ORIGINAL ARTICLE

The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: the role of genome evolution in cold adaptation

Michelle A Allen^{1,7}, Federico M Lauro^{1,7}, Timothy J Williams¹, Dominic Burg¹, Khawar S Siddiqui¹, Davide De Francisci¹, Kevin WY Chong¹, Oliver Pilak¹, Hwee H Chew¹, Matthew Z De Maere¹, Lily Ting¹, Marilyn Katrib¹, Charmaine Ng¹, Kevin R Sowers², Michael Y Galperin⁴, Iain J Anderson⁵, Natalia Ivanova⁵, Eileen Dalin⁵, Michele Martinez⁵, Alla Lapidus⁵, Loren Hauser⁶, Miriam Land⁶, Torsten Thomas^{1,3} and Ricardo Cavicchioli¹

¹School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, New South Wales, Australia; ²Center of Marine Biotechnology, University of Maryland Biotechnology Institute, Baltimore, MD, USA; ³Centre for Marine Bio-Innovation, The University of New South Wales, Sydney, New South Wales, Australia; ⁴NCBI, NLM, National Institutes of Health, Bethesda, MD, USA; ⁵Department of Energy Joint Genome Institute, Walnut Creek, CA, USA; ⁶Oak Ridge National Laboratory, Oak Ridge, TN, USA

Psychrophilic archaea are abundant and perform critical roles throughout the Earth's expansive cold biosphere. Here we report the first complete genome sequence for a psychrophilic methanogenic archaeon, *Methanococcoides burtonii*. The genome sequence was manually annotated including the use of a five-tiered evidence rating (ER) system that ranked annotations from ER1 (gene product experimentally characterized from the parent organism) to ER5 (hypothetical gene product) to provide a rapid means of assessing the certainty of gene function predictions. The genome is characterized by a higher level of aberrant sequence composition (51%) than any other archaeon. In comparison to hyper/thermophilic archaea, which are subject to selection of synonymous codon usage, *M. burtonii* has evolved cold adaptation through a genomic capacity to accommodate highly skewed amino-acid content, while retaining codon usage in common with its mesophilic *Methanosarcina* cousins. Polysaccharide biosynthesis genes comprise at least 3.3% of protein coding genes in the genome, and Cell wall, membrane, envelope biogenesis COG genes are overrepresented. Likewise, signal transduction (COG category T) genes are overrepresented and *M. burtonii* has a high 'IQ' (a measure of adaptive potential) compared to many methanogens. Numerous genes in these two overrepresented COG categories appear to have been acquired from ϵ - and δ -Proteobacteria, as do specific genes involved in central metabolism such as a novel B form of aconitase. Transposases also distinguish *M. burtonii* from other archaea, and their genomic characteristics indicate they have an important role in evolving the *M. burtonii* genome. Our study reveals a capacity for this model psychrophile to evolve through genome plasticity (including nucleotide skew, horizontal gene transfer and transposase activity) that enables adaptation to the cold, and to the biological and physical changes that have occurred over the last several thousand years as it adapted from a marine to an Antarctic lake environment.

The ISME Journal (2009) 3, 1012–1035; doi:10.1038/ismej.2009.45; published online 30 April 2009

Subject Category: evolutionary genetics

Keywords: archaea; cold adaptation; genome plasticity; *Methanococcoides burtonii*; psychrophile

Introduction

Most (~75%) of the Earth's biosphere is cold ($\leq 5^\circ\text{C}$), consisting of polar and alpine regions, the deep ocean, terrestrial and ocean subsurface, caves, the upper atmosphere, seasonally and artificially cold environments. Diverse cold-adapted (psychrophilic) microorganisms have evolved a capacity to

Correspondence: R Cavicchioli, School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, New South Wales 2052, Australia.

E-mail: R.Cavicchioli@unsw.edu.au

⁷These authors contributed equally to this work.

Received 22 January 2009; revised 30 March 2009; accepted 30 March 2009; published online 30 April 2009

proliferate in all these habitats. However, despite the enormous contribution cold environments make to the Earth's biosphere, relatively little is known about the genetic makeup and molecular mechanisms of adaptation of the resident microorganisms and how they drive the critical biogeochemical processes that help to maintain the planet in a habitable state (Cavicchioli, 2006; Murray and Grzymalski, 2007).

The majority of microbiological studies of cold environments have focused on psychrophilic bacteria, and genomic analyses have provided valuable insight into unique characteristics of sea ice, sediment and planktonic species (Rabus *et al.*, 2004; Medigue *et al.*, 2005; Methé *et al.*, 2005; Riley *et al.*, 2008). In addition to bacteria, it is now clearly recognized that the archaea are numerically, phylogenetically and functionally important members of cold aquatic and terrestrial ecosystems (Cavicchioli, 2006; Murray and Grzymalski, 2007). Their role in global biogeochemical cycles is diverse, including critical roles in the nitrification of soil and ocean waters mediated by ammonia oxidation (Konneke *et al.*, 2005; Leininger *et al.*, 2006; Cavicchioli *et al.*, 2007), and the cycling of simple carbon compounds by methanogenesis and reverse methanogenesis (anaerobic oxidation of methane) (Kruger *et al.*, 2005; Cavicchioli, 2006).

Cold anaerobic environments (for example, ocean and lake sediments, permafrost) harbor methanogens that have the capacity to contribute significantly to global carbon emissions through the production of methane as a green house gas. Although presently marine environments only contribute approximately 2% of the world's methane flux, this appears to be kept in check by the microbial communities that anaerobically oxidize methane (Moran *et al.*, 2008). The methane cycle is still poorly understood, and gaining an understanding of how these cellular processes occur in the cold will require fundamental knowledge of the molecular mechanisms of cold adaptation of psychrophilic methane-producing and methane-oxidizing archaea (Cavicchioli, 2006). Psychrophilic archaea have generally proven to be difficult to isolate. However, methanogens have been isolated from Antarctic lakes, marine sediment in Alaska and the Baltic Sea, a freshwater lake in Switzerland and Arctic permafrost (Franzmann *et al.*, 1992, 1997; Simankova *et al.*, 2001; Chong *et al.*, 2002; von Klein *et al.*, 2002; Singh *et al.*, 2005; Kendall *et al.*, 2007; Kotsyurbenko *et al.*, 2007; Morozova and Wagner, 2007).

Methanococcoides burtonii was the first formally characterized archaeal psychrophile and was isolated from cold (1–2 °C) methane-saturated, anaerobic bottom waters (25 m depth) of Ace Lake, Antarctica (Franzmann *et al.*, 1992). *M. burtonii* is a methylotrophic methanogen utilizing methylamines and methanol, but not H₂:CO₂ or acetate for growth. It is a eurypsychrophile with a relatively broad growth temperature range (–2.5 to 29 °C), in

contrast to *Methanogenium frigidum* that is a stenopsychrophile (0–18 °C) also isolated from Ace Lake (Franzmann *et al.*, 1992, 1997; Reid *et al.*, 2006). Unlike *M. frigidum*, which has proven to be very difficult to grow, *M. burtonii* is amenable to laboratory cultivation and has become a model psychrophilic archaeon for dissecting molecular mechanisms of cold adaptation (Cavicchioli, 2006). Studies on *M. burtonii* have examined enzyme structure/function, gene regulation, tRNA modification, membrane lipid composition and proteomics (reviewed in Cavicchioli, 2006), and intracellular solutes (Thomas *et al.*, 2001; Costa *et al.*, 2006). The draft genome sequences of *M. burtonii* and *M. frigidum* have also enabled comparative genomic analyses assessing the compositional and structural basis of thermal adaptation of proteins for archaea spanning growth temperatures ranging from 0 to 110 °C (Saunders *et al.*, 2003).

The genome sequence of *M. burtonii* was recently completed paving the way for determining for the first time, the genomic basis for growth and survival of a psychrophilic archaeon. Due to the value of the genome sequence and importance of *M. burtonii* to the scientific community, the genome sequence was exhaustively manually annotated to generate a high-quality genome sequence that was then used to evaluate the evolution and biology of *M. burtonii*.

Materials and methods

Genome sequencing, assembly, automated annotation

DNA was isolated from *M. burtonii* DSM 6242 grown at 23 °C. The genome of *M. burtonii* was sequenced at the Joint Genome Institute (JGI) using a combination of 3, 8 and 40 kb DNA libraries. All general aspects of library construction and sequencing can be found at the JGI's Web site (<http://www.jgi.doe.gov/>). The Phred/Phrap/Consed software package (www.phrap.com) was used to assemble all three libraries and to assess quality. Possible misassemblies were corrected, and gaps between contigs were closed by editing in Consed, custom primer walks or PCR amplification. The error rate of the completed genome sequence of *M. burtonii* is less than 1 in 50 000. Putative coding regions were identified using Critica, Generation and Glimmer, with automated annotation of proteins according to search results from the databases TIGRFam, PRIAM, Pfam, Smart, COGs, Swiss-Prot/TrEMBL and KEGG, as described for other JGI genomes (Medigue *et al.*, 2005; Chain *et al.*, 2006; Klotz *et al.*, 2006; Ivanova *et al.*, 2007). Additional curation involved the calculation of homolog, paralog and ortholog gene relationships, and generation of genome statistics. A round of manual curation was performed on the predicted genes of *M. burtonii* by IMG staff in which 346 changes were made. This corresponds to 13.9% of predicted genes. The start sites of 136 genes were modified, 62 genes were deleted and 61 new genes

were added. Also, 87 genes were identified as pseudogenes. This includes new pseudogenes that were added as well as previously existing genes that were converted to pseudogenes. IMG's curation staff also performed horizontal annotation involving manual investigation of proteins and biochemical pathways, with 'IMG term' annotations propagated across genomes according to strict homology criteria (http://imgweb.jgi-psf.org/w/doc/img_er_ann.pdf). In the *M. burtonii* genome, 635 proteins were assigned IMG terms by this process.

Manual annotation, evidence rating system and comparison genomes

The protein sequence for each gene was searched against the Swiss-Prot and Protein Data Bank (PDB) databases using BLAST to identify the most closely related, experimentally characterized homolog available in the literature. The curated part of Swiss-Prot database was selected because of the extent and quality of annotations associated with each protein sequence, and the PDB provides an archive of experimentally determined three-dimensional protein structures. BLAST matches were examined sequentially, searching for a match with direct experimental verification of the function. Matching proteins were not considered if they had no function listed, a function determined 'by similarity' or no reference cited. Experimental evidence that was considered acceptable to define function included papers detailing the expression and characterization of the protein, protein crystallography studies and mutation and complementation studies that defined function. Papers that only documented the nucleotide sequence, including genome sequences, were not considered to provide sufficient evidence of function, nor were unpublished protein crystal structures or the (published or unpublished) crystal structures of hypothetical proteins for which function was not known. In most cases papers were read in sufficient depth to establish if conclusions about function were justified. Choosing published experimental evidence as a prerequisite for defining function probably excluded some valid unpublished data; however, it was not possible to assess this data and it was excluded. Careful attention was also given to the recent literature to identify valid experimental data (for example, several methanogenesis-related proteins: Mbur_0808, Mbur_0811) that had not yet been updated in protein databases.

After the best experimentally characterized homolog had been determined, InterPro and Pfam domains were checked for their presence in the *M. burtonii* homolog to confirm that all identifiable functional domains were conserved. The graphical representation on the Swiss-Prot BLAST output was used for assessing the arrangement and extent of domain matches throughout the length of the query and matching proteins, thereby providing a rapid way to assess both the global (whole protein) and

local (domain and motif) similarities between the *M. burtonii* and the functionally characterized proteins.

