Partial Gene Sequences for the A Subunit of Methyl-Coenzyme M Reductase (mcrI) as a Phylogenetic Tool for the Family Methanosarcinaceae

ERIK SPRINGER,† MATTHEW S. SACHS,2 CARL R. WOESE,3 AND DAVID R. BOONE1,2,4

Department of Environmental Science and Engineering1 and Department of Chemistry, Biochemistry, and Molecular Biology2, Oregon Graduate Institute of Science & Technology, Portland, Oregon 97291-1000, and Department of Microbiology, University of Illinois, Urbana, Illinois 618013

Representatives of the family Methanosarcinaceae were analyzed phylogenetically by comparing partial sequences of their methyl-coenzyme M reductase (mcrI) genes. A 490-bp fragment from the A subunit of the gene was selected, amplified by the PCR, cloned, and sequenced for each of 25 strains belonging to the Methanosarcinaceae. The sequences obtained were aligned with the corresponding portions of five previously published sequences, and all of the sequences were compared to determine phylogenetic distances by Fitch distance matrix methods. We prepared analogous trees based on 16S rRNA sequences; these trees corresponded closely to the mcrI trees, although the mcrI sequences of pairs of organisms had 3.01 ± 0.541 times more changes than the respective pairs of 16S rRNA sequences, suggesting that the mcrI fragment evolved about three times more rapidly than the 16S rRNA gene. The qualitative similarity of the mcrI and 16S rRNA trees suggests that transfer of genetic information between dissimilar organisms has not significantly affected these sequences, although we found inconsistencies between some mcrI distances that we measured and previously published DNA reassociation data. It is unlikely that multiple mcrI isogenes were present in the organisms that we examined, because we found no major discrepancies in multiple determinations of mcrI sequences from the same organism. Our primers for the PCR also match analogous sites in the previously published mcrII sequences, but all of the sequences that we obtained from members of the Methanosarcinaceae were more closely related to mcrI sequences than to mcrII sequences, suggesting that members of the Methanosarcinaceae do not have distinct mcrII genes.

The taxonomy of methanogens is strongly guided by 16S rRNA sequences, in part because methanogens have very limited catabolic pathways and few easily determined phenotypic characteristics. Also, most methanogenic species have been described since the time that 16S rRNA sequences began to be used as a way to determine phylogenetic relationships. This unusual degree of reliance on 16S rRNA data has yielded a taxonomy that is strongly phylogenetic (6), but there are at least two major concerns or limitations associated with the exclusive use of 16S rRNA trees for taxonomy.

The first concern is that the number of genetic differences between closely related species may be small and may not measure small phylogenetic distances with sufficient precision to define species (13). Organisms that exhibit levels of 16S rRNA sequence similarity of up to 98% (at least 30 differences per 1,500 bp) have been considered too dissimilar to be placed in the same species (6, 8, 49, but Stackebrandt and Goebel (39) have proposed that only when the level of 16S rRNA sequence similarity is less than 97% should strains be considered members of separate species without other phylogenetic data, such as DNA reassociation data. However, two distinct but closely related species may exhibit a level of similarity greater than 98% (6, 8, 13, 39, 45), and no one has proposed a level of 16S rRNA sequence similarity above which microbes must be placed in the same species. An analogous benchmark of 70% DNA sequence similarity (and a difference between the homologous denaturation temperature and heterologous denaturation temperature of 5°C) has been proposed as the maximum level of similarity (minimum phylogenetic distance) that separates organisms into distinct species, with the stipulation that separate species should have phenotypic differences that are suitable as diagnostic taxonomic tools (41). Another concern about using comparisons of sequences of a single gene for taxonomy is the possibility of recent transfer of genes between dissimilar species, which would give rise to an organism that appeared to be closely related to the donor of the gene and unrelated to its true close relatives (9, 37). Such a discrepancy might be revealed by a recombination analysis (37) or by a comparison of phenotypic data. Alternatively, independent phylogenetic trees based on DNA reassociation data or genetic sequences unrelated to the ribosomal apparatus can be compared with 16S rRNA phylogenetic trees (23). Such comparisons would likely reveal inaccuracies in the 16S rRNA trees that might be caused by transfer of 16S rRNA genes between dissimilar species.

