Complete Genome Sequence of the Genetically Tractable
Hydrogenotrophic Methanogen Methanococcus maripaludis†

e. L. Hendrickson,1 R. Kaul,2,3 Y. Zhou,3 D. Bovee,3 P. Chapman,3 J. Chung,3 E. Conway de Macario,4 J. A. Dodsworth,1 W. Gillett,3 D. E. Graham,3 M. Hackett,6 A. K. Haydock,1 A. Kang,3 M. L. Land,7 R. Levy,3 T. J. Lie,1 T. A. Major,8 B. C. Moore,1 I. Porat,8 A. Palmeiri,3 G. Rouse,3 C. Saenphimmachak,3 D. Söll,9 S. Van Dien,10 T. Wang,1,6 W. B. Whitman,8 Q. Xia,1,6 Y. Zhang,1,6 F. W. Larimer,7 M. V. Olson,2,3,11 and J. A. Leigh1*

Department of Medicine, Division of Medical Genetics,2 and Departments of Microbiology,3 Chemical Engineering,6 and Genome Sciences,11 University of Washington, University of Washington Genome Center,5 and United Metabolics,10 Seattle, Washington; Wadsworth Center, New York State Department of Health, Division of Molecular Medicine, The University at Albany (SUNY), Albany, New York7; Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, Texas9; Genome Analysis and Systems Modeling, Oak Ridge National Laboratory, Oak Ridge, Tennessee7; Department of Microbiology, University of Georgia, Athens, Georgia8; and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut11

Received 12 May 2004/Accepted 13 July 2004

The genome sequence of the genetically tractable, mesophilic, hydrogenotrophic methanogen Methanococcus maripaludis contains 1,722 protein-coding genes in a single circular chromosome of 1,661,137 bp. Of the protein-coding genes (open reading frames [ORFs]), 44% were assigned a function, 48% were conserved but had unknown or uncertain functions, and 7.5% (129 ORFs) were unique to M. maripaludis. Of the unique ORFs, 27 were confirmed to encode proteins by the mass spectrometric identification of unique peptides. Genes for most known functions and pathways were identified. For example, a full complement of hydrogenases and methanogenesis enzymes was identified, including eight selenocysteine-containing proteins, with each being paralogous to a cysteine-containing counterpart. At least 59 proteins were predicted to contain iron-sulfur centers, including ferredoxins, polyferredoxins, and subunits of enzymes with various redox functions. Unusual features included the absence of a Cdc6 homolog, implying a variation in replication initiation, and the presence of a bacterial-like RNase HI as well as an RNase HII typical of the Archaea. The presence of alanine dehydrogenase and alanine racemase, which are uniquely present among the Archaea, explained the ability of the organism to use L- and D-alanine as nitrogen sources. Features that contrasted with the related organism Methanocaldococcus jannaschii included the absence of inteins, even though close homologs of most intein-containing proteins were encoded. Although two-thirds of the ORFs had their highest Blastp hits in Methanocaldococcus jannaschii, lateral gene transfer or gene loss has apparently resulted in genes, which are often clustered, with top Blastp hits in more distantly related groups.

The methanogenic Archaea (methanogens) occupy a unique metabolic niche, as they produce methane, which is a useful energy source and a powerful greenhouse gas. These organisms are found in diverse anaerobic habitats, ranging from aquatic and marine sediments to sewage digesters and the rumens and large intestines of herbivores and other mammals (127). In these habitats, the degradation of organic matter results in the production of H2 and other intermediates by fermentative organisms. By maintaining an extremely low partial pressure of H2, the methanogens keep fermentative pathways energetically favorable. In addition, some methanogens may occupy niches where hydrogen is produced predominately by geothermal reactions.

Metabolically, methanogens are divided into those that specialize in CO2 reduction and those that also use acetate and/or methyl compounds. The former group, the hydrogenotrophs, use H2 as an electron donor to reduce CO2 to methane. Many hydrogenotrophic species can substitute formate or certain low-molecular-weight alcohols and ketones for H2. Complete genome sequences have been published for three hydrogenotrophic methanogens, Methanocaldococcus jannaschii (13), Methanothermobacter thermoautotrophicus (105), and Methanopyrus kandleri (104), all of which are thermophiles or hyperthermophiles. Of the methanogens that utilize acetate and methyl compounds, complete genome sequences have been published for two species, Methanosarcina acetivorans (26) and Methanosarcina mazei (19), both of which are mesophiles. In addition, partial sequences have been published for two psychrophiles, the hydrogenotroph Methanogenium frigidum and the methylotroph Methanolobus burtonii (97).

Genome sequences of methanogens have answered many questions, but they have inspired many others. More than half of the genes in Methanocaldococcus jannaschii lack a predicted function (13), and this proportion has not declined significantly as other methanogen sequences have been determined. The proportions of genes of unknown functions, which are either homologous to other genes of unknown function or have no known homologs at all, are 55% for the Methanothermobacter thermoautotrophicus genome (105) and 51% for the Methanosarcina acetivorans genome (26).
These observations demonstrate a pressing need to identify the functions of genes in the methanogenic Archaea. Many of the most effective approaches involve genetic manipulation, determining the phenotypes of mutants, or affinity tagging proteins in vivo to facilitate their purification. Nevertheless, few genetic tools are available for sequenced species of the methanogenic Archaea. Genetics can be used for *Methanosarcina acetivorans* (88) but not for any previously sequenced hydrogenotrophic methanogenic species. Here we present the genome sequence of the genetically tractable species *Methanococcus maripaludis*.

*M. maripaludis* is a mesophilic hydrogenotrophic methanogen that was isolated from salt marshes (48). Like all methanogens, *M. maripaludis* belongs to the kingdom *Euryarchaeota* in the domain *Archaea*. *M. maripaludis* belongs to the family *Methanococcaceae* in the order *Methanococcales* (12). Although *M. maripaludis* is related to *Methanocaldococcus jannaschii*, it possesses many novel features, and approximately one-third of its genes lack orthologs in *Methanocaldococcus jannaschii* (see below). Extensive studies of physiology and regulation have already been performed with *M. maripaludis*, and many of them used genetic tools (53, 65, 113). The virtues of *M. maripaludis* as a model species are apparent. Dense liquid cultures are obtained overnight and colonies grow on agar medium in 2 days (49). Chemostat cultures can now be established reproducibly (36a). Many important genetic manipulations are routine, including transformation, complementation with shuttle vectors, gene deletions, and insertions of reporters (28). New approaches to genetic manipulation are being implemented (B. C. Moore and J. A. Leigh, submitted for publication), and the first comprehensive expression array and proteomic analyses have been completed (E. L. Hendrickson, M. Hackett, and J. A. Leigh, unpublished data).

**MATERIALS AND METHODS**

**Strain.** *M. maripaludis* strain S2 (120) is a wild-type isolate that has also been designated strain LL.

**Genome sequencing.** *M. maripaludis* strain S2 was sequenced by the use of standard DNA sequencing protocols and data collection tools. Initially, 43,950 small insert shotgun reads and 1,536 fosmid-end sequencing reads were collected by using Big Dye terminator sequencing chemistry. The sequences were assembled and viewed with phred/phrap/consed software. To facilitate opening and viewing of the genome assemblies in consed, we created a phd.ball file from each phd file. The creation of phd.ball files reduced the time to open the genome assembly in consed to <10 min. The initial assembly provided 8.14 × Q20 sequence coverage (Q20, error rate of 0.1% [22]) and provided 99.46% coverage of the 1.66-Mb genome. The *M. maripaludis* genome was finished by using the autofinish tool of consed (30). In all, 770 finishing reads were attempted and four PCR templates were generated to finish the sequence. The fosmid-end sequence reads were tiled along the postshotgun sequence assembly with SeqTile software (W. Gillett, unpublished software tool), which identified two grossly misassembled regions. The misassemblies identified were all due to the presence of nearly identical ribosomal DNA repeats. Two unique fosmid clones that spanned the misassembled regions were selected and sequenced to 8 × Q20 coverage. A third fosmid clone spanning difficult-to-finish regions was dugenized by a transposon mutagenesis protocol suggested by the manufacturer (Epigenic Technologies). Random clones from mutagenesis experiments were picked, and small insert shotgun reads and 1,536 fosmid-end sequencing reads were collected by using standard Big Dye terminator sequencing chemistry. The backbone from these independently assembled fosmid clones were imported into the genome assembly to resolve misassemblies and to improve the sequence quality of difficult-to-finish regions.

