

The Genome of *Methanosarcina mazei*: Evidence for Lateral Gene Transfer Between Bacteria and Archaea

Uwe Deppenmeier^{1,2}, Andre Johann¹, Thomas Hartsch^{1,4}, Rainer Merkl³, Ruth A. Schmitz², Rosa Martinez-Arias¹, Anke Henne¹, Arnim Wiezer¹, Sebastian Bäumer¹, Carsten Jacobi^{1,6}, Holger Brüggemann¹, Tanja Lienard², Andreas Christmann³, Mechthild Bömeke¹, Silke Steckel¹, Anamitra Bhattacharyya⁴, Athanasios Lykidis⁴, Ross Overbeek⁴, Hans-Peter Klenk^{1,7}, Robert P. Gunsalus⁵, Hans-Joachim Fritz^{1,3}, Gerhard Gottschalk^{1,2*}

¹ Göttingen Genomics Laboratory. ² Department of General Microbiology. ³ Department of Molecular Genetics and Preparative Molecular Biology. Institute of Microbiology and Genetics, Georg-August-University, Grisebachstr. 8, D-37077 Goettingen, Germany. ⁴ Integrated Genomics Inc., 2201 West Campbell Park Drive, Chicago, IL 60612, USA. ⁵ Department of Microbiology and Mol. Genetics, University of California, Los Angeles, CA 90095-1489, USA. Present Address. ⁶ BASF Aktiengesellschaft, ZHV - A 030, D-67056 Ludwigshafen, Germany. ⁷ Epidaurus Biotechnologie AG, Am Neuland 1, D-82347 Bernried, Germany

Abstract

The Archaeon *Methanosarcina mazei* and related species are of great ecological importance as they are the only organisms fermenting acetate, methylamines and methanol to methane, carbon dioxide and ammonia (in case of methylamines). Since acetate is the precursor of 60% of the methane produced on earth these organisms contribute significantly to the production of this greenhouse gas, e.g. in rice paddies. The 4,096,345 base pairs circular chromosome of *M. mazei* is more than twice as large as the genomes of the methanogenic Archaea currently completely sequenced (Bult *et al.*, 1996; Smith *et al.*, 1997). 3,371 open reading frames (ORFs) were identified. Based on currently available sequence data 376 of these ORFs are *Methanosarcina*-specific and 1,043 ORFs find their closest homologue in the bacterial domain. 544 of these ORFs reach significant similarity values only in the bacterial domain. They include 56 of the 102 transposases, and proteins involved in gluconeogenesis, proline biosynthesis, transport processes, DNA-repair, environmental sensing, gene regulation, and stress response. Striking examples are the occurrence of the bacterial GroEL/GroES chaperone

system and the presence of tetrahydrofolate-dependent enzymes. These findings might indicate that lateral gene transfer has played an important evolutionary role in forging the physiology of this metabolically versatile methanogen.

Introduction

Methanosarcina species are obligate anaerobic Archaea which are indispensable members of anaerobic food chains (Zinder, 1993). The substrate spectrum of species such as *M. barkeri* and *M. mazei* includes H₂+CO₂, acetate, methanol and methylamines. They coexist with fermentative bacteria, which, for instance, produce acetate from various carbohydrates or trimethylamine from betaine or choline (Hippe *et al.*, 1979). So they are essential for closing the cycle of organic matter on earth. Consortia of *Methanosarcina* species and sulfate-reducing bacteria are thought to even accomplish anaerobic oxidation of methane present in gas-hydrate-rich sediments at the sea floor (Boetius *et al.*, 2000). *Methanosarcina mazei* strain Gö1 played an important role in the elucidation of the energy conserving reactions during methanogenesis. Electron transport processes coupled to proton translocation and subsequent ATP synthesis (Deppenmeier *et al.*, 1999) as well as sodium ion translocation coupled to a novel type of methyltransferase reaction were discovered in this species

Table 1. General features of the *M. mazei* genome*

Length of sequence (bp)		4,096,345
Coding region (%)		75.15 %
G+C (%)		41.50 %
Ribosomal RNAs:	23S	3 copies
	16S	3 copies
	5S	4 copies
tRNAs		58 [‡]
Total number of open reading frames		3,371
ORFs with assigned function		2,450
ORFs without assigned function		921
ORFs with no database match		149
ORFs unique to <i>Methanosarcina</i> species		376
ORFs with best blast score in bacteria		1,043 [¶]
ORFs of the genus <i>Methanosarcina</i> with significant blast scores only in bacteria (bacterial-like proteins)		544
IS elements		185
Transposases		102
Start Codons	ATG	2,372
	TTG	558
	GTG	441
Stop Codons	TAA	1,800
	TGA	1,445
	TAG	97

[‡] includes a putative amber suppressor tRNA for in-frame TAG codons

[¶] Proteins which find their closest homologue in the bacterial domain outside the Methanosarcinales using the ERGO and the genebank data base

* According to the ERGO database

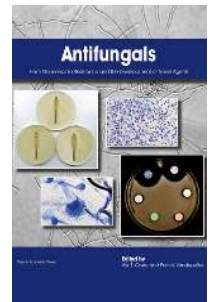
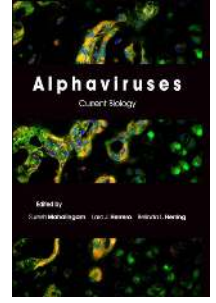
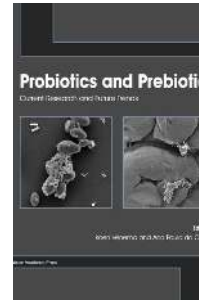
Supplementary information accompanies this paper and can be accessed on the following website: www.horizonpress.com/jmmb/supplementary

*For correspondence. Email ggottsc@gwdg.de; Tel. +49 551 393781; Fax. +49 551 393793.