Phylogenetic trees based on 5S rRNA data (16, 17) and 23S rRNA data (24, 36) compare well with 16S rRNA phylogenetic trees. However, 5S rRNA, 23S rRNA, and 16S rRNA genes are all functionally related, and an independent gene tree would be useful for detecting anomalies. More recently, the gene sequences for the beta subunit of ATPase and for elongation factor Tu showed good agreement with each other (23). The organisms included in these studies were phylogenetically diverse, so the studies would not have been expected to detect genetic exchange among closely related strains. A study of the evolution of niFH genes included closely related organisms, but discrepancies between the trees may have been due to the gene being located on a plasmid (10).
We selected the gene for the A subunit (mcrA) of the methylenecoenzyme M reductase gene cluster for phylogenetic analysis and comparison with the results of a phylogenetic analysis of 16S rRNA genes. Although this reductase is not present outside the methanogens, it is present in all methanogens. The molecular weights of the subunits are consistent with the 16S rRNA phylogeny at the genus level (34), and this large, complex enzyme might have evolved slowly to yield present-day sequences that exhibit enough similarity to construct accurate evolutionary trees. The mcrI sequence of Methanococcus voltae is much more similar to the sequence of Methanococcus vannellii than it is to the sequences of three other, more phylogenetically distinct species (21). Several complete mcrI sequences are available (1, 2, 7, 19, 31, 42) to guide the selection of primers for amplification of sequences by the PCR. Methylcoenzyme M reductase has three distinct subunits (α, β, γ). The genes for these subunits are organized in a single transcription unit, mcrBDCGA, that includes two interspersed additional open reading frames (mcrD and mcrC) whose functions are not known yet. An isoenzyme whose gene sequence
(mcrII) is phylogenetically distinct from the sequence of the more common mcrI (21, 40) present in some methanogens (32); this isoenzyme may occur only in members of the order Methanobacteriales (40).

We selected a relatively conserved region of the mcrA gene from sequences aligned by using their amino acids (43) and determined the corresponding gene sequences for 25 members of the family Methanosarcinaceae. These sequences were analyzed to create phylogenetic trees that were compared with 16S rRNA trees and hybridization data for the same organisms.

(Portions of the results have been described previously [3, 38].)

**MATERIALS AND METHODS**

Source and growth of cultures. Cultures were obtained from the Oregon Collection of Methanogens and were grown at 37°C in the media indicated in Table 1. Methanococcosidaes burtonii was grown at room temperature, and Methanosaurocrina thermophila was grown at 50°C. Strain EBac (= OCM 550 = DSM 8220) was isolated by E. Springer and K.-H. Blotevogel. The culture media used were based on MS medium (4). MS-TMA was MS medium supplemented with 40 mM trimethylamine; MG-H2 was MS medium supplemented with 2.5 g of NaCl per liter and 5 mM acetate and pressurized to 100 kPa with H2; MS-CO2-TMA was MS-TMA with a gas phase consisting of pure CO2; MH-TMA was MS-TMA supplemented with an additional 2 g of NaOH per liter, 1.5 mol of NaCl per liter, 5 g of MgCl2·6H2O per liter, and 20 mmol of KCl per liter (the final pH was 7.3); MSH-TMA was a 2:1 mixture of MS-TMA and MH-TMA; and MSHA-TMA was MSH-TMA equilibrated with an N2 gas phase (final pH, 8.0).

Purification of DNA. Cells were lysed by resuspending pellets in buffer and adding 20% (wt/vol) sodium dodecyl sulfate (final concentration, approximately 0.1 g per liter), and DNA was extracted (26) with chloroform-phenol (1:1).