**Sequence validation.** The final assembly contained 38,601 reads, including reads from autofinish and advanced finishing experiments as well as the backbone from the three independently sequenced fosmid clones. The final validation of the sequence assembly was performed by using SeqTile software and comparing the restriction fingerprint patterns of 417 fosmid clones with the virtual fingerprint pattern of the finished sequence assembly by using three enzymes, BglII, EcoRI, and HindIII. The 417 fosmid clones provided 10 × clone coverage and uninterrupted ×2 fingerprint coverage for the finished sequence assembly.

**Genome analysis and annotation.** The genome sequence was analyzed, and annotations were entered at the Genome Channel facility at the Oak Ridge National Laboratories (http://genome.ornl.gov/microbial/mmar/). Automated annotations were accomplished for all open reading frames (ORFs) by Blastp comparisons to protein databases, Pfam (incorporating Pfam, TIGRFams, SmartHMM, Prosite, Prints, and ProDom algorithms), and Clusters of Orthologous Groups (COGs). Most ORFs were also annotated by hand. In brief, preliminary identifications were first made by Blastp analysis, and high expectation values covering at least 80% of the ORF were sought. The list of Blast hits was then scanned for highly homologous proteins whose functions had been experimentally determined. The other analysis tools mentioned above, as well as the presence of a gene in an operon with functionally related genes, were then examined for supporting evidence. ORFs with clear homologies but uncertain functions were designated members of gene families, relatives of genes of known function, or conserved hypothetical proteins. Putative transporters were checked against M. Saier’s transport protein classification web site (http://www.biology.ucsd.edu/~msaier/transport). Genes were viewed graphically with Artemis (http://www.sanger.ac.uk/Software/Artemis/).

**Proteomics.** During the course of our work, we analyzed 18 protein samples from a variety of *M. maripaludis* cultures. Protein mixtures were digested with trypsin and separated by multidimensional liquid chromatography as described previously (117, 118). “Bottom-up” proteomics was performed by tandem mass spectrometry using a Finnegan LCQ classic quadrupole ion trap mass spectrometer equipped with an electrospray ion source. Peptide sequences derived from proteinic fragments were matched to *M. maripaludis* ORFs by computational reference to the genome sequence by using Sequest (21), DTASelect (109), and d2g (118) software and by manual interpretations of individual collision-induced dissociation mass spectra.

**Nucleotide sequence accession number.** The *M. maripaludis* genome sequence is available at the EMBL/GenBank/DDJB database under accession number BX950229 and at the Oak Ridge National Laboratories Genome Channel at http://genome.ornl.gov/microbial/mmar/.

**RESULTS AND DISCUSSION**

**General features and organization.** The genome of *M. maripaludis* consists of a single circular chromosome of 1,661,137 bp (Table 1). Genome modeling predicts 1,722 protein-coding genes, with 52% carried on the forward strand and 48% carried on the complementary strand. The genome encodes four 5S rRNAs, three 23S rRNAs, three 16S rRNAs, 38 tRNAs, and RNase P. Since no distinct origin of replication can be discerned (see below), nucleotide numbering was begun at the end of an rRNA gene cluster. ORFs were numbered consecutively along the genome and given the prefix “Mmp.” Functional categories of protein-coding genes (ORFs) are listed in Table 2 and mapped in Fig. 1. A complete list of ORFs and their functional annotations is available at http://www.ncbi.nlm.nih.gov/geneomes/altik.cgi?db=G&gi=394. The *M. maripaludis* sequence

<table>
<thead>
<tr>
<th>Table 1. General features of the <em>M. maripaludis</em> genome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameter</strong></td>
</tr>
<tr>
<td>Total no. of bases</td>
</tr>
<tr>
<td>No. of protein-coding genes</td>
</tr>
<tr>
<td>No. of predicted transmembrane proteins</td>
</tr>
<tr>
<td>Gene density (genes/kb)</td>
</tr>
<tr>
<td>Average gene length (bp)</td>
</tr>
<tr>
<td>Protein-coding percentage</td>
</tr>
<tr>
<td>No. of tRNAs</td>
</tr>
<tr>
<td>No. of rRNA operons</td>
</tr>
<tr>
<td>%GC</td>
</tr>
</tbody>
</table>
TABLE 2. Functional categories of proteins encoded by \textit{M. maripaludis} genome

<table>
<thead>
<tr>
<th>Function</th>
<th>No. of ORFs</th>
<th>% of ORFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid biosynthesis</td>
<td>82</td>
<td>4.8</td>
</tr>
<tr>
<td>Biosynthesis of cofactors, prosthetic groups, and carriers</td>
<td>68</td>
<td>3.9</td>
</tr>
<tr>
<td>Cell envelope</td>
<td>8</td>
<td>0.56</td>
</tr>
<tr>
<td>Cellular processes (cell division, chemotaxis, and motility)</td>
<td>30</td>
<td>1.7</td>
</tr>
<tr>
<td>Central intermediary metabolism</td>
<td>112</td>
<td>6.5</td>
</tr>
<tr>
<td>DNA metabolism</td>
<td>40</td>
<td>2.3</td>
</tr>
<tr>
<td>Energy metabolism (methanogenesis, hydrogen metabolism, and ATPase)</td>
<td>81</td>
<td>4.7</td>
</tr>
<tr>
<td>Fatty acid and phospholipid metabolism</td>
<td>6</td>
<td>0.35</td>
</tr>
<tr>
<td>Hypothetical proteins</td>
<td>102</td>
<td>5.9</td>
</tr>
<tr>
<td>Unique proteins of unknown function</td>
<td>27</td>
<td>1.6</td>
</tr>
<tr>
<td>Conserved hypothetical proteins</td>
<td>656</td>
<td>38</td>
</tr>
<tr>
<td>Protein fate</td>
<td>22</td>
<td>1.3</td>
</tr>
<tr>
<td>Protein synthesis</td>
<td>119</td>
<td>6.9</td>
</tr>
<tr>
<td>Purines, pyrimidines, nucleosides, and nucleotides</td>
<td>44</td>
<td>2.6</td>
</tr>
<tr>
<td>Regulatory functions</td>
<td>38</td>
<td>2.2</td>
</tr>
<tr>
<td>Transcription</td>
<td>22</td>
<td>1.3</td>
</tr>
<tr>
<td>Transport and binding proteins</td>
<td>86</td>
<td>5.0</td>
</tr>
<tr>
<td>Unclassified and unknown function</td>
<td>179</td>
<td>10.4</td>
</tr>
<tr>
<td>Total</td>
<td>1,722</td>
<td>100</td>
</tr>
</tbody>
</table>

* Adapted from The Institute for Genome Research’s functional categories.


\textit{M. maripaludis} has a low G+C content, 33.1%, which is fairly homogeneous across the genome (Fig. 1). The only large deviations are in the regions of the rRNAs. Compared to the overall G+C content, the intergenic regions have a lower percentage, 25.7% G+C, while the ORFs contain 34% G+C.

Like those of all \textit{Bacteria} and \textit{Archaea}, most of the genes of \textit{M. maripaludis} appear to be present in polycistronic operons. While some genes with common functionality are clustered into operons in \textit{M. maripaludis}, many are not. Clustered genes include many of those encoding ribosomal components, methanogenic enzymes, conserved hypothetical proteins, and other multicomponent enzymes. However, compared to the case for the \textit{Bacteria}, a striking feature of the \textit{M. maripaludis} genome is the tendency for some functionally related genes to be unlinked. Genes for amino acid, purine, or pyrimidine biosynthesis are rarely linked, but instead are often present in operons with genes of unrelated or unknown function. Tryptophan biosynthetic genes, which are clustered in an operon, are a notable exception.

Although 19 inteins were found in the \textit{Methanocaldococcus jannaschii} genome (84), none were found in \textit{M. maripaludis}. Nevertheless, \textit{M. maripaludis} encodes close homologs of all but two of the intein-containing ORFs.