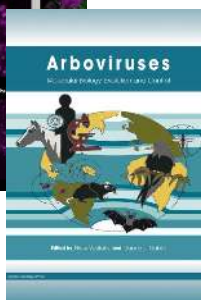
Further Reading

Caister Academic Press is a leading academic publisher of advanced texts in microbiology, molecular biology and medical research. Full details of all our publications at [caister.com](http://www.caister.com)

- **MALDI-TOF Mass Spectrometry in Microbiology**
Edited by: M Kostrzewa, S Schubert (2016)
www.caister.com/malditof
- **Aspergillus and Penicillium in the Post-genomic Era**
Edited by: RP Vries, IB Gelber, MR Andersen (2016)
www.caister.com/aspergillus2
- **The Bacteriocins: Current Knowledge and Future Prospects**
Edited by: RL Dorit, SM Roy, MA Riley (2016)
www.caister.com/bacteriocins
- **Omics in Plant Disease Resistance**
Edited by: V Bhaduria (2016)
www.caister.com/opdr
- **Acidophiles: Life in Extremely Acidic Environments**
Edited by: R Quatrini, DB Johnson (2016)
www.caister.com/acidophiles
- **Climate Change and Microbial Ecology: Current Research and Future Trends**
Edited by: J Marxsen (2016)
www.caister.com/climate
- **Biofilms in Bioremediation: Current Research and Emerging Technologies**
Edited by: G Lear (2016)
www.caister.com/biorem
- **Microalgae: Current Research and Applications**
Edited by: MN Tsaloglou (2016)
www.caister.com/microalgae
- **Gas Plasma Sterilization in Microbiology: Theory, Applications, Pitfalls and New Perspectives**
Edited by: H Shintani, A Sakudo (2016)
www.caister.com/gasplasma
- **Virus Evolution: Current Research and Future Directions**
Edited by: SC Weaver, M Denison, M Roossinck, et al. (2016)
www.caister.com/virusevol
- **Arboviruses: Molecular Biology, Evolution and Control**
Edited by: N Vasilakis, DJ Gubler (2016)
www.caister.com/arbo
- **Shigella: Molecular and Cellular Biology**
Edited by: WD Picking, WL Picking (2016)
www.caister.com/shigella
- **Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment**
Edited by: AM Romani, H Guasch, MD Balaguer (2016)
www.caister.com/aquaticbiofilms
- **Alphaviruses: Current Biology**
Edited by: S Mahalingam, L Herrero, B Herring (2016)
www.caister.com/alpha
- **Thermophilic Microorganisms**
Edited by: F Li (2015)
www.caister.com/thermophile



- **Flow Cytometry in Microbiology: Technology and Applications**
Edited by: MG Wilkinson (2015)
www.caister.com/flow
- **Probiotics and Prebiotics: Current Research and Future Trends**
Edited by: K Venema, AP Carmo (2015)
www.caister.com/probiotics
- **Epigenetics: Current Research and Emerging Trends**
Edited by: BP Chadwick (2015)
www.caister.com/epigenetics2015
- **Corynebacterium glutamicum: From Systems Biology to Biotechnological Applications**
Edited by: A Burkovski (2015)
www.caister.com/cory2
- **Advanced Vaccine Research Methods for the Decade of Vaccines**
Edited by: F Bagnoli, R Rappuoli (2015)
www.caister.com/vaccines
- **Antifungals: From Genomics to Resistance and the Development of Novel Agents**
Edited by: AT Coste, P Vandeputte (2015)
www.caister.com/antifungals
- **Bacteria-Plant Interactions: Advanced Research and Future Trends**
Edited by: J Murillo, BA Vinatzer, RW Jackson, et al. (2015)
www.caister.com/bacteria-plant
- **Aeromonas**
Edited by: J Graf (2015)
www.caister.com/aeromonas
- **Antibiotics: Current Innovations and Future Trends**
Edited by: S Sánchez, AL Demain (2015)
www.caister.com/antibiotics
- **Leishmania: Current Biology and Control**
Edited by: S Adak, R Datta (2015)
www.caister.com/leish2
- **Acanthamoeba: Biology and Pathogenesis (2nd edition)**
Author: NA Khan (2015)
www.caister.com/acanthamoeba2
- **Microarrays: Current Technology, Innovations and Applications**
Edited by: Z He (2014)
www.caister.com/microarrays2
- **Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications**
Edited by: D Marco (2014)
www.caister.com/n2



Order from [caister.com/order](http://www.caister.com/order)