**PCR amplification and cloning.** Degenerate primers (17 bases) were constructed for two conserved amino acid regions of the mcrA gene. These regions were YDQIWL (amino acids 337 to 343) and NYAMNV (amino acids 503 to 509) (43). The primers were obtained from Oligos Etc., Inc., Wilsonville, Oreg. The nucleotide sequences were TAYGAYCARAGTTG (sense) and ACR TTCTATCGTARFT (antisense). The intervening sequences were amplified with VENT-PCR (New England Biolabs, Beverly, Mass.) in an EIRCO MP EasyCycler apparatus (Ericomp, Inc., San Diego, Calif.) by using a low-temperature program (annealing at 57°C for 30 s, extension at 55°C for 60 s, and denaturation at 94°C for 30 s, repeated for 35 cycles). The reaction volume was 100 µl with 300 ng of genomic DNA. No amplification occurred if the extension temperature was above 60°C or if the extension step was omitted. The low-temperature program usually yielded a single major band that was visible on agarose gels.

PCR products were extracted once with a phenol-chloroform mixture (1:1), precipitated with ethanol, and treated with kinase (T4-PolyNucleotide kinase; US Biochemicals, Cleveland, Ohio). The DNA was precipitated with ethanol and dissolved in 15 µl of H2O; 5 µl of the resulting preparation was added to 10 µl of a ligation mixture containing gelpurified, EcoRV-cut pBluescriptIISK– (Stratagene, La Jolla, Calif.) and T4-Ligase (New England Biolabs) as recommended by the manufacturer (polyethylene glycol was not used). This solution was incubated at 15°C overnight, 1 µl was added to 40 µl of competent cells (DH5alpha or XL-1 Blue), and the preparation was electrotoporated. The cells were then spread on Luria-Bertani plates containing ampicillin (100 µg/liter) and X-Gal (5-bromo-4-chloro-3-indolyl--D-galactopyranoside)- IPTG (isopropyl- β-D-thiogalactopyranoside) for blue-white selection. Clones containing DNA fragments of the expected size were selected by using ultrarapid plasmid mini-preps and the boiling method (35). The fragments were released by brief digestion with EcoRI and HindIII.

Plasmid preparation and sequencing. DNA was obtained from 30-ml overnight cultures by alkaline lysis (35) followed by a final polyethylene glycol precipitation step. The DNA yield from each culture was usually sufficient for sequencing both strands of the whole clone. Sequencing reactions were performed with Sequenase 2.0 (US Biochemicals) with the enzyme concentration reduced to 66% of the recommended concentration and with the protocol modified for use with 10-µl microtiter dishes. The dishes were heated at 92°C for 6.5
min, and the DNA in the samples were separated by electrophoresis in 6% acrylamide gels. Fixation in methanol-acetic acid was omitted. For each methanogenic strain, two clones from different PCR preparations were sequenced in most cases; the exceptions were strains OCM 56, OCM 68\textsuperscript{T} (T = type strain), OCM 70\textsuperscript{T}, OCM 150\textsuperscript{T}, OCM 156, and OCM 158. At least 90% (average, 98%) of all sequence information for each clone was confirmed by studying both strands. Three clones of strain OCM 530 were sequenced, but mostly on just one strand (see GenBank entry for details). We also confirmed the previously published sequence (19) of the analogous fragment in Methanococcus voltaei (OCM 70\textsuperscript{T}) (=DSM 1537\textsuperscript{T}).

Some clones were shortened with EcoRI and S1 nuclease to obtain full double-stranded sequence information. Most clones were sequenced without subcloning by using degenerate primers that annealed to a conserved amino acid sequence near the center of the cloned 490-bp sequence. The conserved amino acid sequence was FGGQQR, and we used one of two degenerate primers, TTYG GWGGNTMCRAG (sense) and CTYTGNANCWCCA (antisense), in combination with promoter primers T3 and T7 from pBluescript-Strategene and the appropriate primer for amplification. For larger overlaps, the PCR primers were replaced with two other degenerate primers that were based on the DIATES region (GAYATHGCHGARTC [sense]) and the DLQDOC region (GCNCRCAYGRTCTTG [antisense]) (obtained from the Center for Gene Research and Biotechnology, Oregon State University, Corvallis). Sequences were edited with the MacVector program (International Biotechnologies, Inc., New Haven, Conn.).