Lateral gene transfer and gene loss. Among the protein-coding genes of \textit{M. maripaludis}, the highest frequency (64% of ORFs) of high-scoring Blastp hits occurred with genes of \textit{Methanocaldococcus jannaschii}, the closest relative of \textit{M. maripaludis} with a known genome sequence (13). The frequencies of top Blastp hits with other groups were as follows: other methanogens, 12%; \textit{Euryarchaeota}, 18%; \textit{Crenarchaeota}, 0.2%; \textit{Bacteria}, 9.6%; and \textit{Eukarya}, 0.6% (see the supplemental material). These figures suggest that lateral gene transfer into the \textit{M. maripaludis} lineage from distant lineages has occurred but that it has not been as frequent as in the mesophilic methyl-otroph \textit{Methanosarcina mazei} (19) or \textit{Methanosarcina acetivorans} (26). The lack of any significant deviations from the average mol% G+C among the ORFs implies that any lateral transfers into \textit{M. maripaludis} occurred long ago, allowing the G+C content to equilibrate over time, or were from organisms with similar G+C percentages.

For Fig. 1, the highest-scoring Blastp hits in \textit{Methanocaldococcus jannaschii}, other methanogens, other \textit{Archaea}, and \textit{Bacteria} plus \textit{Eukarya} were color coded around the genome. The distribution was nonrandom. Top Blastp hits to groups other than \textit{Methanocaldococcus jannaschii} were noticeably less frequent in a wide sector centered around base 1,500,000 than in the opposite sector. Furthermore, discrete clusters of genes had top hits to predominantly one or more of the more distant groups at the expense of \textit{Methanocaldococcus jannaschii} (Table 2 and Fig. 1). The most notable of these clusters (Mmp0483 to -0536) contains the genes encoding the molybdenum formylmethanofuran dehydrogenase (Fmd; top hits to other methanogens) as well as a gene for molybdopterin biosynthesis and three ABC transporters, two of which were for molybdate. \textit{Methanocaldococcus jannaschii} lacks Fmd (see below), and the genes for Fmd and molybdenum-related functions could have been transferred laterally to the \textit{M. maripaludis} lineage from outside of the methanococci or could have been present in an ancestor and lost from \textit{Methanocaldococcus jannaschii}. The cluster from Mmp0973 to Mmp0988 contains carbon monoxide dehydrogenase/acetyl coenzyme A (CoA) synthase (Cdh); in this case, \textit{Methanocaldococcus jannaschii} has the enzyme, yet all seven subunits yielded top hits to \textit{Methanothermobacter thermotrophicus}. The clustered nature of these genes is consistent with the idea that clustering can both facilitate and result from the lateral transfer of functionally related genes (61). Interestingly, a family of putative ATPases known only in \textit{Methanocaldococcus jannaschii} (\textit{Methanocaldococcus jannaschii} ORFs MJ0625, MIECL26, MJ1076, and MJ1006, with more distant relatives in \textit{Methanocaldococcus jannaschii} and
Pyrococcus species) is entirely absent from M. maripaludis. Also, ribulose bisphosphate carboxylase, which is present in Methanocaldococcus jannaschii and other methanogens (25), is not encoded in the M. maripaludis genome.

Proteins of known and unknown function. Of the 1,722 predicted proteins, a function was assigned to 758 (44%) of them. Another 835 (48%) ORFs were either homologous to genes of unknown function (conserved hypothetical proteins) or had uncertain affiliations with genes of known function. The remaining 129 (7.5%) were unique to M. maripaludis and had no known homologs.

For the 129 predicted proteins that were unique to M. maripaludis, a function was assigned to 758 (44%) of them. Another 835 (48%) ORFs were either homologous to genes of unknown function (conserved hypothetical proteins) or had uncertain affiliations with genes of known function. The remaining 129 (7.5%) were unique to M. maripaludis and had no known homologs.

For the 129 predicted proteins that were unique to M. maripaludis, a function was assigned to 758 (44%) of them. Another 835 (48%) ORFs were either homologous to genes of unknown function (conserved hypothetical proteins) or had uncertain affiliations with genes of known function. The remaining 129 (7.5%) were unique to M. maripaludis and had no known homologs.

### TABLE 3. Phylogenetic distributions and functions of clustered non-Methanocaldococcus jannaschii top Blastp hit categories

<table>
<thead>
<tr>
<th>ORF interval (Mmp no.)</th>
<th>Predominant group(s)</th>
<th>Represented functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0345-0371</td>
<td>Other Archaea</td>
<td>Conserved hypothetical proteins, carbohydrate metabolism</td>
</tr>
<tr>
<td>0483-0536</td>
<td>Other methanogens and Bacteria</td>
<td>Conserved hypothetical proteins, Fmd, Mo transport, molybdoenzyme biosynthesis, metal chelatase, probable cation transport</td>
</tr>
<tr>
<td>0709-0734</td>
<td>Bacteria and other methanogens</td>
<td>Divalent cation transport, UvrABC</td>
</tr>
<tr>
<td>0753-0762</td>
<td>Bacteria</td>
<td>Conserved hypothetical proteins</td>
</tr>
<tr>
<td>0772-0862</td>
<td>Other methanogens and Bacteria</td>
<td>Conserved hypothetical proteins, Hsp</td>
</tr>
<tr>
<td>0973-0988</td>
<td>Other methanogens</td>
<td>Cdh</td>
</tr>
</tbody>
</table>

*Intervals containing 70% or more top Blast hits for other methanogens, other Archaea, and Bacteria combined.

*Groups containing the top Blast hits for at least 20% of the ORFs in the interval.

FIG. 1. Circular map of M. maripaludis genome. First (outer) double ring, top Blast hits; second double ring, ORFs unique to M. maripaludis; third double ring, functional categories; black single ring, deviation from average mol% G+C; inner ring, GC skew. Top Blast hits are coded as follows: blue, Methanocaldococcus jannaschii; magenta, other methanogens; green, other Archaea; brown, Bacteria and Eukarya. Sectors containing top Blast hits predominately to groups other than Methanocaldococcus jannaschii are shown, with ORF number intervals. Functional categories are coded as follows: red, replication and repair; green, energy metabolism; blue, carbohydrate metabolism; cyan, lipid metabolism; magenta, transcription; yellow, translation; sky blue, cellular processes; orange, amino acid metabolism; pink, metabolism of cofactors; light red, nucleotide metabolism; gray, conserved hypothetical proteins; white, hypothetical proteins; brown, unassigned proteins; black, other; pale green, RNAs.
paludis, the existence of 27 was confirmed by proteomics. Peptides that belong to these proteins were identified unequivocally in samples from M. maripaludis by mass spectrometry, and these proteins were therefore designated unique proteins of unknown function (Table 2; see the supplemental material). The remaining unique proteins were designated hypothetical proteins.

Information systems. (i) Replication. Like most Archaea (8, 9), M. maripaludis contains a subset of the eukaryal replication proteins. However, the M. maripaludis replication apparatus has some distinctive features. At the replication initiation stage, both Methanocaldococcus jannaschii (31) and M. maripaludis lack a homolog of Cdc6, which forms part of the pre-replication complex in Eukarya and other Archaea (75). Both species also lack a discrete transition in GC skew. Since the location of cdc6 and the GC skew transition typically provide the major evidence for the origin of replication, there is no clear indication for the origin in M. maripaludis. In fact, M. maripaludis has many GC skew transitions distributed around the chromosome (Fig. 1). Methanocaldococcus jannaschii is known to maintain 3 to 15 copies of the chromosome during its life cycle (72), and both species may employ multiple origins to achieve this end.

Despite the lack of Cdc6, M. maripaludis has four homologs of the minichromosome maintenance (MCM) proteins (Mmp0030, Mmp0470, Mmp0748, and Mmp1024) that are recruited to the initiation complex and provide helicase activity (62). In contrast, Methanothermobacter thermautotrophicus contains only one MCM protein, which forms a homomultimeric ring structure (100). In M. maripaludis, each protein may act independently to form a helicase, or all four may be required. Like the case for nearly all DNA processes, topoisomerases are also important for replication initiation (116), and a bacterial-type topoisomerase I (Mmp0956) is encoded in M. maripaludis, as in many Archaea. As expected, there is no gene for reverse gyrase, which is known only for hyperthermophiles, including Methanocaldococcus jannaschii.