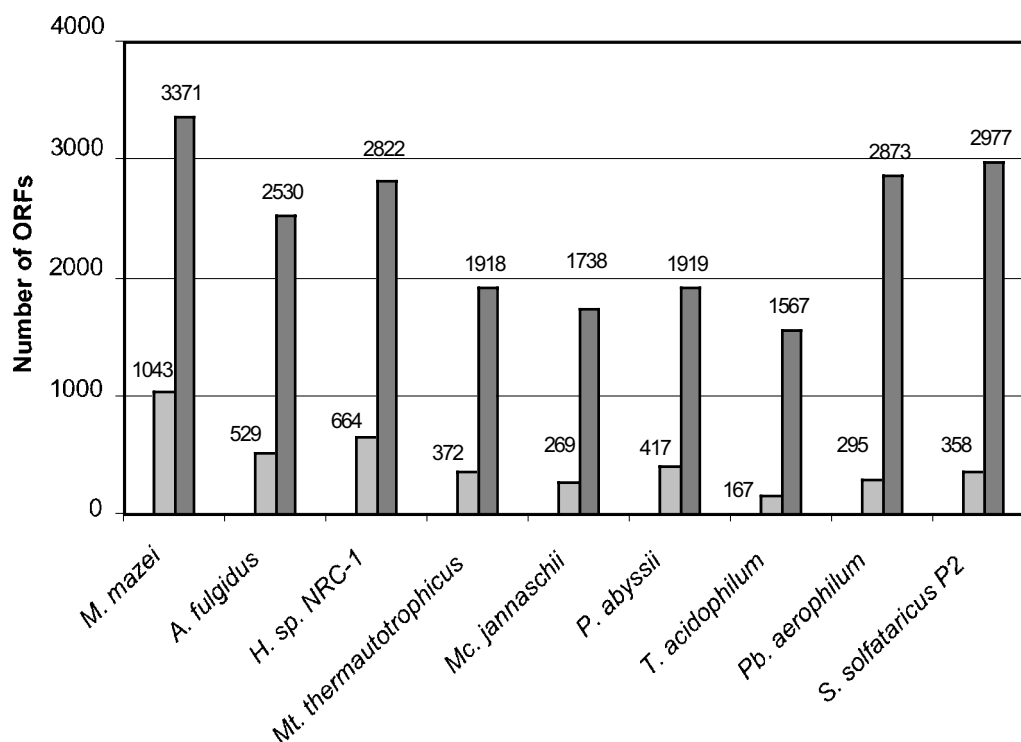


Figure 1. Number of ORFs that find their closest homologue in the bacterial domain outside the genus of the species listed. Values for *Sulfolobus solfataricus* were calculated from data of She *et al.* (2001), all other data were taken from the ERGO database (see Methods). Abbreviations: M., *Methanosarcina*; A., *Archaeoglobus*; H., *Halobacterium*; Mt., *Methanothermobacter*; Mc., *Methanococcus*; P., *Pyrococcus*; T., *Thermoplasma*; Pb., *Pyrobaculum*; S., *Sulfolobus*. Dark grey columns represent the total number of ORFs; light grey columns indicate the number of ORFs that find their closest homologue in the bacterial domain.

(Gottschalk and Thauer, 2001).

Results and Discussion

The genome of *M. mazei* is a single circular chromosome consisting of 4,096,345 base pairs. Some general features of the genome are summarized in Table 1. Of a total of 3,371 ORFs a function could be assigned to 2,450 ORFs: Only 75.15% of the sequences are coding. This is considerably lower than reported for other genomes. In *Vibrio cholerae* 88.6% of chromosome 1 and 86.3% of chromosome 2 are coding (Heidelberg *et al.*, 2000) and this value for *Thermoplasma acidophilum* (Rupp *et al.*, 2000) and *Mc. jannaschii* (Bult *et al.*, 1996) is 87%. Within the non-coding regions (up to 4,000 bp in size), 13 different stretches of direct repeats have been detected. These tracts range in length from 161 to 346 bp and consist of 11 different motifs (see supplementary Table 4 at www.horizonpress.com/jmmb/supplementary for further information). No long direct repeats have been observed applying the approach of Kurtz and Scheiermacher (1999). The G+C content of the intergenic region is 33%, while it is 44% in the coding region. The 185 IS elements range in length from 130 to 3,879 bp. One IS element is present in five copies, 1,514 bp long, three IS elements occur in three and five in two copies. The IS elements account for 4.11% of the *M. mazei* genome (supplementary information Table 3 at www.horizonpress.com/jmmb/supplementary).

The number of tRNAs (58 genes; supplementary information Table 2) exceeds that for *Methanococcus (Mc.) jannaschii* (37 genes) (Bult *et al.*, 1996) and *Methanothermobacter (Mt.) thermoautotrophicus* (formerly *Methanobacterium thermoautotrophicum* strain Δ H) (39 genes) (Smith *et al.*, 1997). This difference is attributed to greater than three tRNAs for ten amino acids in *M. mazei*. Recently, it was described that TAG is an in-frame codon in the monomethylamine methyltransferase of *M. barkeri* (Paul *et al.*, 2000). Evidence has been presented that the amber codon is read through during translation and that a lysine is present at the corresponding position. The genome of *M. mazei* contains seven methyltransferase genes of this type, two for monomethylamine, three for dimethylamine and two for trimethylamine. In addition, we identified in-frame TAG codons within 18 genes encoding transposases and within 4 ORFs of unknown function. Thus, of the 126 TAG codons, only 97 are present in the genome that may function as stops. We further identified a putative amber suppressor tRNA (see supplementary information Figure 2). It is currently not known, how the suppressor tRNA can distinguish between the in-frame amber codons and the TAG stop codons. However, it is tempting to speculate that the in-frame amber codon plus the suppressor tRNA represent a regulatory device, in which suppression of an amber codon acts as a yet unidentified regulatory mechanism for the initiation of methylamine degradation.