An evidence rating (ER) system was developed to enable the confidence of functional assignments to be clearly displayed for each gene (Supplementary Information: manual annotation). ER1 indicates that the protein from *M. burtonii* had been experimentally characterized (a self-match); ER2, the most closely related functionally characterized homolog is not from *M. burtonii* but the BLAST alignments share $\geq 35\%$ sequence identity along the entire length of the protein; ER3, the most closely related functionally characterized homolog shares $< 35\%$ sequence identity along the length of the protein, but all required motifs/domains for function are present and complete; ER4, an experimentally characterized full-length homolog is not available but conserved protein motifs or domains can be identified; ER5 (hypothetical protein), no functionally characterized homolog can be found, and no characterized protein domains above the Pfam and InterProScan cutoff thresholds can be identified. The accession number for the *M. burtonii* genome is CP000300. The manually annotated genome is available through IMG/GEBA (img.jgi.doe.gov/geba) and will become available through the regular IMG system. Genome updates will also be made available through IMG. Comparative genomics (see methods below) was performed using local or external databases (for example, NCBI, IMG) with between 37 and 48 closed archaeal genome sequences. The draft sequence of *M. frigidum* was included in some comparisons (specified in the text).

COG scrambler and ALL scrambler

Two in-house programs, COG_scrambler.pl and ALL_scrambler.pl, were created for identifying significant differences in gene composition between samples. The COG_scrambler program compares all the COGs present in two data sets and calculates which COG categories are statistically over- or underrepresented in a data set by difference of medians analysis. The ALL_scrambler program performs an analogous process but reports which individual COGs are statistically over- or underrepresented. All programs were run with a confidence level setting of 0.99 (99% confidence that the difference is not due to chance) and 10 000 resampling replicates. The sample size was 6000 for the comparison of *M. burtonii* vs methanosarcinal genomes, and 10 000 for all other comparisons. The approach to determine the statistical significance was developed by FML and is similar to that described by Rodriguez-Brito *et al.* (2006).

Phylogenetic analyses

To visualize the phylogenetic relationships between proteins in the genome, we created an in-house PERL script 'phylo_profiler.pl' program. This program takes an input file (all *M. burtonii* proteins) and uses BLAST

to create a matrix of BLAST scores against selected completed genome sequences. The best BLAST bit score for each *M. burtonii* protein was then normalized both within and across genomes as previously described (Enault *et al.*, 2003). Pertinent data such as the COG ID, COG category, annotation and ER value were incorporated in the matrix to facilitate various searches and analyses. The matrix produced by phylo_profiler.pl was viewed using BioLayout Express 3D (Freeman *et al.*, 2007) with typical settings as follows: Graph Size vs Corr. Threshold = 95 (Nodes = 630, Edges = 1981, $R^2 = 0.7987$), Start temperature = 250, No. of iterations = 60, K-value modifier = 1.0, Iterations for Burst = 10, Minimum component size = 3. The protein sequences of transposases occurring multiple times (and those identified by proteomics to be expressed in the cell) were aligned using ClustalW (Thompson *et al.*, 1994) and used to build phylogenetic trees by neighbor-joining (Saitou and Nei, 1987) using the Poisson-corrected distance implemented in MEGA4 (Tamura *et al.*, 2007). The stability of the relationships was assessed by 1000 bootstrap replicates.

Identification of predicted horizontal gene transfer events

The program Alien Hunter (Vernikos and Parkhill, 2006) was used to identify genomic islands with altered nucleotide signatures. These islands were further divided in groups depending on their interpolated variable order motif (IVOM) scores and the open-reading frames (ORFs) contained within these regions were parsed using custom PERL scripts. The ORFs overlapping an island border were eliminated from further analyses. Each ORF was then phylogenetically assigned according to the best two BLAST hits with a minimum bit score of 50 against a customized version of the NCBI nonredundant (nr) database from which all *M. burtonii* proteins had been eliminated. The phylogenetic distribution was computed using MEGAN (Huson *et al.*, 2007). Codon usage was analyzed with CodonW (Peden, 2000). The genes on the primary axis of inertia in the correspondence analysis with values above the mean value plus one standard deviation were considered efficiently expressed; those below the mean value minus one standard deviation were considered not efficiently expressed. Amino-acid percentage composition and principal component analysis were performed using customized PERL scripts and R (Ihaka and Gentleman, 1996) as described by Saunders *et al.* (2003). An artificial genome was generated by removing the coding sequences of efficiently expressed ORFs from the FASTA sequence of the genome using the in-house PERL script 'reduce_genome.pl'.

Identification of hypothetical proteins unique to *M. burtonii*

The IMG Phylogenetic Profiler tool was used to search for genes in the *M. burtonii* genome, which

did not contain a homolog in any of the 814 standard reference genomes (40 completed archaeal genomes, 5 draft archaeal genomes, 19 completed eukaryal genomes, 21 draft eukaryal genomes, 486 completed bacterial genomes and 243 draft bacterial genomes). The cutoff settings for a homolog was minimum amino-acid identity of 30% and maximum *E*-value of 1×10^{-5} using the present/absent algorithm. The list of proteins obtained was then curated further by removing any proteins that were not annotated as a hypothetical protein (ER5). BLAST searches against the NCBI nr database and the CAMERA Global Ocean Survey combined assembly ORF peptides database were performed (2 May 2008), and any proteins with matches of *E*-value smaller than 1×10^{-5} were also discarded to obtain the final list of 117 proteins that are unique to *M. burtonii*. Conserved gene context analysis (Saunders *et al.*, 2005) was also used to infer possible functions for 28 proteins that were known from proteomics analysis to be expressed in the cell (Goodchild *et al.*, 2004a,b, 2005; Saunders *et al.*, 2005). Some of the unique expressed hypothetical proteins were found to be located in pairs (for example, Mbur_2063 and Mbur_2064) or larger clusters with unique hypothetical proteins that (to date) have not been detected in expression studies (for example, Mbur_0640 to Mbur_0643 and Mbur_0275 to Mbur_0278).

Results and discussion

M. burtonii genome overview

The completed genome sequence of *M. burtonii* DSM 6242 contains 257 5032 bp in a single circular chromosome (Figure 1). Manual annotation (see Materials and methods section) resulted in changes to the functional annotation of 1079 of the 2494 genes that had been identified by the auto-annotation process, and functions were assigned to 364 genes that previously had no functional prediction. A complete description of the manual annotation approach is provided in Supplementary Information: manual annotation.

To identify characteristics of the *M. burtonii* genome that distinguishes it from other archaea, we compared the three completed Methanosarcinales genomes, all 14 methanogen genomes or 39 archaeal genomes to the *M. burtonii* genome using COG_scrambler and ALL_scrambler (Table 1). In all three sets of comparisons, the Signal transduction [T] and Replication, recombination and repair [L] COG categories were significantly overrepresented, and the General Function prediction only [R] COG category was underrepresented in the *M. burtonii* genome. Individual COGs that were overrepresented included signal transduction histidine kinases, RecA-superfamily ATPases and CheY-like response regulators [T], and numerous transposases [L]. In addition, Cell wall, membrane, envelope biogenesis

Figure 1 Circular representation of the *Methanococcoides burtonii* genome. The circles show (outermost to innermost): (1) DNA coordinates (black); (2) genes on forward strand, color coded by COG categories; (3) genes on reverse strand, color coded by COG categories; (4) RNA genes, including tRNAs (green), rRNAs (red) and other RNAs (black); (5) genes involved in putative polysaccharide/capsule biosynthesis operons (red); (6) GC content; (7) GC skew.

[M] was overrepresented in all but the Methanosarcinales genomes.

Gene categories associated with cold adaptation

To identify functional gene categories characteristic of cold adaptation in archaea, we compared genomes from archaeal thermophiles and hyperthermophiles (4 methanogens, or 25 archaea in total) to *M. burtonii* using COG_scrambler and ALL_scrambler (Table 1). The psychrophilic genome of *M. burtonii* was overrepresented with Defense mechanisms [V] and Motility [N] and underrepresented with

Translation [J] and Nucleotide metabolism [F] COG categories compared to both the methanogen and total archaeal genome sets.

In the Defense mechanisms category, the individual COGs overrepresented in *M. burtonii* were mainly from Type-I restriction modification (RM) systems (COG0286, COG0610, COG0732 and COG4096), and ABC transporters (see below). The RM COGs were present at four locations on the *M. burtonii* genome (Supplementary Figure S1). One set of restriction endonuclease, methyltransferase and specificity genes was arranged in a typical operon-like RM cluster (Mbur_1841 to Mbur_1843).

Table 1 COG categories with a statistically significant difference in abundance

COG category	<i>M. burtonii</i> atypical ORFs ^a vs whole genome	Comparison with				
		<i>Methanosarcina</i> <i>spp.</i> (3)	<i>Methanogens</i>		<i>Archaea</i>	
			<i>All</i> (14)	<i>Thermophilic</i> (4)	<i>All</i> (39)	<i>Thermophilic</i> (25)
<i>Overrepresented in M. burtonii</i>						
Replication [L]		X	X	X	X	X
Signal transduction [T]	X	X	X	X	X	X
Translation [J]		X				
Coenzyme metabolism [H]		X			X	X
Cell wall [M]	X		X	X	X	X
Intracellular trafficking [U]			X	X	X	X
Motility [N]				X	X	X
Function unknown [S]						X
Defense mechanisms [V]				X		X
<i>Underrepresented in M. burtonii</i>						
General function only [R]		X	X	X	X	X
Inorganic ion metabolism [P]		X	X			
Defense mechanisms [V]		X				
Function unknown [S]	X		X	X		
Energy production [C]			X	X		X
Lipid metabolism [I]					X	X
Amino-acid metabolism [E]					X	X
Carbohydrate metabolism [G]					X	X
Translation [J]	X			X		X
Nucleotide metabolism [F]	X			X		X
Secondary metabolites [Q]						X

^a500 proteins of *M. burtonii* that were not phylogenetically congruent with the domain *Archaea* (combined 'other' and 'NA' categories from Figure 4b). X denotes that the abundance of the COG was statistically significant at the 99% confidence level. Gray shaded boxes highlight differences related to temperature (all methanogens vs hyper/thermophilic methanogens, and all archaea vs hyper/thermophilic archaea).

Another cluster was spaced over 10 genes (Mbur_1213 to Mbur_1222) with numerous genes for hypothetical proteins (ER5) interspersed. Two clusters (Mbur_0506 to Mbur_0509 and Mbur_0480 to Mbur_0484) contain a divergent AAA domain protein (ER4), in addition to the specificity COG0732, methyltransferase COG0286 and a variant helicase as the restriction endonuclease (Mbur_0509 and Mbur_0484). Divergent AAA domain proteins have been implicated in a wide variety of roles, including functioning as transcriptional regulators. The relatively large number of diverse RM systems may be indicative of the exposure of *M. burtonii* to high levels of foreign DNA (see Cold adaptation is associated with specific signatures of genome evolution section).

The other overrepresented COGs from the Defense mechanisms category were COG0577 and COG1136. Proteins belonging to these two COGs were located in six short clusters in the genome, in association with proteins from COG1361 and COG4591 (Figure 2a). COGs 0577 and 4591 both contain proteins that annotated as DUF214-containing protein (ER4). These proteins contain four transmembrane domains each, and one member of the DUF214 family has been characterized as an ABC-transporter permease involved in lipoprotein export (Narita *et al.*, 2003). Interestingly,

although COG1361 is designated as S-layer domain proteins, the hypothetical proteins found in COG1361 form a cluster distinct from the true S-layer proteins, and have similarity to the ABC-transport substrate-binding proteins observed elsewhere in the *M. burtonii* genome (Figure 2b). These proteins each possess a signal peptide motif and one transmembrane domain, similar to ABC-transporter substrate-binding proteins. These clusters of genes may therefore represent novel ABC transporters.