M. maripaludis has all of the expected components for polymerization. Like other Euryarchaeota, M. maripaludis encodes a single family B DNA polymerase (Mmp0380), but it also contains an archaeon-specific two-subunit DNA polymerase (Mmp0008 and Mmp0026) (45). Processivity factors (Mmp1126 and Mmp1711), enabling long-range DNA polymerization, and the clamp-loading proteins (Mmp032 and Mmp0427) that bind the processivity factors are also encoded. Notable distinctions include the observation that M. maripaludis, like Methanocaldococcus jannaschii and Methanothermobacter thermautotrophicus, has only one subunit of the single-stranded DNA binding protein, Mmp1032 (56). Other Euryarchaeota have a protein with three different subunits (56). Like other Archaea (11), M. maripaludis has a single-subunit primase, p48 (Mmp0071), in contrast to Eukarya, which require p58 as well (68). Interestingly, M. maripaludis also has a homolog of DnaG (Mmp1286), the bacterial primase (86). Hence, M. maripaludis may have two separate primase systems.

Like other Archaea, M. maripaludis removes the primers from the lagging strand of DNA replication by using flap endonuclease (Fen1/Rad2 and Mmp1313) and RNase HII (Mmp1374) (58, 89). M. maripaludis is unique among Archaea in that it also encodes a homolog of RNase HI (Mmp0837), the main RNAse in Bacteria (58). The M. maripaludis homolog is similar to RNase HI from Clostridium and may have been acquired by lateral gene transfer. Okazaki fragments are probably ligated together by a homolog of eukaryal ATP-dependent DNA ligase I (Mmp0970) (45).

M. maripaludis encodes a homolog of Smc (structural maintenance of chromosome; Mmp1397), which is believed by analogy with those of the Eukarya to play a part in archaeal chromosome segregation and condensation (7). The archaeal type II topoisomerase (Mmp0989 and Mmp1437), a two-subunit protein that decatenates chromosomes (7), is encoded, as are distant homologs of the Escherichia coli proteins XerC and XerD (Mmp0472 and Mmp0743) (24), suggesting the possibility of two separate systems for chromosome decatenation. M. maripaludis also contains two homologs of the plasmid partitioning gene parA (Mmp0704 and Mmp0593) and parB (Mmp0592) (29).

(ii) Cell division. In bacterial cell division, a ring of proteins forms at the cell center and constricts as the septum grows (94). This system is shared by the Euryarchaeota, which often have multiple ftsZ homologs (8). Two homologs, Mmp1436 and Mmp1500, are found in M. maripaludis. M. maripaludis also carries a homolog of bacterial minD (Mmp1145), which is thought to encode an inhibitor of FtsZ ring formation (44). M. maripaludis lacks homologs of the two E. coli proteins that bind to the FtsZ ring, FtsA and ZipA (94). Surprisingly, M. maripaludis encodes a homolog of Cdc48 (Mmp0176), which in Saccharomyces cerevisiae plays a role in the membrane fusion of organelles (60). Since Archaea do not possess organelles, the role of the Cdc48 homolog is unknown.

(iii) Recombination and repair. The M. maripaludis genome contains several genes that are predicted to code for recombination and repair systems. These include the Mre11-Rad50 double-stranded-break repair system (Mmp1340 to -1341), the archaeal RecA homolog RadA (Mmp1222), the related protein RadB (Mmp0617), which is thought to amplify RadA activity, a RecF homolog (Mmp1682), and the unique archaeal Holliday junction resolvase, Mmp0336 (57). Several base excision repair proteins were found, including ExoA (Mmp1012) and an endonuclease III-related protein (Mmp0586), but no DNA photolyase was present. M. maripaludis, like Methanocaldococcus jannaschii, is missing a homolog of the mismatch repair gene mutS found in other archaeal species. However, unlike Methanocaldococcus jannaschii, M. maripaludis has an E. coli-like excinuclease, UvrABC (Mmp0727 to -0729), which functions as a wide-substrate-range nucleotide excision repair system (82). Weak homologs are also present for the MutT nucleotide diphosphate hydrolase (Mmp0339) and the O6-methylguanine-DNA methyltransferase (Mmp0069), which repair specific damage to nucleotides (101).

(iv) Transcription. M. maripaludis contains a complete set of genes for the archaeal transcriptional machinery. Single homologs of the TATA box binding protein (Mmp0257) and transcription factors B (TFB; Mmp0041) and E (TFE; Mmp0036) are present (38). Homologs are also found for all 13 subunits of the archaeal RNA polymerase (71).

(v) Transcriptional regulators. Like other sequenced Archaea, M. maripaludis encodes a few bacterial regulatory family members, including TetR (the most numerous, with four members), ArsR, LysR, and PadR (see supplemental material).
M. maripaludis also encodes regulators that are found only in Archaea, including the known nitrogen repressor NrpR (Mmp0607) (65) and a member of an Archaea-specific COG that is predicted to be a transcriptional regulator (Mmp0907). Two-component regulators, which are numerous in Methanosarcina acetivorans (26) and Methanothrophicota thermotrophicus (105), are absent from Methanocaldococcus jannaschii and Methanopyrus kandleri (104). Excluding those involved in chemotaxis (see below), only one two-component regulator is encoded by M. maripaludis (Mmp1303 and -1304). In total, 24 transcriptional regulators are predicted with confidence, which is about the number expected for a genome of this size (114).

(vi) Translation. The factors governing translation in Archaea have been determined by homology to eukaryotic and prokaryotic systems (5, 27). M. maripaludis possesses homologs of the four archaeal translation elongation factors (Mmp1131, -1369, -1370, and -1401) and all but one of the archaeal translation initiation factors (Mmp0061, -0284, -0297, -0457, -0603, -0952, -1208, -1618, and -1707). While other Archaea, including Methanocaldococcus jannaschii, encode two subunits of initiation factor 2B (5), M. maripaludis apparently has only subunit 1 (Mmp1618).

Aminoacyl tRNA synthetases were identified for 18 amino acids. Two genes for the alpha subunit of the two-subunit phenylalanyl-tRNA synthetase were found (Mmp0688 and -1496), as is the case for Methanocaldococcus jannaschii and Methanotherapythermota thermotrophicus. No aminoacyl tRNA synthetases were found for asparagine or glutamine. Instead, asparaginyl-tRNA and glutaminyl-tRNA are made by tRNA-dependent amidotransferases, and all of the subunits for the enzyme that forms both asparaginyl-tRNA and glutaminyl-tRNA (GatABC; Mmp1510, -0946, and -0575) and the enzyme specific for glutaminyl-tRNA synthesis (GatDE; Mmp1266 and -1265) are encoded (108).

M. maripaludis makes use of selenocysteine, the 21st co-translationally inserted amino acid. In Bacteria, selenocysteinyl-tRNA synthesis begins with the charging of tRNA^Sec with serine, followed by dehydration and the addition of a selenide moiety from selenophosphate. A homolog of the selenophosphate synthetase was identified with confidence, which is about the number expected for a genome of this size (114).

In total, 24 transcriptional regulators are predicted with confidence, which is about the number expected for a genome of this size (114).

M. maripaludis makes use of selenocysteine, the 21st co-translationally inserted amino acid. In Bacteria, selenocysteinyl-tRNA synthesis begins with the charging of tRNA^Sec with serine, followed by dehydration and the addition of a selenide moiety from selenophosphate. A homolog of the selenophosphate synthetase was identified with confidence, which is about the number expected for a genome of this size (114).

Methanocaldococcus jannaschii possesses only the tungsten form (41), possibly due to its hyperthermophilicity (40). Unlike most methanogens, fvdB (Mmp1691) in M. maripaludis is not in an operon with the genes for the other Fwd subunits but is encoded adjacent to the Vhu hydrogenase (see below). Even more unusual, the fmd operon has two adjacent fmdB (Mmp0511 and -0512) genes, with one encoding a selenocysteine version of the protein.

M. maripaludis contains two sets of enzymes that catalyze the final step in methanogenesis, namely methyl coenzyme M reductases I (Mcr) and II (Mrt) (31, 90). M. maripaludis encodes only one methylreductase complex (Mmp1555 to -1559), and due to the high levels of homology between Mcr and Mrt, sequence similarity was insufficient to distinguish which complex is present. However, the operon configuration and position next to the Mtr operon are characteristic of Mcr (85).