Determination of 16S rRNA sequences. We determined 16S rRNA sequences by the reverse transcriptase method (44).

Phylogenetic analysis. Sequences of 16S rRNA genes were aligned with the program Clustal V (15) called from the GDE editor (20) and checked manually; the mcrl sequences were aligned by hand after the inferred amino acid sequences were aligned. Bootstrap subsets (100 sets) and phylogenetic trees were constructed with the PHYLIP package (11, 12).

RESULTS AND DISCUSSION

Sequence alignment. Most of the partial mcrl sequences of members of the Methanosarcinaceae were 490 bp long; the only exception was the Methanosarcina sp. strain WH-1 sequence, which was 487 bp long. The latter sequence was confirmed with a third clone and was aligned by using a 3-base deletion. The alignment is available in electronic form from our FTP site (ftp.eso.org) as pub/OCM/mcrl.

mcrl isoenes. The sequences of the primers used for amplification of mcrl were identical (antisense primer) or very similar (sense primer) to the sequences of the only previously described isogenes of mcrl, mcrl\textsubscript{I} of Methanothermus fervidus (21) and Methanobacterium thermoautotrophicum \textsubscript{DH} (31), so the degenerate primers should have amplified an mcrl\textsubscript{II} gene if it had been present. However, none of the mcrl genes that we sequenced appeared to be mcrl\textsubscript{II}, which is phylogenetically distinct from mcrl (21, 40). Also, no major differences between clones of the same organism were found. Although it is possible that an isogene of mcrl\textsubscript{I} was present, we did not find one. It is possible that mcrl\textsubscript{II} or some other isogene was present but that it was so distinct that our primers did not recognize it. It is also possible that mcrl\textsubscript{II} or some other distinct isogene was present but rare among members of the Methanosarcinaceae, so that although we sequenced multiple clones, we never selected one containing the alternate sequence. Thus far, mcrl\textsubscript{II} has been detected only in members of the Methanobacteriales, despite attempts to find it in other methanogens, including Methanosarcina barkeri (40).

Phylogenetic tree based on partial mcrl sequences of members of the Methanosarcinaceae. Figure 1 shows the phylogenetic relationships among members of the Methanosarcinaceae based on their partial mcrl sequences. The division of type

FIG. 2. Comparison of Fitch trees based on a 490-bp segment of the mcrl gene (right side) and on 16S rRNA sequences (left side). Abbreviation: M., Methanococcus.
Comparison of phylogenetic trees based on mcrI data and 16S rRNA data. The phylogenetic trees generated from partial mcrI sequences were similar to those generated from 16S rRNA sequences (Fig. 2). Analyses of the two types of sequences resulted in the same division of strains into genera, although there were some minor discrepancies in the relative distances between species of the same genus and some discrepancies in the relative distances of the various genera from each other. We compared the 16S rRNA distance (i.e., the number of differences per nucleotide site) with the mcrI distance for each pair of strains and found that the mcrI distances were 3.01 ± 0.541 times larger (Fig. 3). A linear least-squares fit of the mcrI distance data to the 16S rRNA distance data gave a slope of 2.12 (R² = 0.78) when the line was forced through the origin. However, pairs of organisms belonging to different families had comparatively smaller ratios of mcrI distance to 16S rRNA distance (Fig. 3). A least-squares analysis of the distances between organisms belonging to the same family revealed a slope of 3.03 (R² = 0.71). Figure 3 also shows that the distances between members of different species of the same genus were smaller (the mcrI distances were 0.0937 ± 0.0313, and the 16S rRNA distances were 0.0214 ± 0.0065) than the distances between members of different genera belonging to the same family (the mcrI distances were 0.245 ± 0.038, and the 16S rRNA distances were 0.0807 ± 0.0096). The lack of overlap in these values suggests that within the family Methanosarcinaceae, the assignments of species to genera and genera to family are consistent with evolutionary distances.