The overrepresentation of the putative novel ABC transporters contrasts with the lack of identifiable ABC transporters for peptides in *M. burtonii*. This is one of the major differences between the genomes of *M. burtonii* and the *Methanosarcina* spp., with the latter containing between 6 and 19 substrate binding proteins for peptide ABC transporters. The lack of peptide transporters is consistent with the inability of *M. burtonii* to use peptides for growth. *M. burtonii* also lacks drug exporters from subfamilies 2 and 3 of the major facilitator superfamily (TC 2.A.1) and proteins from the ACR3 family (TC 2.A.59) and the ArsB family (TC 2.A.45) that are found in *Methanosarcina* spp. The susceptibility that *M. burtonii* may face toward antimicrobial compounds and toxic metalloids may be alleviated by the novel ABC transporters, some of

Figure 2 Putative ABC transporters involved in cell defense. (a) COG0577, ABC-type antimicrobial peptide transport system, permease component, all annotated DUF214 protein (ER4); COG4591, ABC-type transport system, involved in lipoprotein release, permease component, all annotated DUF214 protein (ER4); COG1136, ABC-type antimicrobial peptide transport system, ATPase component, all annotated ABC-transporter ATPase (ER2); COG1361, S-layer domain, all annotated hypothetical protein (ER5). (b) Unrooted phylogenetic tree of substrate-binding proteins from ABC transporters constructed using the neighbor-joining method from a multiple alignment of amino-acid sequences of COG1361 hypothetical proteins (cyan), S-layer proteins (pink) and ABC-transporter substrate-binding proteins (black) from *M. burtonii*. The sequence alignments are in given Supplementary Figure S2.

which may also have a role in active transport leading to the formation of extracellular polymeric substances (EPS) that is characteristic of growth of *M. burtonii* in the cold (Reid *et al.*, 2006; see A large genomic commitment to polysaccharide biosynthesis section). It is also noteworthy that, unlike the *Methanosarcina* spp., *M. burtonii* has a coenzyme F₄₂₀-dependent sulfite reductase (Mbur_0619) similar to the one characterized in *Methanocaldococcus jannaschii* (Johnson and Mukhopadhyay, 2005) that allows it to tolerate sulfite in the environment. In the waters where *M. burtonii* was isolated, hydrogen sulfide concentrations reach very high levels (8 mM) (Rankin *et al.*, 1999), indicative of high sulfite reductase activity (also see Linking the evolution of the *M. burtonii* genome to its ecological niche section).

Cold adaptation is associated with specific signatures of genome evolution

All members of the *Methanosarcina* genus (for which *M. burtonii* is closely related) have been shown to possess extremely dynamic genomes (Maeder *et al.*, 2006), where genome rearrangements, acquisition of novel metabolic capabilities (Fournier and Gogarten, 2008) and capacity to express altered or foreign genes (Li *et al.*, 2007b) contribute to their capacity to colonize a wide variety of ecological niches. To examine the importance of horizontal gene transfer (HGT) to the *M. burtonii* genome, we performed gene-independent analysis of potential HGT using IVOM scores calculated from the program Alien Hunter (Vernikos and Parkhill, 2006). An extremely high proportion (51%) of the *M. burtonii* genome (higher than for any other completed archaeal genome) was identified as having aberrant sequence composition compared to the average of the genome (Figure 3a). However, across the 48 archaeal genomes, the number of base pairs implicated in HGT did not correlate with growth temperature (Figure 3a). For example, the mesophilic *Methanococcus maripaludis* C5 had very little predicted HGT, whereas the hyperthermophilic *Methanopyrus kandleri* had almost 30% predicted HGT.

Parametric methods used for the detection of HGT events (such as Alien Hunter program) are prone to false positive predictions (Azad and Lawrence, 2007). To assess the reason for the high level of atypical nucleotide composition and better assess HGT, we performed a range of further analyses. Putative HGT islands were grouped together based on their IVOM scores and their phylogeny determined. For *M. burtonii* 1042 ORFs were found in 125 genomic islands with altered IVOM scores, 1097 were found outside of these regions and 164 were on the border and discarded from further analysis. The islands were 1 313 583 nucleotides in length and represented 51.01% of the genome, with an average island length of 10 509 nucleotides. For the meso-

philic *Methanosarcina* genomes, the islands represented 40.89%, 40.79% and 35.36% for *M. mazei*, *M. acetivorans* and *M. barkeri*, respectively (Supplementary Table S1). Genes with the highest IVOM scores included a high proportion of unclassified genes and genes from unknown organisms (Figure 3b), and may have arisen through HGT events. The weakly atypical islands contained a high number of genes involved in translation, ribosomal structure and biogenesis (COG category J) that were phylogenetically congruent with the *Methanosarcina* genomes.

There are several lines of evidence that suggest that the ORFs within the weakly atypical islands have not arisen by HGT: (1) Genes encoding ribosomal proteins have high barriers to HGT (Sorek *et al.*, 2007). (2) The codon usage of the ORFs within the islands is very similar to that of the rest of the genome (Figure 3c). (3) As evidenced by the proportion of genes most closely related to the *Methanosarcinales* genomes (Figure 3b), the phylogenetic distribution of the ORFs is indicative of vertical rather than horizontal transmission. In addition, the presence of a high number of ribosomal proteins, which often display atypical nucleotide composition (Tsirigos and Rigoutsos, 2005), suggests alternative reasons for the altered IVOM scores. Initially we hypothesized that the high incidence of predicted HGT genomic islands in *M. burtonii* was caused by variations in codon usage. However, codon usage of the ORFs within the islands was similar to that of the whole genome (Figure 3c) and to the *Methanosarcinales* genomes (Figure 3d). Moreover, although the usage of specific synonymous codons has been correlated with the ability to grow at high temperatures (Lynn *et al.*, 2002), the codon usage of the psychrophilic archaea, *M. burtonii* and *M. frigidum*, was indistinguishable from that of the mesophilic *Methanosarcina* genomes (Figure 3d).

We subsequently reasoned that the alteration in IVOM scores might be caused by a high incidence of efficiently expressed genes whose protein products have been strongly selected for specific amino-acid usage. Principal component analysis of amino-acid composition of *M. burtonii* and the three *Methanosarcina* genomes shows a strong correlation (Spearman's correlation $P < 0.01$) between the PC1 loadings of efficiently expressed genes and the PC2 loadings previously observed (Saunders *et al.*, 2003) for whole genomes that was associated with temperature adaptation (Supplementary Figure S3). This correlation was not observed when the same analysis was performed with genes that are not efficiently expressed. We therefore conclude that (1) efficiently expressed genes in *M. burtonii* tend to have a stronger 'psychrophilic' component than those that are not efficiently expressed; (2) *M. burtonii* has unique amino-acid usage due to its psychrophilic lifestyle (for example, in comparison to the mesophilic *Methanosarcina* spp.).

As a further test, we generated an artificial *M. burtonii* genome where the coding regions of efficiently expressed ORFs were removed, reducing the genome size by 322 822 nucleotides (12.54%) and the number of genomic islands from 125 to 105. Of these only 63 were above the original IVOM score and encompass 782 651 nucleotides (34.75%) of the artificial genome. That is, the removal of the islands with efficiently expressed ORFs reduces the incidence of regions with atypical IVOM scores by twice as much as the number of nucleotides removed.

Our data highlight the need to carefully assess inferences about HGT when parametric methods are used. For *M. burtonii*, the exceptional amount of genome plasticity (that is, 51.01% altered IVOM scores) is primarily due to a strong selection for specific amino acids associated with cold adaptation present in efficiently expressed genes. This same coding bias is not seen in other genes. This may indicate that it is a recent evolutionary event and selection is yet to be manifested in other genes, or that low-temperature selection only has a significant effect on genes that need to be efficiently expressed in the cell (for example, high-abundance ribosomal proteins and other proteins critical to growth in the environment) (also see Linking the evolution of the *M. burtonii* genome to its ecological niche section). In comparison to hyper/thermophilic archaea that are subject to selection of synonymous codon usage (Lynn *et al.*, 2002), *M. burtonii* has evolved cold adaptation through a genomic capacity to accommodate highly skewed amino-acid content; an ability it has obtained while retaining codon usage in common with related mesophilic *Methanospirillum* spp.

Another feature of the codon usage plot (Figure 3d) is that the pattern of codon usage for archaea with growth temperatures $\leq 60^\circ\text{C}$ is similar and distinct from archaea with higher growth temperatures. A similar distinction at 60°C was previously noted for the relationship between G + C content of tRNA and growth temperature of archaea (Saunders *et al.*, 2003). In the case of tRNA in *M. burtonii*, high levels of dihydrouridine incorporation provide a mechanism for enhancing flexibility beyond that achievable through G + C content (Noon *et al.*, 2003; Saunders *et al.*, 2003). By analogy for *M. burtonii* proteins, amino-acid usage provides a mechanism for enhancing protein psychrophilicity beyond any requirements for specific codon usage.

To further assess possible HGT events, we analyzed the 500 ORFs that could not be reliably assigned to the domain *Archaea* (combined 'other' and 'NA' categories from Figure 3b). Compared to the whole genome these ORFs were overrepresented in COG categories M (Cell wall, membrane, envelope biogenesis) and T (Signal transduction mechanisms) and underrepresented in S (Function unknown), F (Nucleotide transport and metabolism) and J (Translation ribosomal structure and biogenesis) (Table 1). We note that the majority (44 of 70; 63%) of these ORFs were contained in islands with nonsignificant IVOM scores, illustrating that IVOM scores *per se* do not provide an unambiguous means of identifying HGT events in *M. burtonii*.

The underrepresentation of F and J is consistent with their central role in cell growth and survival and the inherent barriers for HGT of these genes. However, it was striking that category M was associated with HGT as this COG category is statistically overrepresented in the *M. burtonii* genome (Table 1) and the cell wall and lipid membrane have specifically been linked to cold adaptation (see A large genomic commitment to polysaccharide biosynthesis and Lipid biosynthesis sections). The 27 category M ORFs are mainly annotated as different types of glycosyltransferases with varying ERs (ER2–ER4), and a number of them appear to have been inherited from the ϵ -Proteobacteria. In category T, 39 ORFs had atypical phylogenetic assignments. These ORFs mainly belong to COG0642 (Signal transduction histidine kinase) and have ER3 or ER4 confidence ratings. Three that could be reliably assigned to a phylogenetic group (Mbur_1264, Mbur_2201 and Mbur_2108) clustered within the δ -Proteobacteria. This suggests that similar to category M, the overrepresentation of this Signal transduction category in *M. burtonii* (see *M. burtonii* genome overview section) is likely to have involved HGT (also see Signal transduction and adaptive potential section). In addition to these genes in categories M and T, HGT appears to have impacted on specific genes in central metabolism. Phylogenetic analysis of fructose-bisphosphate aldolase (Mbur_1969), serine *O*-acetyltransferase (Mbur_0414) and aconitase (Mbur_0316) shows that they are most closely related to counterparts from δ -Proteobacteria, and interestingly, *M. burtonii* is the first sequenced archaeon to have the B form of aconitase. Aspartate-semialdehyde dehydrogenase

Figure 3 Predictions of genome plasticity in *M. burtonii*. (a) Percentage of genome predicted to have undergone horizontal gene transfer by the Alien Hunter program. Bars are color coded according to the maximum growth temperature of the organism: purple, $< 30^\circ\text{C}$; blue, $40\text{--}49^\circ\text{C}$; green, $50\text{--}59^\circ\text{C}$; yellow, $60\text{--}79^\circ\text{C}$; orange, $80\text{--}99^\circ\text{C}$; red, 100°C and higher. (b) COG category classification (top pie charts), normalized expression (center bar graphs) and phylogenetic assignments (lower pie charts) of the open-reading frames (ORFs) contained in islands with different IVOM (interpolated variable order motif) scores. Norm (normal), $\text{IVOM} \leq 17.223$; low, $17.223 < \text{IVOM} < 26$; medium, $26 \leq \text{IVOM} < 40$; high, $40 \leq \text{IVOM}$. (c) Correspondence analysis of codon usage of ORFs in regions with different IVOM scores. Norm (normal), $\text{IVOM} \leq 17.223$; low, $17.223 < \text{IVOM} < 26$; medium, $26 \leq \text{IVOM} < 40$; high, $40 \leq \text{IVOM}$. (d) 106 446 ORFs from archaeal genomes represented in Figure 3a, and ORFs from the draft sequence of *M. frigidum* were subjected to correspondence analysis of codon usage (Lynn *et al.*, 2002). Growth temperatures: purple, $< 30^\circ\text{C}$; blue, $40\text{--}49^\circ\text{C}$; green, $50\text{--}59^\circ\text{C}$; yellow, $60\text{--}79^\circ\text{C}$; orange, $80\text{--}99^\circ\text{C}$; red, 100°C and higher.