The reduction of methyl-coenzyme M produces a mixed disulfide from coenzymes M and B (37), and heterodisulfide reductase (Hdr) reduces this disulfide to the free coenzymes (18). Like other obligate hydrogenotrophs, M. maripaludis encodes an Hdr with three subunits, A, B, and C. Two hdrBC clusters are present (Mmp0642-Mmp0643 and Mmp1054-Mmp1053), as are two hdrA genes, one for a selenocysteine-type (Mmp1697) protein, adjacent to the Vhu hydrogenase genes, and the other a cysteine-type (Mmp0825) protein, adjacent to the Vhc hydrogenase cluster. In contrast, Methanocaldococcus jannaschii contains only the selenocysteine-type enzyme (31).

(ii) Hydrogenases. M. maripaludis contains six nickel-iron hydrogenases. Like Methanocaldococcus voltae, M. maripaludis contains complete gene clusters for two coenzyme 

H2 and CO2 by the methanogenic pathway (see Fig. 2 for the major pathways in M. maripaludis). The first step in CO2 reduction to methane is catalyzed by formylmethanofuran dehydrogenase, which is found in both tungsten (Fwd) (Mmp1244 to -1249 and -1691) and molybdenum (Fmd) (Mmp0200 and -0508 to -0512) forms in M. maripaludis. In contrast, Methanocaldococcus jannaschii contains only the tungsten form (41), possibly due to its hyperthermophilicity (40). Unlike most methanogens, fvdB (Mmp1691) in M. maripaludis is not in an operon with the genes for the other Fwd subunits but is encoded adjacent to the Vhu hydrogenase (see below). Even more unusual, the fmd operon has two adjacent fmdB (Mmp0511 and -0512) genes, with one encoding a selenocysteine version of the protein.

M. maripaludis contains two sets of enzymes that catalyze the final step in methanogenesis, namely methyl coenzyme M reductases I (Mcr) and II (Mrt) (31, 90). M. maripaludis encodes only one methylreductase complex (Mmp1555 to -1559), and due to the high levels of homology between Mcr and Mrt, sequence similarity was insufficient to distinguish which complex is present. However, the operon configuration and position next to the Mtr operon are characteristic of Mcr (85).

The reduction of methyl-coenzyme M produces a mixed disulfide from coenzymes M and B (37), and heterodisulfide reductase (Hdr) reduces this disulfide to the free coenzymes (18). Like other obligate hydrogenotrophs, M. maripaludis encodes an Hdr with three subunits, A, B, and C. Two hdrBC clusters are present (Mmp0642-Mmp0643 and Mmp1054-Mmp1053), as are two hdrA genes, one for a selenocysteine-type (Mmp1697) protein, adjacent to the Vhu hydrogenase genes, and the other a cysteine-type (Mmp0825) protein, adjacent to the Vhc hydrogenase cluster. In contrast, Methanocaldococcus jannaschii contains only the selenocysteine-type enzyme (31).

(ii) Hydrogenases. M. maripaludis contains six nickel-iron hydrogenases. Like Methanocaldococcus voltae, M. maripaludis contains complete gene clusters for two coenzyme F420-reducing hydrogenases, one of which is a selenocysteine-containing cluster (Fru) (Mmp1382 to -1385) and the other of which is a cysteine-containing cluster (Frc) (Mmp0817 to -0820) cluster, and for two non-F420-reducing hydrogenases, which also contain selenocysteine (Vhu) (Mmp1692 to -1696) and cysteine.
(Vhc) (Mmp0821 to -0824) (6). *M. maripaludis* also encodes two separate multisubunit energy-conserving hydrogenases, Eha (Mmp1448 to -1467) and Ehb (Mmp0400, -0940, -1049, -1073, -1074, -1153, -1074, -1153, -1469, and -1621 to -1629), which are homologous to those first identified in *Methanothermobacter thermautotrophicus* (111). These hydrogenases are thought to couple ion gradients to certain endergonic reduction steps in methanogenesis and biosynthesis. Some members of these gene clusters are predicted to encode polyferredoxins and integral membrane proteins. Like *Methanocaldococcus jannaschii*, *M. maripaludis* Eha is encoded by one cluster that is colinear with the cluster found in *Methanothermobacter thermautotrophicus* (111). In contrast, while some Ehb subunits are encoded by one small cluster, most of the genes appear to be scattered throughout the genome (31). EhbH and EhbI are fused into one ORF (Mmp1626).

(iii) Formate dehydrogenases. *M. maripaludis* contains two formate dehydrogenases (Fdh) (Mmp0138-Mmp0139 and Mmp1297-Mmp1298), either one of which enables growth on formate as an alternative to hydrogen and CO₂ (122). A formate transporter (Mmp1301) is encoded upstream of the latter formate dehydrogenase. Interestingly, while *Methanococcus vannielii* has selenium-dependent and -independent formate dehydrogenases (47), both *M. maripaludis* subunits contain selenocysteine. As a result, *M. maripaludis* cannot grow on formate in the absence of selenocysteine incorporation (91).

(iv) Selenocysteine-containing proteins. Nine selenocysteine-containing proteins are encoded by the genome. Of
these, eight are subunits of methanogenic enzymes, hydrogenses, or formate dehydrogenases. These selenocysteine-containing proteins are the B subunits of the molybdenum (Mmp0511) and tungsten (Mmp1691)-containing formylmethanofuran dehydrogenases, one of the twoHdr subunit A proteins (Mmp1697), subunit A of the Fru hydrogenase (Mmp1382), subunits D (Mmp1696) and U (Mmp1693) of the Vhu hydrogenase, and the A subunits of both formate dehydrogenases (Mmp0138 and -1300). The ninth selenocysteine-containing protein is selenophosphate synthetase (SeID; Mmp9004). These observations are in agreement with the experimental detection of selenocysteine-containing proteins in *M. maripaludis* (92).

(v) Acetyl-coenzyme A synthesis. *M. maripaludis* is capable of autotrophic growth and uses carbon monoxide dehydrogenase/acetyl-coenzyme A synthase (CODH/ACS, or Cdh) to fix CO₂ and form acetyl-CoA. Like many other methanogens (76), the CODH/ACS genes in *M. maripaludis* are found in a single cluster (Mmp0070K to -0985). In addition to genes for CODH/ACS itself, gene Mmp0979 encodes an iron-sulfur protein that may be involved in electron transfer to CODH/ACS (67). Mmp0977, carried in an adjacent operon, is related to the nickel insertion protein for the *Rhodospirillum rubrum* carbon monoxide dehydrogenase and may be involved in biosynthesis or maturation of the prosthetic group (46).

As an alternative to autotrophy, *M. maripaludis* can assimilate acetate. Acetyl-CoA is then synthesized by acetyl-CoA synthetase. *M. maripaludis* has both the ADP-forming enzyme characteristic of *Archaea* and *Eukarya* (Mmp0253) (81) and the AMP-forming type (Mmp0148) found commonly in *Bacteria* and *Eukarya*. *Methanocaldococcus jannaschii* has only the former.

(vi) Reductive branch of TCA cycle. Once acetyl-CoA is produced, it is converted by the incorporation of another CO₂ into pyruvate by a multisubunit pyruvate:ferredoxin oxidoreductase (Por; Mmp1502 to -1507) (67) and thence to oxaloacetate by pyruvate carboxylase (Ryc; Mmp0340 and -0341) (80, 102). Oxaloacetate enters the tricarboxylic acid (TCA) cycle, which proceeds in the reductive direction. All of the enzymes for the reductive arm of the TCA cycle are present, leading from oxaloacetate to 2-oxoglutarate. 2-Oxoglutarate oxidoreductase (Kor; Mmp0003, -1305, -1306, and -1307) belongs to a family of multisubunit ferredoxin oxidoreductases that also includes pyruvate oxidoreductase (67). Unlike the other family members, the genes for the subunits of 2-oxoglutarate oxidoreductase are not all linked: the beta and gamma subunit genes are adjacent, but the alpha and delta (ferredoxin) subunits are each encoded in a different location. The oxidative branch of the TCA cycle is absent.