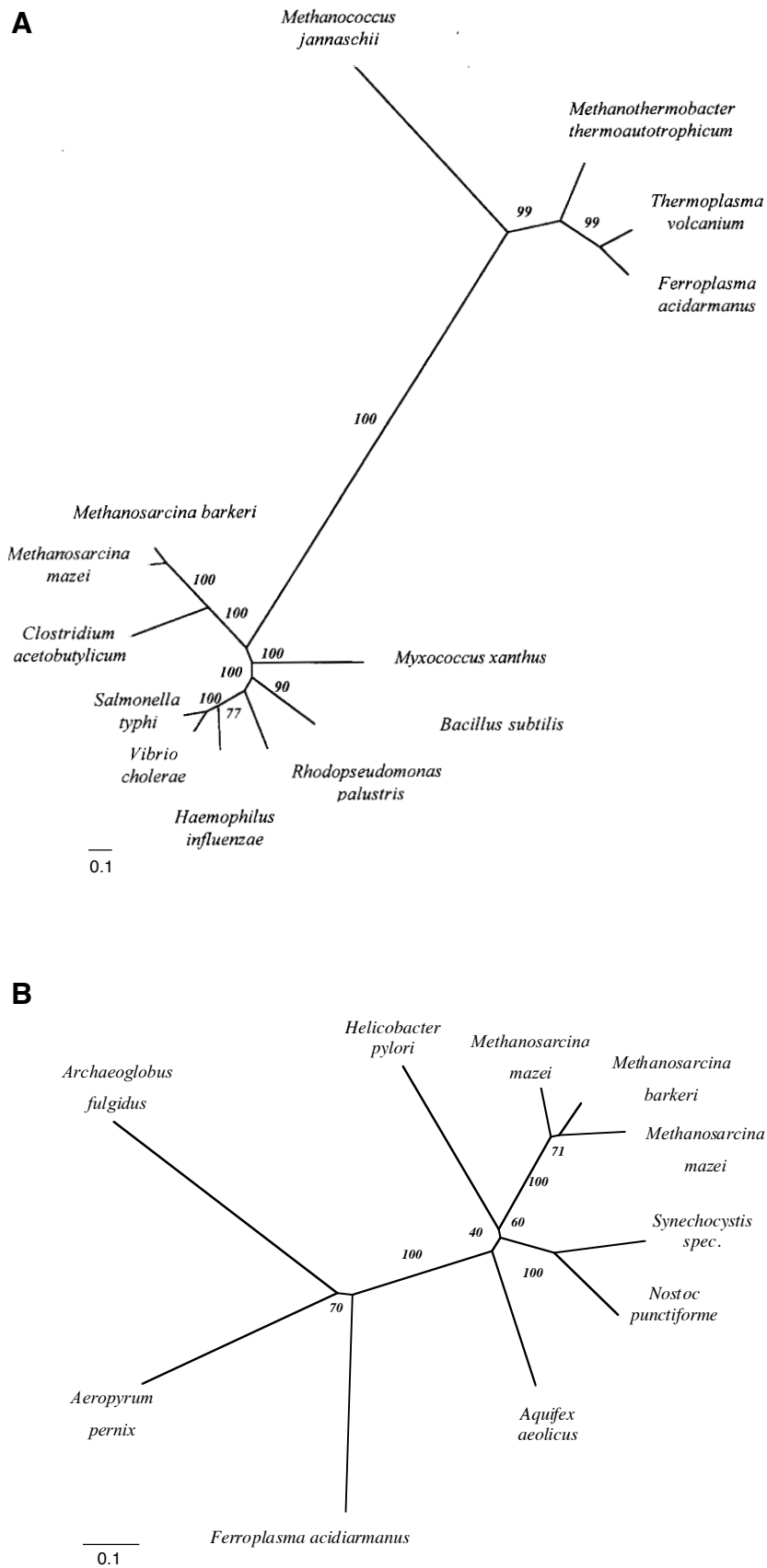


Figure 2. Neighbour-joining trees for two *M. mazei* enzymes: A) ATP-dependent Lon protease, B) Phosphoenol pyruvate synthase. Two main protein clusters are depicted in both cases: one clearly bacterial, and another archaeal. In both cases *M. mazei* proteins cluster within the bacterial group.

The genome size of *M. mazei* exceeds the one of thermophilic methanogens such as *Mc. jannaschii* by a factor of 2.5 (4.09 vs. 1.66 Mbp). This is accounted for by the already mentioned non-coding regions (1.01 Mbp) and by 3,371 ORFs as compared to 1,738 identified ORFs in *Mc. jannaschii* (Bult *et al.*, 1996). A striking feature of the *M. mazei* genome is that 1,043 of the 3,371 ORFs have their best blast scores outside the genus *Methanosarcina* in the bacterial domain. The number of ORFs and also the percentage is larger than in other archaeal genomes (Figure 1). Clearly, these numbers will change somewhat as more sequenced genomes of microorganisms become available. The analysis of these data reveals that ORFs with their best blast score in bacteria are generally more numerous in mesophilic (*M. mazei*, *Halobacterium* sp.) as compared to thermophilic or hyperthermophilic Archaea. They are also more abundant in species exhibiting a broader substrate spectrum, and as a consequence, with a more complex central metabolism (*M. mazei*, *Archaeoglobus fulgidus*, *Halobacterium* sp.). From these 1,043 ORFs on the *M. mazei* genome, 544 ORFs reach significant e-values only in the bacterial domain (referred to as bacterial-like proteins, for definition see Methods, see also supplementary information Table 5). Neighbour-Joining (NJ) trees (Saitou and Nei, 1987) were constructed for a subset of these proteins. In all 20 NJ trees two main clusters are clearly seen, one of bacterial and one of archaeal nature. In all cases the *M. mazei* (and *M. barkeri*) proteins cluster within the group of bacterial proteins (Figure 2, see also supplementary information Figure 3)

376 ORFs were classified as being unique to *Methanosarcina* species. Most of them are hypothetical proteins. Among these ORFs may also be those unknown ones for the synthesis of methanophenazine because this cofactor (Abken *et al.*, 1998) is apparently only present in methanogenic organisms utilizing methylamines and acetate (unpublished results). Only a few of these *Methanosarcina*-specific ORFs could be annotated: they include proteins of the methyltransferase systems for methanogenesis from methylated C1 substrates, components of the anaerobic respiratory chain, ABC transporters and proteins involved in cell wall synthesis (see Figure 3).

Bacterial-like proteins as well as archaeal proteins are combined into the metabolic schemes depicted in Figure 3 and 4. The energy, carbon and nitrogen metabolism