(Mbur_0379) does not cluster with archaeal/eukaryal sequences, but does branch within bacterial enzymes.

The presence of multiple genes in *M. burtonii* originating from ϵ - and δ -Proteobacteria suggests that a close, possibly syntrophic, relationship has

Figure 3 Continued.

existed between *M. burtonii* and members of these proteobacterial groups. In this respect, it is noteworthy that several abundant phylotypes similar to known sulfate-reducing bacteria of the δ -Proteobacteria, and less abundant signatures of ϵ -Proteobacteria have been detected in sediment and anaerobic water samples of Ace Lake (Bowman *et al.*, 2000) (see Linking the evolution of the *M. burtonii* genome to its ecological niche section). Our analysis shows that HGT with ϵ - and δ -Proteobacteria has a specific role for acquiring genomic capabilities important for environmental adaptation.

Transposons are involved in genome evolution

Phylogenetic profiling was used to look for patterns of gene evolution. Each gene in the *M. burtonii* genome was compared using BLAST against a set of completed genomes to create a profile of the gene's phylogenetic relationships. Two comparisons were made: all completed archaeal genomes, and 431 completed bacterial and archaeal genomes. Genes were clustered according to their phylogenetic

occurrence and visualized using BioLayout (Supplementary Figure S4).

Five of the most tightly clustered phylogenetic groups were composed of transposases, with each group having specific phylogenetic occurrence (Table 2; Figure 4). The DDE superfamily of transposases (group 5) appeared to be restricted to members of the Methanosarcinales. Two of the 117 unique hypothetical proteins (Mbur_0556 and Mbur_0576; full list given in Supplementary Table S2) are located in a region adjacent to three cassettes each encoding genes for a DNA-directed DNA polymerase B domain protein and at least two conserved hypothetical proteins. The duplication evident in this region of the genome may also have arisen through transposition, as three transposons are also located nearby. The three cassettes are shared only between *M. burtonii* and *M. mazei*. Several other hypothetical genes appear to have been duplicated, possibly by transposition. Mbur_0169 and Mbur_0074 share 93% sequence identity and both are located adjacent to a Radical SAM family protein and an α/β -hydrolase domain protein, indicative of a duplicated cassette.

Table 2 Clusters formed from phylogenetic profiling using genes from *M. burtonii*

Cluster ID	Proteins represented in clusters	No.	Phylogenetic distribution	Notes
Transposons group 1	Transposase (ER4)	15	<i>Halobacterium</i> sp., <i>M. acetivorans</i> , <i>M. mazei</i> , <i>M. hungatei</i>	
Transposons group 2	Transposase (ER3) with 1 (ER4)	24	<i>M. acetivorans</i> , <i>T. volcanium</i> , <i>T. acidophilum</i>	
Transposons group 3	Transposase (ER4)	10	<i>M. acetivorans</i> , <i>M. barkeri</i> , <i>M. mazei</i> , <i>S. tokodaii</i> , <i>S. solfataricus</i>	
Transposons group 4	Transposase, mutator type (ER3)	4	<i>M. mazei</i> , <i>M. barkeri</i> , <i>S. solfataricus</i>	
Transposons group 5	IS4 Transposase, DDE superfamily (ER4)	8	<i>M. barkeri</i> , <i>M. mazei</i> , <i>M. acetivorans</i>	
<i>Methanosarcina</i> -like cluster	3 hypothetical proteins (ER5), 3 DUF1608 (ER4), 2 DUF1699 (ER4), metalloenzyme superfamily (ER4), sodium/solute symporter (ER3)	10	<i>M. thermophila</i> , <i>M. acetivorans</i> , <i>M. barkeri</i> , <i>M. mazei</i>	DUF1608 (IPR06457)—linked to an InterProDomain associated with archaeal S-layers (IPR06454). DUF1699—archaeal proteins with very conserved sequences
Methanogen cluster	Methanogenesis pathway genes (21), also acetylglutamate kinase, ornithine acetyl transferase, nitrogenase and associated proteins, many ER4 DUF proteins, one hypothetical protein (ER5)	46	All methanogens, <i>A. fulgidus</i>	
<i>M. mazei</i> -specific	8 hypothetical proteins, 3 proteins with DNA-directed DNA-polymerase B domains (ER4), sodium/glutamate symporter (ER3), bacterial Ig-like domain protein (ER4)	13	<i>M. mazei</i>	6 of the genes (3 hypothetical and 3 DNA-polymerase B domain) are in a region of novel hypothetical genes and have possibly duplicated/moved by transposase activity
<i>Methanosarcina</i> -specific	29 hypothetical proteins (ER5), pyrrollysyl-tRNA synthetase (ER2), tRNA-dihydrouridine synthase (ER2), fructose-1,6-bisphosphatase (ER2), restriction endonuclease (ER4), N6 DNA methylase (ER2), nicotinate phosphoribosyl transferase (ER2), 8 DUF or other domain proteins (ER4)	33	<i>M. acetivorans</i> , <i>M. barkeri</i> , <i>M. mazei</i>	
Largest cluster	Several ribosomal proteins, transposases, 3 thermosome subunit (chaperonin), GTP phosphohydrolase, translation elongation factor, RNase L inhibitor, a few DUF or other domain ER4 proteins, but largely all hypotheticals (89)	106	Fairly even across all archaea, moderate peak at <i>Nanoarchaeum equitans</i>	
Signal transduction	Signal transduction histidine kinases (14 were ER4, 4 were ER3)	18	<i>M. aeolicus</i> , <i>M. maripaludis</i> C5, <i>M. maripaludis</i> C7, <i>M. maripaludis</i> S2, <i>M. vanniellii</i> , <i>M. marisnigri</i> , <i>M. thermophila</i> , <i>M. acetivorans</i> , <i>M. mazei</i> , <i>M. barkeri</i> , <i>M. hugatei</i> , <i>N. pharaonis</i> , <i>M. thermoautotrophicum</i> , <i>A. fulgidus</i> , <i>H. marismortui</i> , <i>Halobacterium</i> NRC-1, <i>H. walsbyi</i>	
Chemotaxis	Chemotaxis proteins CheW, CheY, CheB, CheA, CheD and CheR	6	<i>A. fulgidus</i> , <i>H. marismortui</i> , <i>Halobacterium</i> NRC-1, <i>M. maripaludis</i> C5 C7 and S2, <i>M. vanniellii</i> , <i>M. marisnigri</i> , <i>M. acetivorans</i> , <i>M. mazei</i> , <i>M. hungatei</i> , <i>N. pharaonis</i> , <i>P. abyssi</i> , <i>P. horikoshii</i> , <i>T. kodakaraensis</i>	Form an operon Mbur_0357 to Mbur_0364

Table 2 Continued

Cluster ID	Proteins represented in clusters	No.	Phylogenetic distribution	Notes
Methanogenesis	Mono-, di- and trimethylamine corrinoid proteins	6	<i>M. smithii</i> , <i>M. aeolicus</i> , <i>M. maripaludis</i> C5 C7 and S2, <i>M. acetivorans</i> , <i>M. barkeri</i> , <i>M. mazei</i> , <i>M. stadtmanae</i> , <i>S. marinus</i> , <i>T. pendens</i>	<i>S. marinus</i> and <i>T. pendens</i> in this grouping of methanogenesis-related proteins
Methanogenesis 2	2 methanol corrinoid proteins and 2 conserved methanogen proteins with ferredoxin domains (one, Mbur_0809, found very close to the methanol corrinoid proteins Mbur_0812 and Mbur_0814)	4	<i>M. smithii</i> , <i>M. acetivorans</i> , <i>M. barkeri</i> , <i>M. mazei</i> , <i>M. stadtmanae</i>	
Methanogenesis 3	3 methanol-specific methyltransferase MtaA and one methylamine-specific methyltransferase MtbA	4	<i>M. smithii</i> , <i>M. maripaludis</i> C5 C7 and S2, <i>M. acetivorans</i> , <i>M. barkeri</i> , <i>M. mazei</i> , <i>M. stadtmanae</i>	One <i>mtbA</i> gene not clustering with these (Mbur_0957) may have a different phylogenetic occurrence
CRISPR region	7 CRISPR-associated proteins (ER4) and 1 chaperone (ER2)	8	<i>M. hungatei</i>	Appears to have come directly from <i>M. hungatei</i>
Hypotheticals	Hypothetical protein (ER5)	9	<i>M. acetivorans</i>	
Viruses	RNA-directed DNA polymerase (reverse transcriptase) (ER4)	4	<i>M. acetivorans</i> , <i>M. barkeri</i> , <i>M. mazei</i> , <i>M. hungatei</i>	Possibly replicating RNA from prophages
Serpins	Serine protease inhibitor (ER3)	3	<i>M. marisnigri</i> , <i>M. acetivorans</i> , <i>M. mazei</i> , <i>P. aerophilum</i> , <i>P. arsenaticum</i> , <i>P. calidifontis</i> , <i>P. islandicum</i> , <i>T. kodakaraensis</i>	
ER2, ER3 and ER4 all clustered together	O-Phosphoserine-tRNA (Cys) synthetase (ER2), Sep-tRNA:Cys-tRNA synthetase (ER3) and CBS-domain and DUF39-domain containing protein (ER4)	3	<i>A. fulgidus</i> , <i>M. thermoautotrophicum</i> , <i>M. aeolicus</i> , <i>M. jannaschii</i> , <i>M. maripaludis</i> C5 C7 and S2, <i>M. vannielii</i> , <i>M. labreanum</i> , <i>M. marisnigri</i> , <i>M. kandleri</i> , <i>M. thermophila</i> , <i>M. acetivorans</i> , <i>M. barkeri</i> , <i>M. mazei</i> , <i>M. hungatei</i> (all methanogens except <i>M. smithii</i> and <i>M. stadtmanae</i>)	Linkage of an ER4 protein with ER2 and ER3 proteins may infer a common function. These three proteins are linked by phyloprofile in each <i>Methanosarcina</i> genome
ER2 and ER4 proteins	DNA gyrase subunits A and B (ER2) and a protein of unknown function DUF1119 (ER4)	3	<i>A. fulgidus</i> , <i>H. marismortui</i> , <i>Halobacterium</i> NRC-1, <i>H. walsbyi</i> , <i>M. labreanum</i> , <i>M. marisnigri</i> , <i>M. thermophila</i> , <i>M. acetivorans</i> , <i>M. barkeri</i> , <i>M. mazei</i> , <i>M. hungatei</i> , <i>N. pharaonis</i> , <i>P. torridus</i> , <i>T. acidophilum</i> , <i>T. volcanium</i>	DUF protein is possibly associated with DNA-gyrase activity. The proteins are spaced out across the genome for <i>M. burtonii</i> , <i>Methanosarcina</i> and for <i>M. hungatei</i> . All three proteins are expressed in <i>M. burtonii</i> . The DUF1119 domain is a child of IPR006639::Peptidase A22, presenilin signal peptide, suggesting it may have a peptidase function

Due to approaches used for phylogenetic profiling, more distantly related phylogenetic groupings of transposons were detected (Table 2) compared to using BLAST against NCBI nr database to identify best matches (Figure 4). For example, group 2 included matches to *Thermoplasma* species, and groups 3 and 4 to Sulfolobales (Table 2). However, sequences from other methanogens always produced the best BLAST matches (Figure 4), indicating that *M. burtonii* transposases did not appear to have been transferred across distant phylogenetic boundaries. Additional signatures of genes shared between *M. burtonii* and other methanogens include an entire *M. hungatei* CRISPR region, and a cluster

of RNA-directed DNA polymerase genes (for example, possibly viral genes) common to the Methanosarcinales and *M. hungatei*.