(vii) Glycolysis and gluconeogenesis. As in *Methanocaldococcus jannaschii*, most of the genes for glycolysis and gluconeogenesis are present in *M. maripaludis* (99). These genes include those for two noncanonical phosphoglycerate mutases, Mmp0112 and Mmp1439 (33), an unusual ADP-dependent enzyme (Mmp1296) with both glucokinase and phosphofructokinase activities (95), and an archaeal-type fructose bisphosphate aldolase (Mmp0686) (103). Genes for glycogen synthesis and degradation are also present, including glycogen synthase (Mmp1294).

(viii) Nitrogen metabolism. *M. maripaludis* can meet its nitrogen needs from several sources, including ammonia assimilation, the fixing of diatomic nitrogen, and the assimilation of alanine, the last of which is unusual for an archaeon (119). Ammonia is assimilated by a Lα-type glutamine synthetase (Mmp1209) (15). Glutamate synthase provides glutamate. As in other *Archaea*, the glutamate synthase large chain is encoded in three separate subunits (Mmp0080 to -0082), which correspond to domains of a single protein in *Bacteria*. The glutamate synthase small chain seems to be absent. The presence of an alanine dehydrogenase (Mmp1513), an alanine racemase (Mmp1512), and an alanine permease (Mmp1511) account for the unusual ability of *M. maripaludis* to use L- and D-alanine (Moore and Leigh, submitted).

*M. maripaludis* has a nitrogenase operon (51) that contains *nifH*, -D, -K, -E, -N, and -X (Mmp0853 and -0856 to -0860), encoding the nitrogenase complex and proteins that participate in the synthesis of the nitrogenase cofactor, as well as *nifL* (Mmp0854) and *nifL*₂ (Mmp0855), encoding proteins that regulate nitrogenase activity (52, 53). Homocitrate, a component of the nitrogenase cofactor, is synthesized by NifV in *Bacteria*. The NifV homolog in *M. maripaludis* that is responsible for homocitrate synthesis is probably AksA (Mmp0153), which is also involved in the synthesis of 2-oxosuberate, an intermediate in biotin and coenzyme B synthesis (42).

(ix) Ferredoxins and iron-sulfur proteins. Several steps in the pathways mentioned above require low-potential electrons. To transport these electrons, *M. maripaludis* encodes numerous iron-sulfur proteins. A total of 59 proteins are predicted to have 4Fe-4S centers, characterized by the motif CXXCXXCXXXC. These proteins include ferredoxins whose sole predicted function is to carry low-potential electrons and subunits of enzymes that catalyze low-potential redox reactions. Of particular interest are ferredoxins associated with hydrogenases and oxidoreductases, since several of these enzymes require low-potential electrons to drive enzymatic reactions. Among the hydrogenases, three ferredoxins are found in the Eha cluster, namely Mmp1463 (a polyferredoxin, containing many iron-sulfur centers), Mmp1464, and Mmp1465. The Ehb cluster contains two ferredoxins, Mmp1623 and Mmp1624 (a polyferredoxin). Vh contains Mmp0824 (a polyferredoxin), Vhu contains Mmp1692 (a polyferredoxin), Fru contains Mmp1384, and Fr contains Mmp0818. Among the multisubunit oxidoreductases, one subunit for each enzyme contains 4Fe-4S motifs as follows: indolepyruvate oxidoreductases (Ior) 1 and 2 (Mmp0316 and Mmp0713), 2-oxoal阀arate oxidoreductase (Vor; Mmp1273), 2-oxoglutarate oxidoreductase (Kor; Mmp1687), and pyruvate oxidoreductase (Por; Mmp1506). In addition, the fifth and sixth Por subunits, PorE and PorF (Mmp1503 and Mmp1502) (66) contain iron-sulfur motifs, as does Mmp0979, a PorE homolog predicted to be the electron carrier associated with CODH/ACS. The CODH/ACS alpha subunit (Mmp0985) also contains a 4Fe-4S motif that is presumably involved in electron transfer from the carrier to the active site.

Several additional enzymes have subunits containing 4Fe-4S motifs, which is indicative of electron transfer via unknown ferredoxins. These include the A and C subunits of the hetrodisulfide reductases (Mmp0825, Mmp1697, Mmp154, and Mmp1054), subunits H, F, and G of the formylmethanofuran
dehydrogenase (Mmp1244 to -1246), the β subunits of the formate dehydrogenases (Mmp0139 and Mmp1297), and the large subunit of glutamate synthase (Mmp0081). Additional enzymes that include 4Fe-4S motifs are succinate dehydrogenase/fumarate reductase (Mmp1067) and one of the two thymidylate synthases (Mmp0986). In addition, the functions of many iron-sulfur proteins found in the genome are not known.

Flagella and chemotaxis. M. maripaludis has the same arrangement of archaeal flagellar genes as that found in *Methanococcus voltae* (Mmp1666 to -1676) (50). Also present is an almost complete set of bacterial chemotaxis homologs (Mmp0925 to -0933). As in the α-Proteobacteria (2), no cheZ gene is present. Four homologs of sensory methyl-accepting chemotaxis proteins are present (Mmp0413, -0487, -0788, and -0929), suggesting the ability to respond to many different chemotactants.

Transporters. The *M. maripaludis* genome encodes 86 predicted transporter and binding proteins comprising approximately 48 transporter systems (see the supplemental material). The majority fall into the ABC transporter class. Iron transporters are highly prevalent, with one ferric iron ABC transporter cluster (Mmp0108 to -0110) and two iron-chelating ABC transporter clusters (Mmp0196 to -0198 and Mmp1181 to -1183) as well as a cluster of three ORFs that are predicted to be iron binding periplasmic proteins (Mmp1176 to -1178; however, Mmp1176 and -1177 encode separate amino and carboxyl ends of an iron binding protein and may represent a nonfunctional frame shift). There is also a homolog of a non-ABC ferrous iron uptake protein (Mmp0630) that is believed to be powered by ATP or GTP hydrolysis. *M. maripaludis* employs both molybdenum and tungsten as metal cofactors, and besides iron transporters, putative molybdenum transporters make up the other large group of transporters. There are four clusters of molybdenum ABC transporters (Mmp0205 to -0207, Mmp0504 to -0506, Mmp0514 to -0516, and Mmp1650 to -1652) as well as a separately encoded periplasmic molybdenum binding protein (Mmp1111). Two ORFs homologous to ABC sulfate transporter proteins (Mmp1518 to -1519) may actually comprise a tungsten transporter. *M. maripaludis* also has two members (Mmp0711 and -1108) of the CorA family of aqueous pore transporters, which transport divalent metal ions.

*M. maripaludis* can assimilate both ammonia and alanine as nitrogen sources, and two ammonium transporters (Mmp0065 and Mmp0068) and an alanine-cation symporter (Mmp1511) are encoded. *M. maripaludis* strains have been reported to take up a variety of amino acids (121), and several ORFs may encode additional amino acid transporters. Scattered around the genome are homologs of an ABC polar amino acid transporter system. While only one homolog of an ATP binding protein (Mmp0229) and a permease (Mmp0551) are seen, there are six putative periplasmic amino acid binding proteins (Mmp0455, -0550, -0712, -0770, -1224, and -1225). There is also a proline-Na⁺ symporter (Mmp0221) as well as a member of the amino acid-polyamine symporter-antiporter family (Mmp0850). An ABC phosphate transporter system is also present (Mmp1095 to -1099).

A series of three ORFs (Mmp0165 to -0167) showed homology to genes for drug efflux transporters, encoding the two components of an ABC drug efflux system separated by a predicted Na⁺-drug antiporter gene. Finally, a predicted Na⁺-H⁺ antiporter (Mmp0587) may allow for the interconversion of a sodium-motive force (produced by the methyltransferase step of methanogenesis) with a proton-motive force.

**S layer.** *M. maripaludis* encodes an S-layer precursor (Mmp0383) with high sequence similarity to that of *Methanococcus vannielii* (3).

Amino acid synthesis. (i) Glutamate family. As mentioned above, glutamine and glutamate are synthesized by glutamine synthetase and an archaeal-type glutamate synthase. Arginine is synthesized from glutamate via the intermediate ornithine. Homologs to all of the enzymes of the pathway except the initial enzyme are present (Mmp0013, -0063, -0073, -0116, -0553, -0897, -1013, -1101, and -1589). A homolog of the argF gene (Mmp0897) is also present, which is characteristic of the acetyl cycle version of the pathway of ornithine biosynthesis (73) that is known to occur in *Methanococcus vannielii* (78).