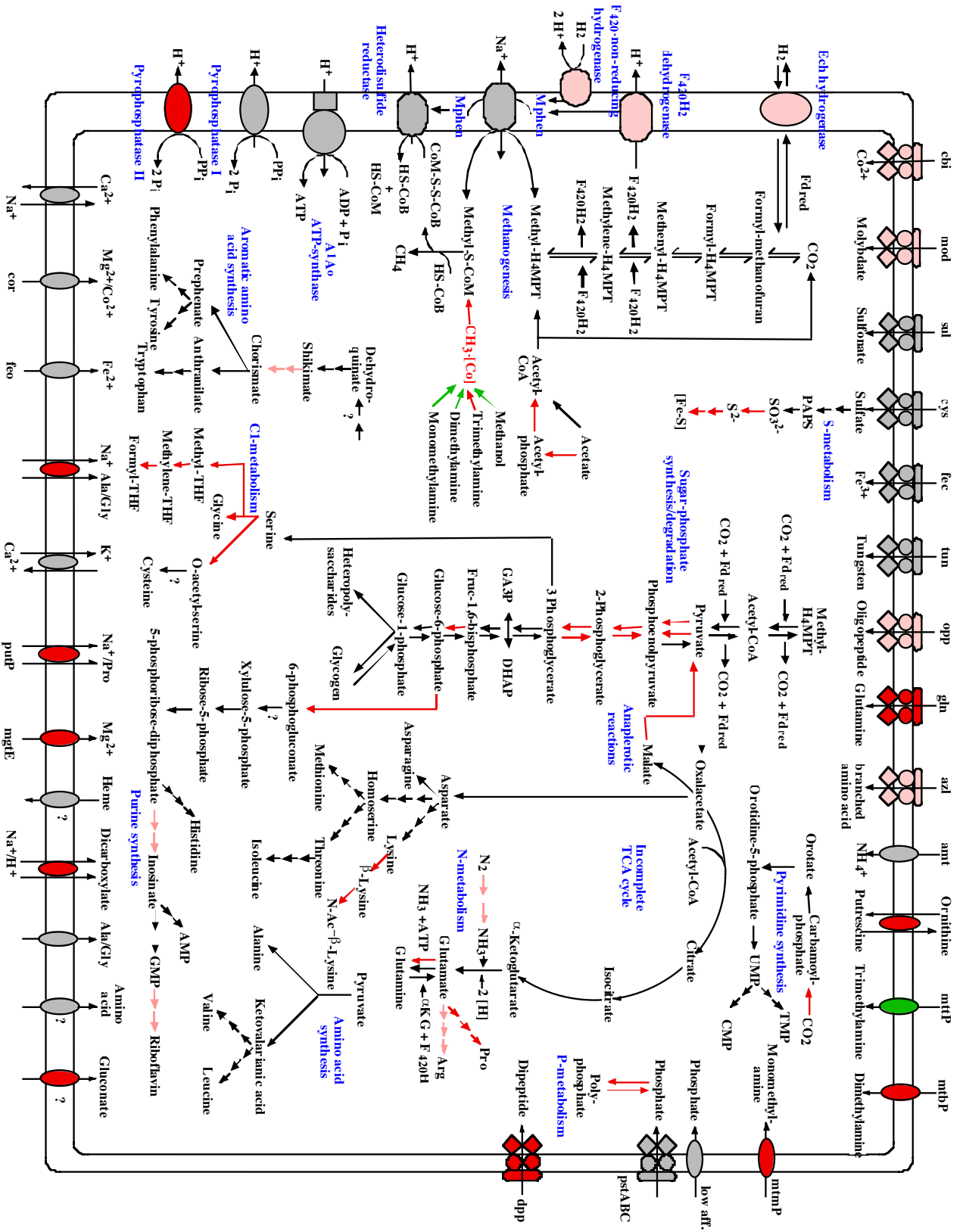
pathways as well as membrane transport systems of *M. mazei* are summarized in Figure 3. Typically, the genes for the CO₂ reduction pathway to methane are archaeal. This is also true for the heterodisulfide reductase and the A₁A₀-ATP synthase. Genes for two acetate-activation pathways are present: a bacterial type acetate kinase + phosphotransacetylase (not found in *A. fulgidus* and *Halobacterium* sp. NRC-1) and the archaeal ADP-forming acetyl-CoA-synthetase. Two copies of the genes encoding the CO dehydrogenase / acetyl-CoA-synthase with best blast e-values in Archaea were detected. Proteins involved in methyl group transfer from the methylamines and methanol to coenzyme M exhibit a mosaic-type of origin. The methyltransferases with in-frame TAG codons which demethylate the substrates and methylate the corrinoid proteins are unique for *Methanosarcina* except the ORF for trimethylamine which is otherwise only found in bacteria (*Desulfitobacterium halfniense*). The corrinoid proteins and the methyltransferases, which demethylate the corrinoid proteins and methylate coenzyme M, are clearly bacterial-like. This is also true for permeases that take up dimethylamine and monomethylamine.

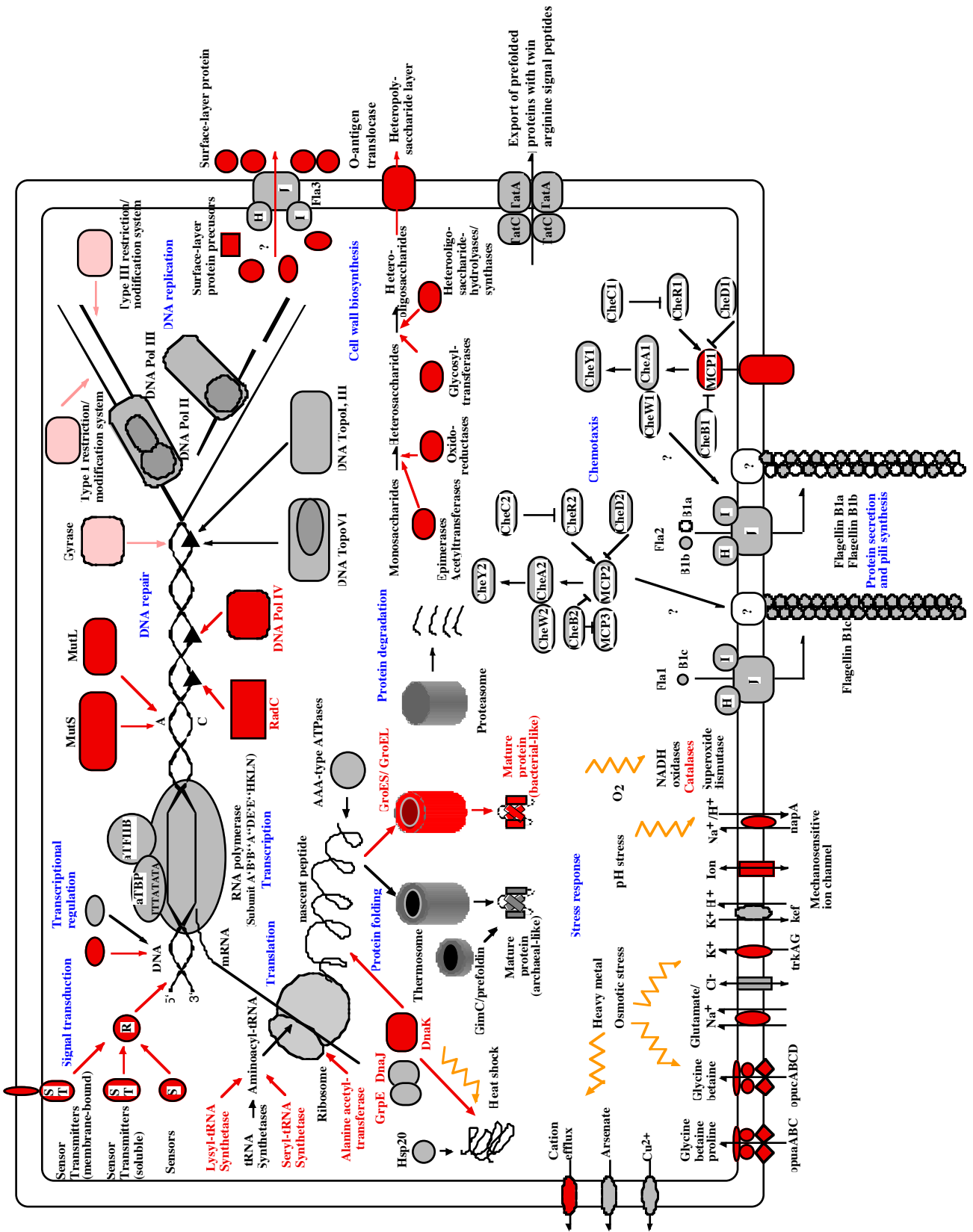
Hydrogenases and the F₄₂₀H₂ dehydrogenase, a proton pump functionally analogous to complex I of the respiratory chain, are of a mosaic type. The latter enzyme complex consists of 13 subunits organized on the chromosome like complex I on the *Escherichia coli* genome (Bäumer *et al.*, 2000). Most of the subunits of the F₄₂₀H₂ dehydrogenase are closely related to their counterparts in Bacteria. Exceptions are those subunits that interact with the methanogenic substrates reduced coenzyme F₄₂₀ (FpoF, archaeal origin) and methanophenazine (FpoO, *Methanosarcina*-specific).