Seven transposases have been found to be expressed in the cell during growth (Figure 4, triangles), indicating that transposases are active and not just markers of past genomic changes (Goodchild *et al.*, 2004b). The three largest groups (1, 2 and 3) were all represented in the expressed list. A further indicator of the activity of transposons is the presence of seven ORFs interrupted by the coding region of a transposase, with five of the ORFs known to be expressed (detected by proteomics) (Table 3). Fragments of five transposases are also present in the

Figure 4 Phylogeny of transposases in *M. burtonii*. Sixty-three transposases grouped in six major clusters: Group 1 is found in environmental genomic fragments from deep-sea sediments (GZfos19A5) and has multiple hits to *M. acetivorans* (3) and *M. mazei* (5). Group 2 is found in environmental genomic fragments from deep-sea sediments (GZfos27B6, GZfos9C4, GZfos26D6) and to *M. acetivorans* (1). Group 3 has multiple hits to *M. acetivorans* (11), *M. mazei* (1) and *M. barkeri* (5) and is probably an IS1-like element. Group 4 has hits to *M. mazei* (2) and is likely to belong to the mutator type. Group 5 contains the DDE superfamily IS4 transposases and has hits to *M. acetivorans* (1), *M. barkeri* (2) and *M. mazei* (1). Group 6 was not detected by phylogenomic profiling and was found in *M. acetivorans* (2) and in genomic fragments from deep-sea sediments (GZfos34A6). Triangles indicate transposases identified during proteomic analysis to be expressed in the cell. Best matches to *M. burtonii* transposases were found by performing BLAST against a customized version of the NCBI nonredundant (nr) database from which all *M. burtonii* proteins had been eliminated. The sequence alignments are in given Supplementary Figure S5.

Table 3 Genes interrupted by transposases

Gene number, function and ER	Transposon
Mbur_0914 ^a ; histidine kinase/PAS domain; ER4	Mbur_0915 (group 1)
Mbur_1066; NADPH-dependent FMN reductase; ER4	Mbur_1067 (group 4)
Mbur_1638 ^a ; hypothetical; ER5	Mbur_1639 (group 2)
Mbur_1664 ^a ; hypothetical; ER5	Mbur_1668 (group 3)
Mbur_2136 ^a ; histidine kinase/PAS domain; ER4	Mbur_2137 (group 2)
Mbur_2172; hypothetical; ER5	Mbur_2173 (group 2)
Mbur_2291 ^a ; dimethylamine methyltransferase; ER4	Mbur_2290 (group 3)

^aDetected by mass spectrometry in proteomics study to be expressed in the cell.

genome and are likely to be remnants of transposition events that are gradually being purged from the genome; Mbur_0774 and Mbur_0265 are fragments of a group 2 transposase; Mbur_0771 is a fragment of a group 1 transposase; Mbur_2251 is related to either Mbur_800 or Mbur_2016; and Mbur_0442 is a fragment of a group 6 transposase. The coding region of two of the five transposase fragments have themselves been interrupted by transposable elements; Mbur_0771 possibly by Mbur_0774 (also truncated), and Mbur_2251 by Mbur_2252.

Transposases [L] were one of the main COG categories distinguishing the psychrophilic *M. burtonii* from other archaea (see *M. burtonii* genome overview section). In association with the above genomic and proteomic data, it appears that transposases have an important role in evolving the genome of *M. burtonii*. Transposase inactivation has previously been linked to cold sensitivity in the deep sea, cold-adapted bacterium *Photobacterium profundum* SS9 (Lauro *et al.*, 2008), and transposases have been found to be one of the most overrepresented COG categories in metagenomic data of cold, deep (4000 m) ocean water samples (DeLong *et al.*, 2006).

Functional capacity is associated with distinct phylogenetic clusters of genes

The largest phylogenetic cluster (106 proteins) was common to all archaeal species and included ribosomal proteins, chaperonins, translation machinery and 89 hypothetical proteins (Table 2). Other clusters with phylogenetic distribution across many archaeal species include groups of signal transduction proteins, chemotaxis proteins, serine protease inhibitors, *O*-phosphoserine-tRNA (Cys) synthetase-linked proteins and DNA-gyrase-linked proteins. The latter two groups are particularly interesting as they link a well-characterized protein function (ER2) with a protein of unknown function in close phylogenetic association, even though the genes are physically separated on the genome. From this, it can be inferred that the DUF39-domain containing protein Mbur_0311 may be involved in production of Cys-tRNA^{Cys} by phosphoserine, and based on the child-parent relationship of the DUF1119 and peptidase A22 InterPro domains the

DUF1119 protein Mbur_1910 may have peptidase function involved in assisting the action of DNA gyrase. In six cases, poorly characterized genes (ER4 and ER5) clustered with well-characterized methanogenesis genes (only present in methanogenic archaea). The phylogenetic profile suggests that all the genes in the clusters are involved in methanogen-specific metabolic processes. A similar approach has been used to characterize hypothetical genes known to be expressed in *M. burtonii* (Saunders *et al.*, 2005).

General metabolism and transport

Central metabolism appears overall to be similar to other methanogens. Complete pathways for glycolysis (through a modified Embden–Meyerhof pathway) and gluconeogenesis are present. Although no glycogen phosphorylase could be found, other enzymes of starch metabolism are present suggesting that *M. burtonii* can store carbon as starch or glycogen. Glycogen serves as a polysaccharide storage reserve in many methanogens. The only pentose synthesis pathway present is the reverse ribulose monophosphate pathway, similar to other archaea.

M. burtonii possesses carbon monoxide dehydrogenase/acetyl-coenzyme A synthase (CODH/ACS; see Methanogenesis section) to produce acetyl-CoA from methyl-tetrahydrosarcinapterin and endogenously generated carbon dioxide. Although in a previous report *M. burtonii* was suggested to also be capable of carbon fixation by RubisCO as it possesses a type III ribulose-1,5-bisphosphate carboxylase/oxygenase (Mbur_2322) (Goodchild *et al.*, 2004b), no identifiable gene for phosphoribulokinase has been found in the completed genome. Further, *M. burtonii* does not grow autotrophically (Franzmann *et al.*, 1992), and there is no evidence for operation of either the reductive acetyl-CoA pathway or Calvin–Benson–Bassham cycle for carbon fixation. Archaeal (type III) RubisCO was recently demonstrated to be involved in AMP metabolism in the archaeon *Thermococcus kodakarensis* (Sato *et al.*, 2007). In this pathway, AMP phosphorylase (Mbur_0255) and ribose-1,5-bisphosphate isomerase (Mbur_1938) supply RubisCO substrate (ribulose-1,5-bisphosphate, RuBP) from

AMP and phosphate. RuBP can then be converted by RubisCO to two 3-phosphoglycerate (3-PGA) molecules, which can then enter central carbon metabolism. As homologous genes for all three enzymes are present, we suggest the *M. burtonii* RubisCO functions in the AMP phosphorylase pathway similar to *T. kodakaraensis*.

M. burtonii also has ADP-dependent (AMP-forming) sugar kinases that are used by many archaea in glycolysis. Given that 3-PGA can be used for ATP generation, these archaea may use the above pathway under anaerobic conditions to utilize AMP when energy levels are low and/or intracellular AMP levels are high (Sato *et al.*, 2007). Alternatively, AMP can be produced from 5-phosphoribosyl-1-pyrophosphate by adenine phosphoribosyltransferase (Mbur_1435), which also recycles the adenine generated by AMP phosphorylase.

Pyruvate synthase (Mbur_2155 to Mbur_2158) and pyruvate carboxylase (Mbur_2425 to Mbur_2426) provide a pathway for synthesis of oxaloacetate from acetyl-CoA. Synthesis of 2-oxoglutarate from oxaloacetate and acetyl-CoA is likely to take place through a truncated oxidative tricarboxylic acid (TCA) cycle. The genes involved in the reductive steps of the TCA cycle from oxaloacetate to 2-oxoglutarate are almost completely missing, with only a heterodimeric fumarase (Mbur_0250 to Mbur_0251) present. In the oxidative direction, citrate synthase was initially thought to be missing on the basis of the automated genome annotation. However, a citrate synthase with alternate stereospecificity (known as Re-citrate synthase) has been characterized in a *Clostridium* species, and an ortholog in *M. burtonii* was identified during the manual annotation process (Mbur_1075) (Li *et al.*, 2007a). Aconitase (Mbur_0316) is present (see Cold adaptation is associated with specific signatures of genome evolution section), and Mbur_1073 (isocitrate/isopropylmalate dehydrogenase) is a likely candidate to catalyze the remaining step from isocitrate to 2-oxoglutarate.

M. burtonii has biosynthetic pathways for pyrrolysine in addition to the 20 standard amino acids. Akin to *Methanosarcina* spp., *M. burtonii* has two pathways for cysteine synthesis: the tRNA-dependent pathway (Sauerwald *et al.*, 2005) and the O-acetylserine pathway. Cysteinyl-tRNA synthetase is absent, so the tRNA-dependent pathway is probably the only way to make cysteine for incorporation into proteins. Pyrrolysyl-tRNA synthetase (Mbur_2086) is located adjacent to the putative gene cassette for pyrrolysine synthesis (Mbur_2083 to Mbur_2085) (Longstaff *et al.*, 2007); thus pyrrolysine is synthesized before being attached to its tRNA. Evidence for the incorporation of pyrrolysine in methyltransferase enzymes has been provided from proteomic studies (Goodchild *et al.*, 2004a).

The way in which *M. burtonii* fixes nitrogen is unclear from the genome sequence. One *nifH* and one *nifD* homolog are present, but they cluster

phylogenetically with group IV nitrogenase-related genes that are not expected to be involved in nitrogen fixation (Raymond *et al.*, 2004). To test the ability of *M. burtonii* to fix nitrogen, we grew cells in defined media with N₂ gas as the sole source of nitrogen. Growth was observed after several generations and repeated passaging, although rates of growth and growth yield were lower than in media supplemented with an organic or inorganic nitrogen source (not shown). Preliminary analysis of *nifH* mRNA levels indicates that gene expression occurs even when cells are grown in complex media (not shown). These data indicate that *M. burtonii* fixes nitrogen through an as yet, undefined pathway.

Nitrate reductase and nitrate transporters cannot be identified in the genome sequence. The only member of the ammonium transporter family in *M. burtonii* (Mbur_0933) is situated adjacent to the gene for the regulatory PII protein (GlnB) (Mbur_0934). Mbur_0933 has 64% sequence identity to the characterized transporter Amt-3 from *Archaeoglobus fulgidus*. However, it is truncated at both the N terminus and the C terminus by approximately 88 amino acids so its functional status is uncertain. Given that ammonia concentrations increase with depth in Ace Lake (Butler *et al.*, 1988; Rankin *et al.*, 1999), it is possible that the ammonia requirements of *M. burtonii* can be met by inefficient facilitated diffusion of ammonium ion, and/or by the passive diffusion of ammonia. After uptake, *M. burtonii* may assimilate ammonia using, (1) the two-step glutamine synthetase (Mbur_1975) and glutamate synthase (Mbur_0092) pathway, and/or (2) glutamate dehydrogenase (Mbur_1973). 2-Oxoglutarate serves as the carbon skeleton for both pathways, and can be provided by a partial oxidative TCA cycle (see above). The glutamate synthase of *M. burtonii* appears to be a ferredoxin-dependent species. All three proteins are expressed in *M. burtonii* under laboratory growth conditions (Goodchild *et al.*, 2004a,b, 2005).