Like *Methanocaldococcus jannaschii*, the enzyme that catalyzes the first step in ornithine biosynthesis is unknown for *M. maripaludis* (73).

*Methanocaldococcus jannaschii* generates proline by the cyclization of ornithine, but the enzyme is evidently not a homolog of any known ornithine cyclodeaminase and has not been characterized (34). As in *Methanocaldococcus jannaschii*, the genes for proline biosynthesis in *M. maripaludis* are unknown.

(ii) Pyruvate family. Alanine is produced from pyruvate by a type I aminotransferase (77). Five type I aminotransferases (Mmp0096, -1072, -1216, -1396, and -1527; see below) are encoded, and experiments are needed to determine their specificities. Unlike other Archaea, *M. maripaludis* also has the potential to use the alanine dehydrogenase pathway since this enzyme is present; however, the gene is required only for alanine utilization, not alanine synthesis (Moore and Leigh, submitted).

As in other methanogens (20), isoleucine is synthesized by the citramalate pathway by use of the enzyme (R)-citramalate synthase (CimA; Mmp1018), which was first identified in *Methanocaldococcus jannaschii* (43).

Leucine and valine are synthesized by standard pathways. Recently, the leuA gene encoding isopropylmalate synthase (Mmp1063) was distinguished from its paralogs in the citramalate and α-ketosuberate pathways by mutagenesis in *M. maripaludis* (36a). The presence of 2-oxoisovalerate oxidoreductase (Mmp1271 to -1273) suggests the additional ability to produce the branched-chain amino acids from the corresponding branched-chain fatty acids.

(iii) Aspartate family. Aspartate is evidently synthesized by an aspartate aminotransferase (AspC) orthologous to the one identified in *Methanothermobacter thermautotrophicus* (Mmp0391) (110), Asparagine is synthesized by glutamine-hydrolyzing asparagine synthase (Mmp0918) (59); no homolog of ammonia-utilizing asparagine synthase is present. Hence, as in certain other organisms (79), a tRNA-independent pathway for asparagine synthesis appears to exist, despite the lack of an asparaginyl-tRNA synthetase (see above).

Threonine, methionine, and lysine are synthesized largely by standard pathways. Threonine and methionine share a common intermediate, homoserine, whose biosynthesis (Mmp1017, -1391, and -1702) is distinguished by separate enzymes for aspartate
kinase (Mmp1017) and homoserine dehydrogenase (Mmp1702), in contrast to the bifunctional enzymes typical of Bacteria. From homoserine, threonine is synthesized by standard enzymes (Mmp0135 and -0295). Like the case for many other archaeal genomes, only one ORF (MetE; Mmp0401) for the methionine biosynthesis pathway was found (39). Nevertheless, labeling studies with Methanocaldococcus jannaschii showed that methionine was formed from aspartate (107). Lysine is evidently synthesized with glycine hydroxymethyltransferase (GlyA), and while missing a homolog for SerC. Glycine is normally formed from phosphatidic acid pathway (Mmp0576, -0917, -0923, -1200, and -1398) (4), despite the lack of known orthologs for certain steps (99, 105).

(iv) Serine family. Three steps synthesize serine from 3-phosphoglycerate. Like Methanothermobacter thermautotrophicus and Methanocaldococcus jannaschii (99, 105), M. maripaludis has SerA (Mmp1588) and SerB (Mmp0541) but is missing a homolog for SerC. Glycine is normally formed from serine by glycine hydroxymethyltransferase (GlyA), and while homologs were found in Methanothermobacter thermautotrophicus and Methanocaldococcus jannaschii (99, 105), no homolog is present in the M. maripaludis genome.

As in Methanocaldococcus jannaschii (99) and Methanothermobacter thermautotrophicus (105), no genes involved in cysteine synthesis were identified.

(v) Aromatic amino acids. Chorismate is the branch point for phenylalanine, tyrosine, and tryptophan synthesis. Five of the seven enzymes in the known pathway of chorismate synthesis (Mmp0320, -0936, -1205, -1333, and -1394) were found in M. maripaludis. Like the case for many Euryarchaea, homologs for the initial steps are not apparent (17, 99). Because erythrose-4-phosphate is not a precursor for chorismate in M. maripaludis (112), it seems likely that the initial steps in this pathway are different from those found in Bacteria. Recently, the presence of a dehydroquinase dehydratase, which catalyzes the third step in the pathway, was confirmed by the construction of a deletion mutation of Mmp1394 (87). Phenylalanine and tyrosine biosynthesis are initiated by chorismate mutase (Mmp0578), followed by separate prephenate dehydratase (PheA; Mmp1528) and prephenate dehydrogenase (TyrA; Mmp1514) enzymes (70). In contrast, in many other organisms aroQ (encoding chorismate mutase) is fused with another aromatic amino acid biosynthetic genes or a regulatory domain (14). The entire standard pathway for tryptophan synthesis is also present (Mmp1002 to -1008). In addition to the de novo pathway, M. maripaludis can synthesize the aromatic amino acids by reductive carboxylation of the aryI acids phenylacetate, p-hydroxyphenylacetate, and indoleacetate (87). Indolepyruvate oxidoreductase catalyzes the key step in this pathway, and M. maripaludis contains two homologs of this enzyme system, Mmp0315-Mmp0316 and Mmp0713-Mmp0714.

(vi) Histidine. All of the genes for the biosynthesis of histidine (Mmp051, -0256, -0280, -0417, -0548, -0947, -0968, -1082, -1083, -1216, -1690, and -1722) were found except that for histidinol phosphate phosphatase (HisJ).

Nucleotide synthesis. (i) Purines. Almost all of the genes for the biosynthesis of purines are present in M. maripaludis, although there are variations from the pathways of Bacteria and Eukarya. Ribose phosphate is synthesized by the nonoxidative pentose phosphate pathway, as in Methanocaldococcus jannaschii (99, 126), and phosphoribosylpyrophosphate is synthesized by phosphoribosylpyrophosphate synthase (Mmp0410). Like the genomes of several other Archaea (55), the genome of M. maripaludis is missing the purN homolog for phosphoribosylglycinamide formyltransferase, but it does have the alternative enzyme purT (Mmp0123) (74). As with certain other Archaea and Bacteria, M. maripaludis encodes a two-subunit phosphoribosylformylglycinamidase synthase (PurL [Mmp0179] and PurQ [Mmp0178]) (98). M. maripaludis does not encode PurS even though it is encoded by Methanocaldococcus jannaschii and is required for phosphoribosylformylglycinamidase synthase activity in Bacillus subtilis (98).

As in other methanogens, N5-carboxamimidomimidazole ribonucleotide synthetase (PutK) and N6-carboxamimidomimidazole ribonucleotide mutase (PutE) activities seem to be fused into a single PurE homolog constituting phosphoribosylamimidazole carboxylase (Mmp0282) (106). For the final two steps in the de novo synthesis of IMP, which are normally catalyzed in Bacteria and Eukarya by a single bifunctional enzyme, only the archaean IMP cyclohydrolyase (PurO; Mmp1310) (35) has been identified; no homolog of PurH is encoded, and the gene for aminomimidazole carboxamidine ribonucleotide transformylase is not yet to be identified. All of the genes necessary for converting IMP to AMP and GMP are present (PurA, Mmp1432; PurB, Mmp0971; GuaA, Mmp1445; and GuaB, Mmp0133). Many of the genes involved in the biosynthesis of ATP, dATP, GTP, and dGTP from AMP and GMP are present (AdkA, Mmp1031; Ndk, Mmp0283; and NrdD, Mmp0227). However, neither bacterial nor archaean (55) ribonucleotide diphosphate reductase is present, and M. maripaludis presumably generates any dADP from dATP. No guanylate kinase, which catalyzes the formation of GDP from GMP, has been found in any of the Archaea (55).