It is also apparent from Figure 3 that most amino acid biosynthetic pathways are archaeal except proline synthesis which involves bacterial-like proteins (e.g. glutamate-5-kinase, γ -glutamyl phosphate reductase and pyrroline-5-carboxylate reductase). Tetrahydrofolate-dependent C1 metabolism from serine/glycine and many transporters have their best blast e-values in the bacterial domain. There are two membrane-bound pyrophosphatases. One closely related to archaeal / eukaryotic pyrophosphatases, the other homologous to corresponding enzymes of Bacteria.

Aspects of *M. mazei* macro molecular synthesis and assembly, sensory transduction and stress response are

Figure 3. Summary of central metabolism and solute transport. Reactions catalyzed by bacterial-like enzymes are marked by red arrows (based on the definition of bacterial-like ORFs as described in Methods, see also Tab. 1; bacterial-like ORFs with assigned function but no asserted pathway and hypothetical ORFs are not shown). The corresponding enzymes are: PEP synthetase, pyruvate-orthophosphate dikinase, enolase, phosphoglycerate mutase, glucose-6-phosphate dehydrogenase (sugar-phosphate synthesis/degradation); sulfite reductase (sulfur metabolism); components of the mono-, di-, and trimethylamine methyltransferases (methanogenesis from methylamines); acetate kinase, phosphotransacetylase (methanogenesis from acetate); serine hydroxymethyltransferase, pyrroline-5-carboxylate reductase, glutamate-5-kinase, γ -glutamyl phosphate reductase, serine transacetylase, shikimate kinase (amino acid biosynthesis); methylene-tetrahydrofolate- (THF) reductase, methylene-THF dehydrogenase, methenyl-THF cyclohydrolase (C₁-metabolism), phosphoribosyl-glycinamide formyltransferase, phosphoribosyl-aminoimidazolecarboxamide formyl-transferase (inosinate biosynthesis); polyphosphate kinase, exopolyphosphatase (phosphate metabolism); glutamine synthetase (N-metabolism); L-lysine-2,3 aminomutase, β -lysine-N-acetyl transferase for synthesis of N-acetyl- β -lysine (osmoregulation). Pathways that are in part catalyzed by bacterial-like enzymes are indicated by pink arrows. Reactions and pathways catalyzed by archaeal proteins are marked with black arrows. Bacterial-like transporters and proteins containing bacterial-like subunits are shown in red and pink, respectively. Examples are hydrogenases, the coenzyme F₄₂₀H₂ dehydrogenase and several transport systems. Proteins found in other Archaea are shown in grey. Transporters and enzymes only found in *Methanosarcina* species are shown in green. Permeases are represented as ovals; ABC transporters are shown as composite figures of ovals, diamonds and circles, channels are shown as cylinders. Export or import of solutes is designated by the direction of the arrows. Abbreviations: [FeS], iron-sulfur clusters; Fd_{red}, reduced ferredoxin; H₄MPT, tetrahydromethanopterin, F₄₂₀H₂, reduced coenzyme F₄₂₀, HS-CoM, coenzyme M, HS-CoB, coenzyme B, CoM-S-S-CoB, heterodisulphide of HS-CoM and HS-CoB; Mphen, methanophenazine; PPI, pyrophosphate.





depicted in Figure 4. The genome contains two complete sets of chemotaxis genes, three flagellin B1 genes, one twitching motility gene (*pilT*), and three *flaHJI* systems for type II protein secretion. Since pili structures have been previously observed in *M. mazei* (Lai *et al.*, 2000), we hypothesize that this strain employs at least one set of its *che/flaHJI/flaB1* genes for twitching motility related to biofilm formation (Zellner *et al.*, 1996). The third set of FlaHJI proteins (without any associated *flaB1* genes) might be involved in general protein export processes including export of surface layer proteins. Protein export also occurs by a sec-independent TAT system (twin arginine transport system). The *M. mazei* genome contains four clusters of cell envelope related genes that constitute 5.2 % of the genome. These four cell envelope islands consist of 160 genes needed for the cell surface S-layer proteins plus the thick heteropolysaccharide layer (Kandler and König, 1998). Many of the proteins have their best blast e-value in bacteria, while many others are unique to *Methanosarcina* strains.

The *M. mazei* replication, transcription and translation machinery is typically archaeal (Figure 4). In contrast, components highly homologous to bacterial proteins are the MutS/MutL and RadC DNA repair systems as well as a gyrase and the DNA polymerase IV. More than 50 transcriptional regulators including iron dependent repressors and members of the ArsR, AscN and MarR families were identified which have their best score in bacteria. *M. mazei* contains a complete archaeal set of aminoacyl-tRNA synthetases plus a second class II lysyl-tRNA and a second seryl-tRNA synthetase, both of which are closely related to the corresponding proteins of bacteria. It is of special interest that the genetic information for the archaeal thermosome and the bacterial GroES/GroEL system is present; this raises the question as to their involvement in the folding of archaeal and bacterial proteins.