Methanogenesis

As an obligately methylotrophic methanogen, *M. burtonii* obtains energy from the oxidation of methyl groups to carbon dioxide and reduction to methane (Figure 5). Methyltransferases required to initiate metabolism by demethylation of methanol, monomethylamine, dimethylamine and trimethylamine and subsequent methylation of their respective corrinoïd-binding polypeptides were detected. Multiple copies of the each substrate-specific methyltransferase and corrinoïd protein were present. However, the second gene encoding dimethylamine methyltransferase (Mbur_2291) is disrupted by a putative transposon and is likely to be nonfunctional. Three copies of the methyl CoM methyltransferase for transfer of the methyl group from the methanol-specific corrinoïd, and two copies of the common

Figure 5 Methanogenesis and biomass production in *M. burtonii*. Color coding indicates proteins involved in methanogenesis from methanol (red), monomethylamine (MMA; green), dimethylamine (DMA; blue) and trimethylamine (TMA; purple). Abbreviations are as follows: CH₃OH, methanol; (CH₃)₂NH, DMA; (CH₃)₃N, TMA; CH₃-CP, methyl-corrinoid protein; CoM, coenzyme M; CoB, coenzyme B; MePh, oxidized methanophenazine; MePhH₂, reduced methanophenazine; F₄₂₀, coenzyme F₄₂₀; F₄₂₀H₂, reduced coenzyme F₄₂₀; CH₃-H₄SPT, methyl-tetrahydrosarcinapterin; CH₂=H₄SPT, methylene-tetrahydrosarcinapterin; CH=H₄SPT, methenyl-tetrahydrosarcinapterin; CHO-H₄SPT, formyl-tetrahydrosarcinapterin; CHO-MFR, formyl-methanofuran; Fd, ferredoxin; MtaB and MtaC, methanol methyltransferase and corrinoid protein; MttB and MttC, TMA methyltransferase and corrinoid protein; MtbB and MtbC, DMA methyltransferase and corrinoid protein; MtmB and MtmC, MMA methyltransferase and corrinoid protein; Mcr, methyl-CoM reductase; MtaA, methanol:CoM methylase; MtbA, methylamine:CoM methylase; Mcr, methyl-CoM reductase; Hdr, heterodisulfide reductase; Fpo, F₄₂₀H₂ dehydrogenase; Mtr, methyl-H₄SPT:CoM methyltransferase; Mer, methylene-H₄SPT reductase; Mtd, methylene-H₄SPT dehydrogenase; Mch, methenyl-H₄SPT cyclohydrolyase; Ftr, formyl-methanofuran:H₄SPT formyltransferase; Fmd or Fwd, formyl-methanofuran dehydrogenase; CODH/ACS, carbon monoxide dehydrogenase/acetyl-CoA synthase.

methyl CoM methyltransferase for methyl transfer from the methylated amine-specific corrinoid proteins were identified. This is similar to observations in the closely related methanosarcinal genomes,

which also have multiple copies of initiating methyltransferases and cognate corrinoid proteins (Galagan *et al.*, 2002; Deppenmeier *et al.*, 2002; Maeder *et al.*, 2006). However, in the methanosarcinal genomes only single copies of each CoM methyltransferase are present. Methyltransferases specific for methylthiols, which are found in some methanosarcinal species, were not present in *M. burtonii* (Tallant and Krzycki, 1997; Tallant *et al.*, 2001). During methylotrophic growth, methyl CoM is disproportionated 3:1 by oxidation to carbon dioxide and reduction to methane, respectively. All of the genes encoding for enzymes required for methyl reduction to methane and oxidation of methyl CoM through the reversal of the carbon dioxide reduction pathway are present in the *M. burtonii* genome (Figure 5).

Methanosarcina spp. are capable of growth by all four known methanogenic pathways: hydrogen reduction of carbon dioxide, hydrogen reduction of methylated compounds, fermentation of acetate and dismutation of methylotrophic compounds. The genome of *M. burtonii* provides the first direct insight into the minimal genes required for obligate methylotrophy by a methanogen. Growth with hydrogen requires three hydrogenases: Ech, which catalyzes ferredoxin reduction by hydrogen for subsequent reduction of carbon dioxide to the formyl level; Frh/Fre, which reduce coenzyme F₄₂₀ for reduction of methenyl and methylene groups; Vho, which catalyzes the reduction of the coenzyme CoM reducing electron carrier methanphenazine (Meuer *et al.*, 2002). Except for one protein with similarity to a coenzyme F₄₂₀-reducing hydrogenase β -subunit (Mbur_2261), genes encoding these enzymes were not detected in *M. burtonii*; consistent with the inability of this species to grow with hydrogen (Franzmann *et al.*, 1992). During growth, F₄₂₀ dehydrogenase (Fpo) could reduce methanphenazine from reduced F₄₂₀ generated during the oxidation steps of methyl and methylene groups in the methanogenic pathway. Reduced ferredoxin generated by oxidation of formylmethanofuran to carbon dioxide could be providing reducing potential for biosynthesis. *M. acetivorans*, which grows with acetate or methylotrophic substrates but does not grow with hydrogen, also lacks the Ech hydrogenase and although it has *frh* and *vho* genes they are not expressed (Guss *et al.*, 2005). This latter observation combined with the lack of hydrogenase genes in *M. burtonii* confirms that hydrogenases are not required for the methylotrophic pathway.

Hydrogen reduction of methylated compounds requires the specific methyltransferases and the reductive methanogenic pathway after methanopterin. However, this pathway also requires electrons from oxidation of hydrogenase to reduce methyl groups, which are not present in *M. burtonii*. Aceticlastic methanogenesis requires CODH/ACS to catalyze the methylation of methanopterin by dismutation of acetyl CoA. In contrast to two

gene copies in aceticlastic *Methanosarcina* spp., *M. burtonii* genome has only one copy of the CODH/ACS operon and lacks genes encoding acetate kinase and phosphotransacetylase for generating acetyl CoA from acetate. This is consistent with other nonaceticlastic methanogens that likely use CODH/ACS for carbon assimilation from carbon dioxide. Overall, the limited catabolic capabilities of *M. burtonii* are consistent with the small genome size relative to *Methanosarcina* spp.; a characteristic shared with the genomes of other methanogens with a single catabolic pathway.

In one respect the ability of *Methanosarcina* spp. to grow with all known methanogenic substrates provides them with the ability to adapt their catabolism to changes in substrate availability. In contrast *M. burtonii* is obligately methylotrophic and restricted to methanol and methylamines as exogenous carbon sources, which raises the question of how this is an advantage for the survival of this species as a specialist. One phenomenon observed for *Methanosarcina* spp. is increased hydrogen production on acetate and methylotrophic substrates when grown in the presence of a sulfate-reducing species (Phelps *et al.*, 1985). This also results in a reduction in methane yield per mole of substrate as more substrate is oxidized and reducing equivalents are diverted to hydrogen. This shift in the pathway would effectively reduce the energy yield for the methanogen and enable sulfate-reducing bacteria to utilize energy from the noncompetitive methylotrophic substrate through the reverse methanogenic pathway of the methanogen. The lack of hydrogenases in *M. burtonii* allows this specialist to utilize methylotrophic substrates without diversion of electrons to sulfate reducers and prevent the indirect utilization of these substrates by the latter. This may have provided a growth advantage to obligate methylotrophs such as *M. burtonii* in Ace Lake when the lake was (in the past) a sulfate-rich environment (see Linking the evolution of the *M. burtonii* genome to its ecological niche section).

Signal transduction and adaptive potential

The signal transduction system of *M. burtonii* is similar to that of its *Methanosarcina* spp. relatives, despite *M. burtonii* having a much smaller genome size. *M. burtonii* is motile by means of a single flagellum (Franzmann *et al.*, 1992) and encodes a chemotaxis system that includes a single methyl-accepting chemotaxis protein (Mbur_0356), its methylase and demethylase (Mbur_0360, Mbur_0399), chemotaxis histidine kinase CheA (Mbur_0361) and chemotaxis response regulator CheY (Mbur_0359). In addition to CheA, *M. burtonii* encodes 29 other two-component histidine kinases, of which 10 contain a (predicted) extracytoplasmic sensor domain and are probably involved in environmental sensing (Galperin, 2006). The remaining histidine kinases are intracellular; most of them

contain PAS domains and can be involved in sensing of the cellular level of oxygen, carbon monoxide, NO and other molecules. This could be particularly important for *M. burtonii* because this strict anaerobe might have to cope with the higher solubility of oxygen at low temperatures. Four of the *M. burtonii* histidine kinases have an N-terminal receiver (REC) domain similar to the proteins described recently in the haloarchaea (Galperin, 2006), and are likely to participate in complex phosphorelay signal transduction cascades.

Of the 14 response regulators encoded in the *M. burtonii* genome, eight consist of a single stand-alone REC domain and two more have an REC-PAS domain architecture. All these response regulators are likely involved in protein–protein interactions. Similar to other archaea, *M. burtonii* encodes no response regulators of the OmpR, NarL, NtrC, PrrA or LytR families. However, it encodes a response regulator (Mbur_0695) with a DNA-binding output domain of the GlpR type, which forms a typical operon-like structure with an environmental sensor histidine kinase Mbur_0694. This regulator is more abundant in cells grown at low temperature, and may form a temperature responsive regulatory system with the histidine kinase (Goodchild *et al.*, 2004a). Two more response regulators (Mbur_0878 and Mbur_2185) contain previously uncharacterized output domains, one unique for *M. burtonii* and the other found only in *Methanosarcina* spp. In addition, *M. burtonii* encodes an adenylate cyclase (Mbur_1935) and five predicted Ser/Thr protein kinases, one of the ABC1/AarF family, two of the RIO family, one unusual protein kinase and one Kae1-associated serine threonine kinase. Similar to other archaea, the role of cAMP, presumably synthesized by the adenylate cyclase, remains unknown, as are the cellular targets of the (predicted) protein kinases.

The large number of two-component regulatory systems in *M. burtonii* with limited similarity to known functional proteins may reflect a requirement for complex internal regulation (Ashby, 2006), and *M. burtonii* has a high 'IQ' (a measure of the adaptive potential of an organism) compared to many methanogens (Galperin, 2005). The majority of signal transduction genes (35 of 45 proteins) have a functional evidence rating of ER4 (indicating a lack of experimentally characterized full-length homologs). In addition, five ORFs listed as pseudogenes contain histidine kinase domains, and at least two of these have been interrupted by transposons and appear to be nonfunctional. Given the general lack of experimentation on two component regulatory systems in archaea and the specific importance of signal transduction systems to *M. burtonii* (see Cold adaptation is associated with specific signatures of genome evolution section), there is a compelling reason to experimentally characterize the sensing and response mechanisms of these phosphorelay networks.

Lipid biosynthesis

An important pathway associated with cold adaptation in *M. burtonii* is isoprenoid lipid biosynthesis (Nichols *et al.*, 2004). Synthesis of unsaturated lipids increases during growth at low temperature and is thought to maintain the fluidity, and hence functionality of the membrane in the cold. The digeranylgeranylgeranyl phosphate (DGGGP) synthase that was not present in the draft genome was identified as Mbur_1679 in the closed genome (Figure 6). Phosphomevalonate kinase was also not previously identified, and to date the gene has only

been identified in the Sulfolobales (Boucher, 2007). Recently *M. jannaschii* was hypothesized to use a modified mevalonate pathway, with the two steps required for conversion of mevalonate phosphate to isopentenyl diphosphate catalyzed by a novel (although not yet experimentally verified) mevalonate phosphate decarboxylase, followed by isopentenyl phosphate kinase (Grochowski *et al.*, 2006). The corresponding enzymes in *M. burtonii* are Mbur_2394 and Mbur_2396, respectively. This mevalonate pathway differs from the steps previously proposed for *M. burtonii*, which requires diphospho-

Figure 6 Lipid biosynthesis in *M. burtonii*. Figure based on analysis of the closed genome sequence (including new gene numbers), incorporating information from previous studies (Nichols *et al.*, 2004; Grochowski *et al.*, 2006; Boucher, 2007). Abbreviations are as follows: CoA, coenzyme A; HMG-CoA, 3-hydroxy-3-methylglutaryl-CoA; P, phosphate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GPP, geranyl diphosphate; FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate; GGGP, geranylgeranylgeranyl phosphate; DGGGP, digeranylgeranylgeranyl phosphate; CDP, cytidine diphosphoglycerol; G-1-P, glycerol-1-phosphate; DHAP, dihydroxyacetone phosphate.

sphomevalonate kinase and diphosphomevalonate decarboxylase (Nichols *et al.*, 2004). In the *M. burtonii* genome, Mbur_2394 and Mbur_2396 lie within a putative gene cluster associated with lipid biosynthesis, including mevalonate kinase (Mbur_2395), isopentenyl-diphosphate δ -isomerase (Mbur_2397), and a bifunctional enzyme comprising farnesyl pyrophosphate synthetase and geranyltransferase (Mbur_2399). Other important members of the lipid biosynthesis pathway include an experimentally characterized geranylgeranyl reductase responsible for conversion of DGGGP to archaetidic acid (Mbur_1077) (Murakami *et al.*, 2007). This enzyme is thought to have an important role in facilitating the regulation of unsaturation levels by performing selective saturation (similar to plants) (Nichols *et al.*, 2004). Other genes characterized on the basis of COG groupings involved in lipid formation include acetyl-CoA synthetases (3 proteins), acyl-CoA synthetase, pyruvate decarboxylase (α - and β -subunits located adjacent to each other in the genome), phosphatidyltransferases and cytidyltransferases.