Comparison of mcrI distance data with DNA reassociation values. The evolutionary distances computed by using mcrI sequences were also compared with DNA-DNA reassociation data. The mcrI tree is only partially consistent with the DNA reassociation tree for the genus Methanohalophilus (5). The high levels of DNA sequence similarity among FDF-1, FDF-2, Ret-1, and SF-1 were supported by similar mcrI sequences. DNA reassociation values were not always consistent with mcrI distances, however. The mcrI distances indicated that Methanohalophilus halophilus was more similar to Methanohalophilus portucalensis than to Methanohalophilus mahii, whereas DNA reassociation data indicated the opposite (5). DNA reassociation data indicated that strain Cas-1 is distantly related to other Methanohalophilus strains (5), whereas the mcrI sequence data for this strain indicated that it is similar to Methanohalophilus sp. strains Ret-1 and SF-1. Also, according to DNA reassociation data, strains CAS-1 and SD-1 are closely related (5); however, on the mcrI sequence tree, strain SD-1 is phylogenetically related to a different genus, the genus Methanohalobium.

DNA reassociation data were the basis for the proposal to consider Methanosarcina frisia a junior subjective synonym of Methanosarcina mazeii (25). This proposal is supported by 16S rRNA data but not by mcrI data (Fig. 2). The phylogenetic distance between the type strains of these two species may beg the definition of species level phylogenetic distance. Two of the three measures of phylogenetic distance indicate a distance small enough that the two species should be considered synonymous, but the third suggests that they could be considered separate species.

Inconsistencies between 16S rRNA distances and DNA reassociation values also occur in the genus Aeromonas (27). Some strains that have nearly identical 16S rRNA sequences exhibit low levels of DNA reassociation, and in one case two strains that exhibited a high level of DNA reassociation were separated on the 16S rRNA tree (27).

Both 16S rRNA trees and mcrI trees separate species belonging to the Methanosarcinaceae into genera with a high degree of certainty on distance matrix trees (with bootstrap values of 99% or higher). A parsimony analysis of the same 16S rRNA sequences also gave bootstrap values of 100%, and most of the mcrI bootstrap values were 100%; the only exceptions were the genera Methanococcoides (95%) and Methanoblobus (87%). When the bootstrap analyses included additional strains (such as those shown in Fig. 1), the bootstrap values were slightly lower.

ACKNOWLEDGMENTS

We thank Yitai Liu, Jane Boone, and Pierre Rouvière for growing cultures and extracting DNA from some of them. We also thank William B. Whitman for providing cells of Methanosarcina vacuolata and Methanosarcina mazeii S-6 for rRNA sequencing. We thank Don Buchholz for installing and configuring software for sequence analysis at the Oregon Graduate Institute of Science & Technology and Andreas Bartelt for installing software at the University of Oldenburg, and we thank Jim Cregg and Karl-Heinz Blotsewogel for helpful discussions. This work was supported by grant 14-08-001-G2098 from the U.S. Geological Survey, by the U.S. Department of Energy Subsurface Science Program (Deep Microbiology Subprogram) under U.S. Department of Energy grant DE-FG05-90ER1039, and by grant NAGW-2554 from the National Aeronautics and Space Administration. E.S. was supported by stipend 332-4-03-101/515-402-498-3 from the German Academic Exchange Service (DAAD).

REFERENCES

2. Bokranz, M., and A. Klein. 1987. Nucleotide sequence of the methyl coen-

FIG. 3. Scatter plot of 16S rRNA distances versus reductase distances (based on a 490-bp segment). Symbols: □, distances between pairs of strains belonging to the same genus; Δ, distances between pairs of strains belonging to different genera but the same family; δ, distances between pairs of strains belonging to different families. M. fervidus, Methanothermus fervidus; M. thermoaototrophicum, Methanobacterium thermoautotrophicum; M. voltaei, Methanococcus voltaei; M. vannieli, Methanococcus vannieli.