Stress response is important for *M. mazei* living in environments where the conditions might change. For osmotic adaptation the organism contains genetic information to produce N-acetyl- β -lysine, transporters for glycine-betaine, potassium ion uptake systems, sodium ion dependent symporters/ antiporters, an archaeal chloride channel for uptake, and extrusion of compatible solutes. The organism also contains two catalases, a superoxide dismutase and several NADH oxidases plus metal ion exporters and heat shock proteins to cope with various kinds of stress. Many of these proteins are not found in other Archaea or are closely related to bacterial proteins.

How did a metabolically versatile methanogen such as *M. mazei* evolve? It contains all the biochemical abilities and pathways of the $H_2 + CO_2$ utilizing methanogens. Therefore, it has not lost genetic information. Rather, the core genome of a simple methanogen ($H_2 + CO_2$ utilizer) was supplemented by acquiring genes for utilizing new

substrates. This was accomplished by vertical evolution and by lateral gene transfer which has been recognized as an important secondary evolutionary mechanism (Woese, 1998). Along with the acquisition of new pathways for energy generation from acetate and methylamines, additional genes came for environmental adaptation such as oxidative, pH and metal stress and osmotic adaptation. Even genes for bacterial chaperons are present which have not been detected yet in other Archaea. As far as can be seen from the annotation, this new genetic information is not present on the genome in form of large islands of bacterial genes. Rather, it is more or less scattered over the whole genome (supplementary Figure 1) indicating that the new abilities of *Methanosarcina* species were the result of multiple events involving numerous donors.

Claims for lateral gene transfer between members of the bacterial and the archaeal domain are not new. Evidence for its importance comes from the analysis of various completely sequenced genomes and from comparisons of a number of genes. The bacterium *Thermotoga maritima* contains 24% of ORFs most similar to archaeal ones (Nelson, 1999). They are present in a number of islands, many of them being involved in transport processes. The archaeon *Halobacterium spec.* NRC-1 contains proteins uniquely homologous to *Bacillus subtilis* and *Deinococcus radiodurans* which is attributed by the authors to lateral gene transfer (Ng *et al.*, 2000). ORFs for fatty acid degradation are present in *A. fulgidus* which are unknown in other sequenced archaeal genomes but have homologies in bacteria (Klenk *et al.* 1997, Doolittle, 1999). Another interesting example is that bacterial methane oxidation and archaeal methanogenesis involve common genes for the C_1 oxidation/reduction pathways (Chistoserdova *et al.*, 1998).

M. mazei and its closest relatives thrive in fresh water and marine sediments. They are surrounded by bacterial partners and competitors. Since evolution is influenced by the size and the genetic diversity of local populations, it is of interest that many of the bacterial-like genes of *M. mazei* are closely related to genes of obligate anaerobic (e.g. *Clostridium spec.*, *Desulfitobacterium halfrniense*) and facultative anaerobic bacteria (e.g. enteric bacteria) followed by cyanobacteria (e.g. *Anabaena spec.*, *Nostoc punctiforme*). The great number of IS elements and transposases might indicate that *M. mazei* took, and continues to take advantage of the genetic diversity of its environment by lateral gene transfer. The step from hydrogenotrophic to methylotrophic growth of methanogens apparently proceeded at moderate temperatures because an extremely thermophilic methanogen utilizing acetate or methylamines is not known, the most thermophilic species being *M. thermophila* with an optimum growth temperature of 50°C (Boone *et al.*, 1993).

Figure 4. Overview of different cellular functions of *Methanosarcina mazei*. Bacterial-like proteins and proteins containing bacterial-like subunits are shown in red and pink, respectively. Proteins found in other Archaea are shown in grey. Orange arrows indicate various stress conditions. Question marks indicate hypothetical subunits of protein complexes or hypothetical interactions of proteins. Abbreviations: ST, sensor and transmitter domains of sensor protein transmitters (total number 31), S, proteins with sensor domain only (total number 3); R, receiver regulators (total number 9); aTBP, archaeal TATA-box binding protein (the letters within the protein indicate the consensus sequence of archaeal promoter structures), aTfIIIB, archaeal transcription initiation factor IIB; DNA Pol, DNA polymerase; Topo, Topoisomerase.

Experimental Procedures

The *M. mazei* Gö1 (DSMZ 3647) genome was determined by a whole-genome shotgun approach. Total genomic DNA of *M. mazei* was sheared randomly and DNA fragments of a size of 2 – 3 kb were used to construct small insert libraries in pTZ19R. A large insert library was constructed from *Sau3aI* partially digested genomic DNA, which was cloned in the cosmid vector system SuperCos1. Insert ends of the recombinant plasmids were sequenced by using ABI PRISM 377 or LICOR IL 4200 automated DNA sequencer. All generated sequences were assembled into contigs with P. Greens phrap assembling tools and have been edited with GAP which is part of the Staden package software (Staden *et al.*, 2000). More than 40,000 sequences (8.6-fold coverage) were aligned. Problems with misassembled regions caused by repetitive sequences were solved using orientation of sequence readings, comparison of predicted sizes of inserts, sequencing of overlapping cosmid inserts and PCR technologies. Closure of sequencing gaps was accomplished through primer-walking on plasmid and cosmid templates. Sequences from cosmids served for a verification of the orientation, order, and linkage of the contigs. Initial ORF prediction was accomplished automatically by the CRITICA software (Badger *et al.*, 1999) implemented in the ERGO software (Integrated Genomics, Inc., <http://www.IntegratedGenomics.com>) (Overbeek *et al.*, 2000). The ERGO ORFs annotation tool uses FASTA searches against its own non-redundant database (version of 24.07.01, 258 genomes, 17 Archaea, 83 eukaryotes, 158 Bacteria). The ERGO annotation was verified and refined by blasting all ORFs against the public Genbank/EMBL databases. With these tools a first identification of putative functions of gene products was made. Further decisions were performed manually through ERGO analysis tools. The threshold value for the 1,043 ORFs which had their best blast *e*-values in the bacterial domain was $e \leq 10^{-5}$. ORFs were referred to as bacterial-like proteins under the following conditions: 1. Blast hits were only obtained for proteins from Bacteria taking into account a total number of 200 hits. 2. If an archaeal protein was found among the 200 hits the difference of the *e*-value exponents between the bacterial first hit and the first hit inside the archaeal domain was calculated. This difference of *e*-value exponents had to be ≥ 14 . The accuracy of the second definition was verified by generating Neighbour-Joining trees (see below), which indicated a clear clustering of the *M. mazei* protein with bacterial proteins and a significant separation from archaeal proteins. Putative frameshifts were checked and corrected manually and where frameshifts remained they were considered as authentic.