A large genomic commitment to polysaccharide biosynthesis

Saccharide or polysaccharide moieties are important modifiers of the isoprenoid lipid membrane (Jahn *et al.*, 2004) and S-layer (Karcher *et al.*, 1993) of archaea, and can be secreted as EPS (Parolis *et al.*, 1996; Paramonov *et al.*, 1998). In *M. burtonii*, genes involved in polysaccharide biosynthesis comprise at least 3.3% of protein coding genes in the genome. In contrast to the 81 *M. burtonii* genes, the mesophilic *M. acetivorans* only contains approximately 30 polysaccharide biosynthesis genes representing 0.6% of its genome (Galagan *et al.*, 2002). Four operon-like clusters of polysaccharide biosynthesis genes (containing 39, 16, 11 and 10 genes) and five single glycosyl-transferase genes are distributed around the *M. burtonii* genome (Figure 1). In addition, five unique hypothetical proteins are encoded in proximity to the polysaccharide biosynthesis gene clusters. Five homologs (Mbur_0724, Mbur_0725, Mbur_0726, Mbur_1581 and Mbur_2023) of a glycosyl transferase (COG0438), and four homologs (Mbur_0727, Mbur_1593, Mbur_2028 and Mbur_2225) of a membrane protein involved in the export of O-antigen and teichoic acid (COG2244) are present in four main regions of the genome where they are adjacent to a number of other genes involved in polysaccharide biosynthesis, such as sugar- and *N*-acetylglucosamine epimerases and sugar dehydratases (Figure 1). The proteins represented by COG0438 and COG2244 are homologs to *P. profundum* SS9 proteins, PBPA2678 and PBPA2684, respectively. Mutation of the *P. profundum* SS9 genes produces a cold-sensitive phenotype (Lauro *et al.*, 2008; Ferguson *et al.* personal communication). EPS, including

polysaccharides appear to have a role in cell aggregation and biofilm formation of archaea at low temperatures (Reid *et al.*, 2006). Higher levels of EPS are produced by *M. burtonii* growing at low (compared to high) temperatures (Reid *et al.*, 2006), and approximately half of the *M. burtonii* polysaccharide biosynthesis genes are known to be expressed, having been shown to be abundant proteins in proteomic analyses (Goodchild *et al.*, 2004a,b, 2005; Saunders *et al.*, 2005). Collectively these data strongly suggest that polysaccharide biosynthesis has an important role in the cold adaptation of *M. burtonii*.

Linking the evolution of the M. burtonii genome to its ecological niche

Ace Lake is located on Long Peninsula in the Vestfold Hills, East Antarctica (Rankin *et al.*, 1999). During the early Holocene (13 000–9400 years ago), Ace Lake was an aerobic freshwater system, becoming a seasonally isolated marine basin (9400–9000 years ago) and subsequently open marine basin (9400–5700 years ago) (Rankin *et al.*, 1999; Coolen *et al.*, 2004; Cromer *et al.*, 2005). Approximately 5100 years ago Ace Lake became a permanently isolated saline lake, and developed meromixis, with an active methane cycle in existence for last ~3000 years. The microbiota in the lake is clearly marine derived. However, isotopic data indicate that all the water now present in the lake is of meteoric origin and that the lake has been mixed for considerable periods before its present stable meromixis. Nutrient input presently into the lake is very limited and it is a cold (average ~0 °C), oligotrophic system, although inorganic carbon levels are sufficient to lead to carbon dioxide efflux (Rankin *et al.*, 1999). The concentrations of most trace metals are higher than the ocean and are not limiting. Gradients of nutrients occur throughout the lake and concentrations can vary significantly; for example, manganese concentration varies from 78 to 1460 nM within 5 m in the anaerobic zone and this concentration is orders of magnitudes higher than the 5.1 nM of the ocean. The anoxic waters (12–25 m depth) support stable increasing gradients of salt (up to 4.3%), methane (saturated below 20 m, ~5 mM) and H₂S (up to 8 mM), and decreasing gradients of sulfate (essentially depleted below 19 m) (Rankin *et al.*, 1999).

Close relatives of *M. burtonii* have been identified in several cold ocean water locations in south and north polar marine waters and the deep sea, including *Methanococcoides alaskense* that has 99.8% 16S rRNA identity (Li *et al.*, 1999; Purdy *et al.*, 2003; Singh *et al.*, 2005; Cavicchioli, 2006). Cold, deep-sea environments in the Atlantic Ocean have been found rich in extracellular DNA, and may promote opportunities for DNA exchange (Dell'Anno and Danovaro, 2005). The adaptation of *M. burtonii* to the cold will have occurred over

millennia in the cold marine environment. However, the large environmental changes that have taken place in Ace Lake since the Holocene are likely to have provided strong selection pressure for ecotypes with genomic variation better suited to the new environment. In this regard it would be valuable to sequence the genome of *M. alaskense* to assess this. The specific capacity of *M. burtonii* to evolve through genome plasticity (including nucleotide skew, HGT and transposase activity) appears to have placed it in a strong position to not only adapt to the cold (for example, polysaccharide synthesis, lipid composition, amino-acid composition), but to the particular biotic and abiotic conditions that have changed in the lake throughout its recent several thousand year history (for example, central metabolism, novel ABC transporters, coenzyme F₄₂₀-dependent sulfite reductase).

Acknowledgements

The Australian contingent was supported by funding from the Australian Research Council. The work of IJA, NI, ED, MM, AL, LH and ML was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, Los Alamos National Laboratory under Contract No. DE-AC02-06NA25396, and Los Alamos National Laboratory under Contract No. DE-AC05-00OR22725. The work of KRS was supported by funding from the US Department of Energy's Office of Science, Biological and Environmental Research Program Grant No. DE-FG02-07ER64502 and the National Science Foundation, Division of Cellular and Bioscience Grant No. MCB0110762. MYG was supported by the NIH Intramural Research Program at the National Library of Medicine.

References

- Ashby MK. (2006). Distribution, structure and diversity of 'bacterial' genes encoding two-component proteins in the Euryarchaeota. *Archaea* **2**: 11–30.
- Azad RK, Lawrence JG. (2007). Detecting laterally transferred genes: use of entropic clustering methods and genome position. *Nucleic Acids Res* **35**: 4629–4639.
- Boucher Y. (2007). Lipids: biosynthesis, function and evolution. In: Cavicchioli R (ed). *Archaea: Molecular and Cellular Biology*. ASM Press: Washington, DC, pp 341–353.
- Bowman JP, Rea SM, McCammon SA, McMeekin TA. (2000). Diversity and community structure within anoxic sediments from marine salinity meromictic lakes and a coastal meromictic marine basin. Vestfold Hills, Eastern Antarctica. *Environ Microbiol* **2**: 227–237.
- Butler ECV, Burton HR, Smith JD. (1988). Iodine distribution in an Antarctic meromictic saline lake. *Hydrobiologia* **165**: 97–101.
- Cavicchioli R. (2006). Cold adapted Archaea. *Nat Rev Microbiol* **4**: 331–343.
- Cavicchioli R, DeMaere M, Thomas T. (2007). Metagenomic studies reveal the critical and wide-ranging ecological importance of uncultivated archaea: the role of ammonia oxidizers. *BioEssays* **29**: 11–14.
- Chain PSG, Denev VJ, Konstantinidis KT, Vergez LM, Agullo L, Reyes VL et al. (2006). *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proc Natl Acad Sci USA* **103**: 15280–15287.
- Chong SC, Liu Y, Cummins M, Valentine DL, Boone DR. (2002). *Methanogenium marinum* sp. nov., a H₂-using methanogen from Skan Bay, Alaska, and kinetics of H₂ utilization. *Antonie Van Leeuwenhoek* **81**: 263–270.
- Coolen MJL, Hopmans EC, Rijpstra WIC, Muyzer G, Schouten S, Volkman JK et al. (2004). Evolution of the methane cycle in Ace Lake (Antarctica) during the Holocene: response of methanogens and methanotrophs to environmental changes. *Org Geochem* **35**: 1151–1167.
- Costa J, Empadinhas N, Gonçalves L, Lamosa P, Santos H, da Costa MS. (2006). Characterization of the biosynthetic pathway of glucosylglycerate in the archaeon *Methanococcoides burtonii*. *J Bacteriol* **188**: 1022–1030.
- Cromer L, Gibson JAE, Swadling KTM, David RA. (2005). Faunal microfossils: indicators of Holocene ecological change in a saline Antarctic lake. *Palaeogeogr Palaeoclimatol Palaeoecol* **221**: 83–97.
- Dell'Anno A, Danovaro R. (2005). Extracellular DNA plays a key role in deep-sea ecosystem functioning. *Science* **309**: 2179.
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N-U et al. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.
- Deppenmeier U, Johann A, Hartsch T, Merkl R, Schmitz RA, Martinez-Arias R et al. (2002). The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between Bacteria and Archaea. *J Mol Microbiol Biotechnol* **4**: 453–461.
- Enault F, Suhre K, Abergel C, Poirot O, Claverie JM. (2003). Annotation of bacterial genomes using improved phylogenomic profiles. *Bioinformatics* **19**: i105–i107.
- Fournier GP, Gogarten JP. (2008). Evolution of acetoclastic methanogenesis in *Methanosarcina* via horizontal gene transfer from cellulolytic *Clostridia*. *J Bacteriol* **190**: 1124–1127.
- Franzmann PD, Liu Y, Balkwill DL, Aldrich HC, Conway de Macario E, Boone DR. (1997). *Methanogenium frigidum* sp. nov., a psychrophilic, H₂-using methanogen from Ace Lake, Antarctica. *Int J Syst Bacteriol* **47**: 1068–1072.
- Franzmann PD, Springer N, Ludwig W, Conway de Macario E, Rohde M. (1992). A methanogenic archaeon from Ace Lake, Antarctica: *Methanococcoides burtonii* sp. nov. *Syst. Appl Microbiol* **15**: 573–581.
- Freeman TC, Goldovsky L, Brosch M, van Dongen S, Mazière P, Grocock RJ et al. (2007). Construction, visualisation, and clustering of transcription networks from microarray expression data. *PLoS Comput Biol* **3**: 2032–2042.