(ii) Pyrimidines. All of the genes involved in the biosynthesis of UTP, the precursor of pyrimidines, are found in the genome of M. maripaludis. Like some other archaeal genomes, the M. maripaludis genome has two ORFs encoding orotate phosphoribosyltransferase (OrotE; Mmp0079 and Mmp1492) (10, 55). M. maripaludis contains all of the genes required for the conversion of UTP to CTP (PyrG; Mmp0893) and thence to GDP (Ndk; Mmp0283). In addition, CTP is converted to dCTP by ribonucleoside triphosphate diphosphatase (NrdD; Mmp0227). dTTP is evidently made from dCTP by a pathway that was recently elucidated in Methanocaldococcus jannaschii that avoids the production of toxic dUTP as an intermediate (64) as follows. A bifunctional dCTP deaminase-dUTP diphosphatase (Mmp1426) converts dCTP to dUMP, which is then converted to dTMP by thymidylate synthase (ThyA; Mmp0986 and Mmp1379). dTMP is converted to dTTP by thymidylate kinase (Tmk; Mmp0103) and thence to dTTP by nucleoside diphosphate kinase (Ndk; Mmp0227). dUTP diphosphatase (Dut; Mmp1075) may be present merely to scavenge dUTP produced by the spontaneous deamination of dCTP. For dCDP synthesis, neither the typical ribonucleotide-diphosphate reductase nor the alternative enzyme found in some Archaea is present. Thus, as with dADP, any dCDP presumably comes entirely from the triphosphate.

For dCDP synthesis, neither the typical ribonucleotide-diphosphate reductase nor the alternative enzyme found in some Archaea is present. Thus, as with dADP, any dCDP presumably comes entirely from the triphosphate.

(iii) Salvage pathways. As in Methanothermobacter thermautotrophicus, M. maripaludis has a hyoxanthine phosphoribosyltransferase (Hpt; Mmp0145) that may also serve as a guanine phosphoribosyltransferase (96). M. maripaludis also has homologs of adenine phosphoribosyltransferase (Mmp0660) and uracil phosphoribosyltransferase (Mmp0680).
Coenzymes. Numerous genes for the known biosynthetic pathways of the conventional coenzymes were identified. Among the coenzymes of methanogenesis, all of the coenzyme M biosynthesis genes in Methanocaldococcus jannaschii (32) were found to have orthologs in *M. maripaludis*. Several genes encoding steps in the synthesis of coenzyme F₄₂₀ have been identified in *Methanocaldococcus jannaschii* (32), and orthologs were found in *M. maripaludis*. Unlike *Methanocaldococcus jannaschii*, *M. maripaludis* has three homologs of F₃₉₀ synthetase, encoded by Mmp0160, Mmp0314, and Mmp0715 (115). However, methanococci are not known to produce F₃₉₀, which suggests that these genes must have some other purpose. Only a few genes in methanopterin biosynthesis have been identified. Although a dihydropteroate synthase (MptH) was identified in *M. maripaludis* (125), no clear ortholog could be found in *M. maripaludis*. A series of condensation (AksA; Mmp0153), spontaneous hydration, and oxidative decarboxylation (AksF; Mmp0880) reactions are involved in the synthesis of 2-oxoasuberate, an intermediate in coenzyme B as well as coenzyme F₄₄₀ (42).

Aminotransferases. *M. maripaludis* contains 11 ORFs that have been identified as aminotransferases. In general, aminotransferases are divided into four subgroups based on structural similarity (77). *M. maripaludis* has five aminotransferases from subgroup I (Mmp0096, -1072, -1216, -1396, and -1527), which is generally the most common subgroup, comprising aspartate, alanine, tyrosine, phenylalanine, and histidinol phosphate aminotransferases (77). Because the substrate specificities of the subgroup I aminotransferases are highly variable, it is often difficult to assign a specific function based only on homology to a characterized enzyme. However, Mmp1216 could be assigned unambiguously as histidinol phosphate aminotransferase (HisC) based on its homology to the enzyme from *Halobacterium volcanii* (16), in which the function was demonstrated by genetic complementation of a histidine auxotroph.

*M. maripaludis* has three subgroup II aminotransferases, encoded by Mmp0224, -0865, and -1101. Mmp0224 was assigned as glutamate-1-semialdehyde aminotransferase (HemL) based on its homology to the enzyme from *Staphylococcus solfataricus* (83). One subgroup III aminotransferase (Mmp0132) was identified whose homology suggests that it could account for the branched-chain amino acid aminotransferase activity detected in *Methanococcus* spp. (123, 124). One subgroup IV aminotransferase, Mmp0391, was assigned as aspartate aminotransferase (AspC) based on its homology to the enzyme from *Methanothermobacter thermautotrophicus* (110). A final aminotransferase, Mmp1680, does not belong to any of the recognized subgroups and was assigned as glucosamine-fructose-6-phosphate aminotransferase (GlmS) based on its homology to the enzyme from *Thermus thermophilus* (23).

Conclusions. The *M. maripaludis* genome reveals much about the organism, with nearly half of the genes having assignable functions. Many of these functions will no doubt be confirmed as experiments continue with this genetically manipulable species. However, the absence of assignable functions is at least equally striking, with nearly half of the proteins designated as conserved hypothetical proteins. Genes with unassignable functions include 129 predicted proteins with no homologs in any other species. For these cases, the availability of genetic tools will also be important. In addition, the small genome, which is typical of many Archaea, facilitates studies of global regulation, which should complement genetic analyses.

As expected, the genome size of *M. maripaludis* is in line with those of its hydrogenotrophic relatives *Methanocaldococcus jannaschii*, *Methanothermobacter thermautotrophicus*, and *Methanopyrus kandleri*, and it is much smaller than the genome sequence of the nutritionally versatile *Methanosarcina* species. The *M. maripaludis* genome provides ample opportunity for comparisons with its nearest relative with a sequenced genome, *Methanocaldococcus jannaschii*. Approximately two-thirds of the *M. maripaludis* ORFs had their highest-scoring Blastp hits in *Methanocaldococcus jannaschii*, and many pathways and functions are held in common. The two species also share some novel features, including an unusual mechanism of replication initiation implied by the absence of a Cdc6 protein and the lack of a discrete transition in GC skew. However, the contrasts between *M. maripaludis* and *Methanocaldococcus jannaschii* are interesting as well. Some differences can be attributed to the growth temperatures, with *M. maripaludis* being mesophilic and *Methanocaldococcus jannaschii* being hyperthermophilic. Reverse gyrase, which is present only in *Methanocaldococcus jannaschii*, is needed only in hyperthermophiles, and the presence in *M. maripaludis* of only a molybdenum formylmethanofuran dehydrogenase also agrees with the correlation of molybdenum- and tungsten-containing enzymes with the temperature (40). A systematic difference in amino acid preferences for homologs between the two species has already been reported (36). Other differences include the presence of only one methyl-coenzyme M reductase in *M. maripaludis* and the absence of inteins from *M. maripaludis*. Insights into some of the contrasts between *M. maripaludis* and *Methanocaldococcus jannaschii* came from our analysis of top Blast hit categories. ORFs with top Blast hits to groups more distant than *Methanocaldococcus jannaschii* are often clustered, possibly due to the lateral transfer of functionally related genes (61) or the loss of clustered functions in the *Methanocaldococcus jannaschii* lineage.

ACKNOWLEDGMENTS

This work was supported by grant GM60403 from the National Institutes of Health, grant NCC 2-1273 from the NASA Astrobiology Institute, and grant DE-FG03-01ER15252 from the Department of Energy’s Microbial Cell Program.

We thank Alberto J. L. Macario and Mona Malz for their contributions to this work.

REFERENCES

5. Bell, S. D., and S. P. Jackson. 1998. Transcription and translation in Ar-
METHANOCOCCUS MARIPALUDIS

Genome sequence of Methanococcus maripaludis

VOL. 186, 2004

GENOME SEQUENCE OF METHANOCOCCUS MARIPALUDIS

6967


15.


14.


13.


12.

Bernander, R.

11.


10.


9.


8.

Berger, P. S., C. Blank, and J. A. Leigh.

7.


6.

Berger, P. S., C. Blank, and J. A. Leigh.

5.

Berghofer, P. S., C. Blank, and J. A. Leigh.

4.

Berger, P. S., C. Blank, and J. A. Leigh.

3.

Berger, P. S., C. Blank, and J. A. Leigh.

2.

Berger, P. S., C. Blank, and J. A. Leigh.

1.

Berger, P. S., C. Blank, and J. A. Leigh.

0.

Berger, P. S., C. Blank, and J. A. Leigh.


