearson, T., Hegseth, E.N., Kovacs, K.M., Wiencke, C., and Kwasniewski, S. (2002) The marine ecosystem of Kongsfjorden, Svalbard. *Polar Res* **21:** 167–208.

Jørgensen, B.B. (1982) Mineralization of organic matter in the sea-bed – the role of sulphate reduction. *Nature* **296:** 643–645.

Klenk, H.-P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E., Ketchum, K.A., *et al.* (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390:** 364–370.

Knoblauch, C., Sahm, K., and Jørgensen, B.B. (1999) Psychrophilic sulfate-reducing bacteria isolated from permanently cold Arctic marine sediments: description of *Desulfofrigus oceanense* gen. nov., sp. nov., *Desulfofrigus fragile* sp. nov., *Desulfofaba gelida* gen. nov., sp. nov., *Desulfotalea psychrophila* General nov., sp. nov. and *Desulfotalea arctica* sp. nov. *Int J Syst Bacteriol* **49:** 1631–1643.

Könneke, M., and Widdel, F. (2003) Effect of growth temperature on cellular fatty acids in sulphate-reducing bacteria. *Environ Microbiol* **5:** 1064–1070.

Koretke, K.K., Lupas, A.N., Warren, P.V., Rosenberg, M., and Brown, J.R. (2000) Evolution of two-component signal transduction. *Mol Biol Evol* **17:** 1956–1970.

Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* **29:** 4633–4642.

Lao, P.J., and Forsdyke, D.R. (2000) Thermophilic bacteria strictly obey Szybalski's transcription direction rule and politely purine-load RNAs with both adenine and guanine. *Genome Res* **10:** 1–20.

Lemos, R.S., Gomes, C.M., Santana, M., LeGall, J., Xavier, A.V., and Teixeira, M. (2001) The 'strict' anaerobe *Desulfovibrio gigas* contains a membrane-bound oxygen-reducing respiratory chain. *FEBS Lett* **496:** 40–43.

Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., *et al.* (2002) Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res* **30:** 242–244.

Lim, J., Thomas, T., and Cavicchioli, R. (2000) Low temperature regulated DEAD-box RNA helicase from the Antarctic archaeon, *Methanococcoides burtonii*. *J Mol Biol* **297:** 553–567.

Llobet-Brossa, E., Rabus, R., Böttcher, M.E., Könneke, M., Finke, N., Schramm, A., *et al.* (2002) Community structure and activity of sulfate-reducing bacteria in an intertidal surface sediment: a multi-methods approach. *Aquat Microb Ecol* **29:** 211–226.

Lumppio, H.L., Shenvi, N.V., Summers, A.O., Vourdouw, G., and Kurtz, D.M., Jr (2001) Rubrerythrin and rubredoxin oxidoreductase in *Desulfovibrio vulgaris*: a novel oxidative protection system. *J Bacteriol* **183:** 101–108.

McCloskey, J.A. Graham, D.E. Zhou, S. Crain, P.F. Ibba, M. Konisky, J., *et al.* (2001) Posttranscriptional modification in archaeal tRNAs: identities and phylogenetic relations of nucleotides from mesophilic and hyperthermophilic *Methanoccales*. *Nucleic Acids Res* **29:** 4699–4706.

Malki, S., Saimmaime, I., De Luca, G., Rousset, M., Dermoun, Z., and Belaich, J.P. (1995) Characterization of an operon encoding an NADP-reducing hydrogenase in *Desulfovibrio fructosovorans*. *J Bacteriol* **177:** 2628–2636.

Masau, R.J.Y., Oh, J.K., and Suzuki, I. (2001) Mechanism of oxidation of inorganic sulfur compounds by thiosulfate-grown *Thiobacillus thiooxidans*. *Can J Microbiol* **47:** 348–358.

Methe, B.A., Nelson, K.E., Eisen, J.A., Paulsen, I.T., Nelson, W., Heidelberg, J.F., *et al.* (2003) Genome of *Geobacter sulfurreducens*: metal reduction in subsurface environments. *Science* **302:** 1967–1969.

Mewes, H.W., Amid, C., Arnold, R., Frishman, D., Guldener, U., Mannhaupt, G., *et al.* (2004) MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res* **32 (**Database issue): D41–44.

Neretin, L.N., Schippers, A., Pernthaler, A., Hamann, K., Amann, R., and Jørgensen, B.B. (2003) Quantification of dissimilatory (bi)sulphite reductase gene expression in *Desulfobacterium autotrophicum* using real-time RT-PCR. *Environ Microbiol* **5:** 660–671.

Noon, K.R., Guymon, R., Crain, P.F., McCloskey, J.A., Thomm, M., Lim, J., and Cavicchioli, P. (2003) Influence of temperature on tRNA modification in Archaea: *Methanococcoides burtonii* (optimum growth temperature [Topt], 23 degrees C) and *Stetteria hydrogenophila* (Topt, 95 degrees C). *J Bacteriol* **185:** 5483–5490.