- Galagan JE, Nusbaum C, Roy A, Endrizzi MG, Macdonald P, FitzHugh W *et al.* (2002). The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res* **12**: 532–542.
- Galperin MY. (2005). A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol* **5**: 35.
- Galperin MY. (2006). Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *J Bacteriol* **188**: 4169–4182.
- Goodchild A, Raftery M, Saunders NFW, Guilhaus M, Cavicchioli R. (2004a). Biology of the cold adapted archaeon, *Methanococcoides burtonii* determined by proteomics using liquid chromatography-tandem mass spectrometry. *J Proteome Res* **3**: 1164–1176.
- Goodchild A, Raftery M, Saunders NFW, Guilhaus M, Cavicchioli R. (2005). Cold adaptation of the Antarctic archaeon, *Methanococcoides burtonii* assessed by proteomics using ICAT. *J Proteome Res* **4**: 473–480.
- Goodchild A, Saunders NFW, Ertan H, Raftery M, Guilhaus M, Curmi PMG. (2004b). A proteomic determination of cold adaptation in the Antarctic archaeon, *Methanococcoides burtonii*. *Mol Microbiol* **53**: 309–321.
- Grochowski LL, Xu H, White RH. (2006). *Methanocaldococcus jannaschii* uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. *J Bacteriol* **188**: 3192–3198.
- Guss AM, Mukhopadhyay B, Zhang JK, Metcalf WW. (2005). Genetic analysis of mch mutants in two *Methanosarcina* species demonstrates multiple roles for the methanopterin-dependent C-1 oxidation/reduction pathway and differences in H₂ metabolism between closely related species. *Mol Microbiol* **55**: 1671–1680.
- Huson DH, Auch AF, Qi J, Schuster SC. (2007). MEGAN analysis of metagenomic data. *Genome Res* **17**: 377–386.
- Ihaka R, Gentleman R. (1996). R: a language for data analysis and graphics. *J Comp Graph Stat* **5**: 299–314.
- Ivanova NN, Mavromatis K, Chen I-MA, Markowitz VM, Kyrpide NC. (2007). Standard operating procedure for the annotations of genomes and metagenomes submitted to the integrated microbial genomes expert review (IMG-ER) system. Available: http://imgweb.jgi-psf.org/w/doc/img_er_ann.pdf.
- Jahn U, Summons RE, Sturt H, Grosjean E, Huber H. (2004). Composition of the lipids of *Nanoarchaeum equitans* and their origin from its host *Ignicoccus* sp. strain KIN4/1. *Arch Microbiol* **182**: 404–413.
- Johnson EF, Mukhopadhyay B. (2005). A new type of sulfite reductase, a novel coenzyme F₄₂₀-dependent enzyme, from the methanarchaeon *Methanocaldococcus jannaschii*. *J Biol Chem* **280**: 38776–38786.
- Karcher U, Schroder H, Haslinger E, Allmaier G, Schreiner R, Wieland F *et al.* (1993). Primary structure of the heterosaccharide of the surface glycoprotein of *Methanothermobacter feravidus*. *J Biol Chem* **268**: 26821–26826.
- Kendall MM, Wardlaw GD, Tang CF, Bonin AS, Liu Y, Valentine DA. (2007). Diversity of Archaea in marine sediments from Skan Bay, Alaska, including cultivated methanogens, and description of *Methanogenium boonei* sp. nov. *Appl Environ Microbiol* **73**: 407–414.
- Klotz MG, Arp DJ, Chain PSG, El-Sheikh AF, Hauser LJ, Hommes NG *et al.* (2006). Complete genome sequence of the marine, chemolithoautotrophic, ammonia-oxidizing bacterium *Nitrosococcus oceani* ATCC 19707. *Appl Environ Microbiol* **72**: 6299–6315.
- Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**: 543–546.
- Kotsyurbenko OR, Friedrich MW, Simankova MV, Nozhevnikova AN, Golyshin PN, Timmis KN *et al.* (2007). Shift from acetoclastic to H₂-dependent methanogenesis in a west Siberian peat bog at low pH values and isolation of an acidophilic *Methanobacterium* strain. *Appl Environ Microbiol* **73**: 2344–2348.
- Kruger M, Treude T, Wolters H, Nauhaus K, Boetius A. (2005). Microbial methane turnover in different marine habitats. *Palaeogeogr Palaeoclimatol Palaeoecol* **227**: 6–17.
- Lauro FM, Tran K, Vezzi A, Vitulo N, Valle G, Bartlett DG. (2008). Large-scale transposon mutagenesis of *Photobacterium profundum* SS9 reveals new genetic loci important for growth at low temperature and high pressure. *J Bacteriol* **190**: 1699–1709.
- Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW *et al.* (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**: 806–809.
- Li F, Hagemeyer CH, Seedorf H, Gottschalk G, Thauer RK. (2007a). Re-citrate synthase from *Clostridium kluyveri* is phylogenetically related to homocitrate synthase and isopropylmalate synthase rather than to Si-citrate synthase. *J Bacteriol* **189**: 4299–4304.
- Li L, Kato C, Horikoshi K. (1999). Microbial diversity in sediments collected from the deepest cold-seep area, the Japan Trench. *Mar Biotechnol* **1**: 391–400.
- Li L, Li Q, Rohlin L, Kim U, Salmon K, Rejtar T *et al.* (2007b). Quantitative proteomic and microarray analysis of the Archaeon *Methanosarcina acetivorans* grown with acetate versus methanol. *J Proteome Res* **6**: 759–771.
- Longstaff DG, Larue RC, Faust JE, Mahapatra A, Zhang L, Green-Church KB *et al.* (2007). A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc Natl Acad Sci USA* **104**: 1021–1026.
- Lynn DJ, Singer GAC, Hickey DA. (2002). Synonymous codon usage is subject to selection in thermophilic bacteria. *Nucleic Acids Res* **30**: 4272–4277.
- Maeder DL, Anderson I, Brettin TS, Bruce DC, Gilna P, Han CS *et al.* (2006). The *Methanosarcina barkeri* genome: comparative analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* reveals extensive rearrangement within methanosarcinal genomes. *J Bacteriol* **188**: 7922–7931.
- Medigue C, Krin E, Pascal G, Barbe V, Bernsel A, Bertin PN *et al.* (2005). Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudomonas haloplanktis* TAC125. *Genome Res* **15**: 1325–1335.
- Méthé BA, Nelson KE, Deming JW, Momen B, Melamud E, Zhang X *et al.* (2005). The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. *Proc Natl Acad Sci USA* **102**: 10913–10918.
- Meuer J, Kuettner HC, Zhang JK, Hedderich R, Metcalf WW. (2002). Genetic analysis of the archaeon *Methanosarcina barkeri* Fusaro reveals a central role

- for Ech hydrogenase and ferridoxin in methanogenesis and carbon fixation. *Proc Natl Acad Sci USA* **99**: 5632–5637.
- Moran JJ, House CH, Vrentas JM, Freeman KH. (2008). Methyl sulfide production by a novel carbon monoxide metabolism in *Methanosarcina acetivorans*. *Appl Environ Microbiol* **74**: 540–542.
- Morozova D, Wagner D. (2007). Stress response of methanogenic archaea from Siberian permafrost compared with methanogens from nonpermafrost habitats. *FEMS Microbiol Ecol* **61**: 16–25.
- Murakami M, Shibuya K, Nakayama T, Nishino T, Yoshimura T, Hemmi H. (2007). Geranylgeranyl reductase involved in the biosynthesis of archaeal membrane lipids in the hyperthermophilic archaeon *Archaeoglobus fulgidus*. *FEBS J* **274**: 805–814.
- Murray AE, Grzymski JJ. (2007). Diversity and genomics of Antarctic marine microorganisms. *Phil Trans R Soc B* **362**: 2259–2271.
- Narita S-I, Kanamaru K, Matsuyama S-I, Tokuda H. (2003). A mutation in the membrane subunit of an ABC transporter LolCDE complex causing outer membrane localization of lipoproteins against their inner membrane-specific signals. *Mol Microbiol* **49**: 167–177.
- Nichols DS, Miller MR, Davies NW, Goodchild A, Raftery M, Cavicchioli R. (2004). Cold adaptation in the Antarctic archaeon *Methanococcoides burtonii* involves membrane lipid unsaturation. *J Bacteriol* **186**: 8508–8515.
- Noon KR, Guymon R, Crain PF, McCloskey JA, Thomm M, Cavicchioli R. (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.
- Paramonov NA, Parolis LAS, Parolis H, Boan IF, Anton J, Rodriguez-Valera F. (1998). The structure of the exocellular polysaccharide produced by the archaeon *Haloferax gibbonsii* (ATCC 33959). *Carbohydr Res* **309**: 89–94.
- Parolis H, Parolis LAS, Boan IF, Rodriguez-Valera F, Widmalm G, Manca MC et al. (1996). The structure of the exopolysaccharide produced by the halophilic archaeon *Haloferax mediterranei* strain R4 (ATCC 33500). *Carbohydr Res* **295**: 147–156.
- Peden J. (2000). CodonW. Available: <http://codonw.sourceforge.net/>.
- Phelps TJ, Conrad R, Zeikus JG. (1985). Sulfate-dependent interspecies H₂ transfer between *Methanosarcina barkeri* and *Desulfovibrio vulgaris* during coculture metabolism of acetate or methanol. *Appl Environ Microbiol* **50**: 589–594.
- Purdy KJ, Nedwell DB, Embley TM. (2003). Analysis of the sulfate-reducing bacterial and methanogenic archaeal populations in contrasting Antarctic sediments. *Appl Environ Microbiol* **69**: 3181–3191.
- Rabus R, Ruepp A, Frickey T, Rattei T, Fartmann B, Stark M et al. (2004). The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. *Environ Microbiol* **6**: 887–902.
- Rankin LM, Gibson JAE, Franzmann PD, Burton HR. (1999). The chemical stratification and microbial communities of Ace Lake, Antarctica: a review of the characteristics of a marine-derived meromictic lake. *Polarforschung* **66**: 33–52.
- Raymond J, Siefert JL, Staples CR, Blankenship RE. (2004). The natural history of nitrogen fixation. *Mol Biol Evol* **21**: 541–554.
- Reid IN, Sparks WB, Lubnow S, McGrath M, Livio M, Valenti J et al. (2006). Terrestrial models for extra-terrestrial life: methanogens and halophiles at Martian temperatures. *Int J Astrobiol* **5**: 89–97.
- Riley M, Staley JT, Danchin A, Wang TZ, Brettin TS, Hauser LJ et al. (2008). Genomics of an extreme psychrophile, *Psychromonas ingrahamii*. *BMC Genomics* **9**: 210.
- Rodriguez-Brito B, Rohwer F, Edwards RA. (2006). An application of statistics to comparative metagenomics. *BMC Bioinformatics* **7**: 162.
- Saitou N, Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.
- Sato T, Atomi H, Imanaka T. (2007). Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science* **315**: 1003–1006.
- Sauerwald A, Zhu W, Major TA, Roy H, Palioura S, Jahn D et al. (2005). RNA-dependent cysteine biosynthesis in archaea. *Science* **307**: 1969–1972.
- Saunders NFW, Goodchild A, Raftery M, Guilhaus M, Curmi PMG, Cavicchioli R. (2005). Predicted roles for hypothetical proteins in the low-temperature expressed proteome of the Antarctic archaeon *Methanococcoides burtonii*. *J Proteome Res* **4**: 464–472.
- Saunders NFW, Thomas T, Curmi PMG, Mattick JS, 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* **13**: 1580–1588.
- Simankova MV, Parshina SN, Tourova TP, Kolganova TV, Zehnder AJ, Nozhevnikova AN. (2001). *Methanosarcina lacustris* sp. nov., a new psychrotolerant methanogenic archaeon from anoxic lake sediments. *Syst Appl Microbiol* **24**: 362–367.
- Singh N, Kendall MM, Liu Y, Boone DR. (2005). Isolation and characterization of methylotrophic methanogens from anoxic marine sediments in Skan Bay, Alaska: description of *Methanococcoides alakenese* sp. nov., and emended description of *Methanosarcina baltica*. *Int J Syst Evol Microbiol* **55**: 2531–2538.
- Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. (2007). Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**: 1449–1452.
- Tallant TC, Krzycki JA. (1997). Methylthiol:Coenzyme M methyltransferase from *Methanosarcina barkeri*, an enzyme of methanogenesis from dimethylsulfide and methylmercaptopropionate. *J Bacteriol* **179**: 6902–6911.
- Tallant TC, Paul L, Krzycki JA. (2001). The MtsA subunit of the methylthiol:coenzyme M methyltransferase of *Methanosarcina barkeri* catalyses both half-reactions of corrinoid-dependent dimethylsulfide: coenzyme M methyl transfer. *J Biol Chem* **276**: 4485–4493.
- Tamara K, Dudley J, Nei M, Kumar S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.
- Thomas T, Kumar N, Cavicchioli R. (2001). Effects of ribosomes and intracellular solutes on activities and stabilities of elongation factor 2 proteins from psychrotolerant and thermophilic methanogens. *J Bacteriol* **183**: 1974–1982.

- Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Tsirigos A, Rigoutsos I. (2005). A new computational method for the detection of horizontal gene transfer events. *Nucleic Acids Res* **33**: 922–933.
- Vernikos GS, Parkhill J. (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* **22**: 2196–2203.
- von Klein D, Arab H, Völker H, Thomm M. (2002). *Methanosarcina baltica* sp. nov., a novel methanogen isolated from the Gotland Deep of the Baltic Sea. *Extremophiles* **6**: 103–110.

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