Neighbour-joining (NJ) trees were constructed from characteristic bacterial-like proteins (Saitou and Nei, 1987). NJ trees were built from amino acid sequences extracted from Ergo database (<http://www/integratedgenomics.com>) and aligned by means of the Clustal W algorithm. Sequence distance matrices were computed with the PROTDIST program, in the Phylip 3.5c package (Felsenstein, 1989). Tree robustness was assessed by 1000 bootstrap iterations.

For the identification of repeats in intergenic regions the program REPuter was used (Kurtz and Schleiermacher, 1999). Input was the complete DNA sequence, with coding elements masked according to the annotation.

The complete *M. mazei* genome sequence has been deposited in the ERGO database (<http://www.integratedgenomics.com>), in Genbank (<http://www.ncbi.nlm.nih.gov>, Acc. No. AE008384) and the Göttingen Genomics Laboratory database (<http://www.g2l.bio.uni-goettingen.de>).

Acknowledgements

This work was supported by a grant of the “Niedersächsisches Ministerium für Wissenschaft and Kultur” to the Göttingen Genomics Laboratory and by grants of the “Deutsche Forschungsgemeinschaft”.

References

- Abken, H.J., Tietze, M., Brodersen, J., Bäumer, S., Beifuss, U., and Deppenmeier, U. 1998. Isolation and characterization of methanophenazine and function of phenazines in membrane-bound electron transport of *Methanosarcina mazei* Gö1. *J. Bacteriol.* 180: 2027-2032.
- Bäumer, S., Ide, T., Jacobi, C., Johann, A., Gottschalk, G., and Deppenmeier, U. 2000. The F₄₂₀H₂ dehydrogenase from *Methanosarcina mazei* Gö1 is a redox-driven proton pump closely related to NADH dehydrogenases. *J. Biol. Chem.* 275: 17968-17973.
- Badger, J.H. and Olsen, G.J. 1999. CRITICA: coding region identification tool invoking comparative analysis. *Mol. Biol. Evol.* 16: 512-524.
- Boetius, A., *et al.* 2000. A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* 407: 623-626.
- Doolittle, W.F. 1999. Phylogenetic classification and the universal tree. *Science* 284: 2124-2128.
- Boone, D.R., Whitman, W.B., and Rouvière, P. 1993. Diversity and taxonomy of methanogens. In: *Methanogenesis – Ecology, Physiology, Biochemistry and Genetics*. J.G. Ferry, ed. Chapman and Hall, New York – London. p. 35-80.
- Bult, C.J. *et al.* 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273: 1058-1073.
- Chistoserdova, L., Vorholt, J.A., Thauer, R.K., and Lidstrom, M.E. 1998. C₁ transfer enzymes and coenzymes linking methylotrophic Bacteria and methanogenic Archaea. *Science* 281: 99-102.
- Deppenmeier, U., Lienard, T., and Gottschalk, G. 1999. Novel reactions involved in energy conservation by methanogenic archaea. *FEBS Lett.* 457: 291-297.
- Felsenstein, J. 1989. PHYLIP-phylogeny inference package (version 3.2). *Cladistics* 5: 164-166.
- Gottschalk, G. and Thauer, R.K. 2001. The Na⁺-translocating methyltransferase complex from methanogenic archaea. *Biochim. Biophys. Acta* 1505: 28-36.
- Heidelberg, J.F. *et al.* 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406: 477-483.
- Hippe H., Caspari, C., Fiebig, K., and Gottschalk, G. 1979. Utilization of trimethylamine and other N-methyl compounds for growth and methane formation by *Methanosarcina barkeri*. *Proc. Natl. Acad. Sci. USA* 76: 494-498.
- Kandler, O. and König, A. 1998. Cell wall polymers in Archaea (Archaeobacteria). *Cell Mol. Life Sci.* 54: 305-308.
- Klenk, H. P. *et al.* 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 390: 364-370.
- Kurtz, S. and Schleiermacher, C. 1999. REPuter-Fast Computation of Maximal Repeats in Complete Genomes. *Bioinformatics* 15, 426-427.
- Lai, M.C., Shu, C.M., Chen, S.C., Lai, L.J., Chiou, M.S., and Hua, J.J. 2000. *Methanosarcina mazei* strain O1M9704, methanogen with novel tubule isolated from estuarine environment. *Curr. Microbiol.* 41: 15-20.
- Nelson, K.E., *et al.* 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399: 323-329.
- Ng, W. V., *et al.* 2000. Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl. Acad. Sci. USA* 97: 12176-12181.

- Overbeek, R., *et al.* 2000. WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Res.* 28: 123-125.
- Paul, L., Ferguson, D.J., and Krzycki, J.A. 2000. The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of *Methanosarcina barkeri* contain in-frame and read-through amber codons. *J. Bacteriol.* 182: 2520-2529.
- Rupp, A. *et al.* 2000. The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* 407: 508-513.
- She, Q. *et al.* 2001. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. USA* 98: 7835-7840.
- Saitou, N. and Nei, M. 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.
- Smith, D.R. *et al.* 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: Functional analysis and comparative genomics. *J. Bacteriol.* 179: 7135-7155.
- Staden, R., Beal, K.F., and Bonfield, J.K. 2000. The Staden package. *Methods Mol. Biol.* 132: 115-130.
- Woese, C. 1998. The universal ancestor. *Proc. Natl. Acad. Sci. USA* 95: 6854-6859.
- Zellner, G., Macario, A.J.L., and Conway de Macario, E. 1996. Microbial subpopulations in the biofilm attached to the substratum and in the free flocs of a fixed-bed anaerobic bioreactor. *Appl. Microbiol. Biotechnol.* 46: 443-449.
- Zinder, S.H. 1993. Physiological ecology of methanogens. In: *Methanogenesis – Ecology, Physiology, Biochemistry and Genetics*. J.G. Ferry, ed. Chapman and Hall, New York – London. p. 35-80.