Pearson, W.R., and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85:** 244–2448.

Pedersen, A.G., Jensen, L.J., Brunak, S., Staerfeldt, H.-H., and Ussery, D.W. (2000) A DNA structural atlas for *Escherichia coli*. *J Mol Biol* **299:** 907–930.

Pfennig, N., and Biebl, H. (1976) *Desulfuromonas acetoxidans* gen. nov. and sp. nov., a new anaerobic, sulfur-reducing, acetate-oxidizing bacterium. *Arch Microbiol* **110:** 3–12.

Poole, R.K. (1994) Analysis of cytochromes. In *Chemical Methods in Prokaryotic Systematics*. Goodfellow, M., and O'Donnell, A.G. (eds). Chichester, UK: Wiley, pp. 311–344.

Rabus, R., Jack, D.L., Kelly, D.J., and Saier, M.H., Jr (1999a) TRAP transporters: an ancient family of extracytoplasmic solute-receptor-dependent secondary active transporters. *Microbiology* **145:** 3431–3445.

Rabus, R., Hansen, T., and Widdel, F. (1999b) Dissimilatory sulfate- and sulfur-reducing bacteria. In *The Prokaryotes – an Evolving Electronic Resource for the Microbiological Community*. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E. (eds). Heidelberg: Springer Science Online. [WWW document] URL http://www.prokaryotes.com.

Rabus, R., Kube, M., Beck, A., Widdel, F., and Reinhardt, R. (2002) Genes involved in the anaerobic degradation of ethylbenzene in a denitrifying bacterium, strain EbN1. *Arch Microbiol* **178:** 506–516.

Ravenschlag, K., Sahm, K., Knoblauch, C., Jørgensen, B.B., and Amann, R. (2000) Community structure, cellular rRNA content, and activity of sulfate-reducing bacteria in marine arctic sediments. *Appl Environ Microbiol* **66:** 3592–3602.

Russel, N.J., and Hamamoto, T. (1998) Psychrophiles. In *Extremophiles: Microbial Life in Extreme Environments*. Horikoshi, K., and Grant, W.D. (eds). New York, USA: John Wiley & Sons, pp. 25–45.

Sagemann, J., Jørgensen, B.B., and Greeff, O. (1998) Temperature dependence and rates of sulfate reduction in cold sediments of Svalbard, Arctic Ocean. *Geomicrobiol J* **15:** 85–100.

Sahm, K., Knoblauch, C., and Amann, R. (1999) Phylogenetic affiliation and quantification of psychrophilic sulfate-reducing isolates in marine arctic sediments. *Appl Environ Microbiol* **65:** 3976–3981.

Saier, M.H., Jr (2000) A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol Mol Biol Rev* **64:** 354–411.

Saunders, N.F.W., Thomas, T., Curmi, P.M.G., Mattick, J.S., Kuczek, E., Slade, R., *et al.* (2003) Mechanisms of thermal adaptation revealed from the genomes of the Antarctic archaea *Methanogenium frigidum* and *Methanococcoides burtonii*. *Genome Res* **64:** 15801588.

Sauter, M., Böhm, R., and Böck, A. (1992) Mutational analysis of the operon (*hyc*) determining hydrogenase 3 formation of *Escherichia coli*. *Mol Microbiol* **6:** 1523–1532.

Shaw, J.G., and Kelly, D.J. (1991) Binding protein dependent transport of C_4-dicarboxylates in *Rhodobacter capsulatus*. *Arch Microbiol* **155:** 466–472.

Sørensen, J., Christensen, D., and Jørgensen, B.B. (1981) Volatile fatty acids and hydrogen as substrates for sulfate-reducing bacteria in anaerobic marine sediment. *Appl Environ Microbiol* **42:** 5–11.

Staden, R., Beal, K.F., and Bonfield, J.K. (2000) The Staden Package, 1998. *Methods Mol Biol* **132:** 115–130.

Stetter, K.O. (1988) *Archaeoglobus fulgidus* gen. nov., sp. nov. a new taxon of extremely thermophilic archaebacteria. *Syst Appl Microbiol* **10:** 172–173.

Straub, K.L., and Schink, B. (2003) Evaluation of electron-shuttling compounds in microbial ferric iron reduction. *FEMS Microbiol Lett* **220:** 229–233.

Thomas, T., and Cavicchioli, R. (2003) Cold adaptation of archaeal elongation factor 2 (EF-2) proteins. *Curr Protein Pept Sci* **2:** 223–230.

Widdel, F., and Bak, F. (1992) Gram-negative mesophilic sulfate-reducing bacteria. In *The Prokaryotes*, 2nd edn. Balows, A., Trüper, H.G., Dworkin, M., Harder, W., and Schleifer, K.-H. (eds). New York: Springer-Verlag, pp. 583–624.

Williams, K.P., and Bartel, D.P. (1996) Phylogenetic analysis of tmRNA secondary structure. *RNA* **2:** 1306–1310.