GENOME-WIDE SURVEY OF GENE FUNCTIONALITY FOR THE METHANOGENIC ARCHAEON *METHANOCOCCUS MARIPALUDIS*

by

Felipe Andres Sarmiento Boban

(Under the Direction of William B. Whitman)

ABSTRACT

Methanogens are obligately anaerobic Archaea that obtain most of their energy required for growth from the biosynthesis of methane, or methanogenesis, a process performed only by these prokaryotes. In addition to this distinctive metabolism, methanogens exhibit many other features unique to this group of microorganisms. As one of the earliest forms of life on our planet, the study of methanogens and their particular physiology, biochemistry and genetics is not only necessary to understand this specific biological group and the Archaea domain but also provides clues about the origin and the evolutionary history of all other forms of life.

A comprehensive whole-genome survey of gene function was performed in the hydrogenotrophic methanogen *Methanococcus maripaludis* by the Tn-seq methodology. About 30% of the genome was classified as possibly essential or strongly advantageous for growth in rich medium. Many of these genes were homologous to eukaryotic genes that encode fundamental processes in replication, transcription and translation, providing direct evidence for a close relationship between Archaea and Eukaryotes in the information processing system. However, some genes classified as possibly essential were unique to the archaeal or methanococcal lineages. Of special interest, the gene polD encoding a unique replicative DNA polymerase found in archaea was essential, indicating that it is the major replicative polymerase. In contrast, the archaeal homolog to the gene encoding the ubiquitous DNA polymerase B was not essential for growth. Interestingly, PolD is absent from the genomes of the crenarchaeotes, suggesting a clear

evolutionary division within the archaeal domain. These results demonstrate a fundamental change in the replication mechanism among domains and suggest an unanticipated variability in archaeal DNA replication.

Additionally, the Tn-seq technology was used to identify unknown genes involved in specific metabolic processes of *M. maripaludis*. Some genes possibly involved in biosynthesis of vitamins and coenzymes and the aromatic amino acids aminotransferase, among others, were detected. To get the complete picture, the next step is to corroborate the involvement of these genes in the proposed pathways by detailed genetic and biochemical studies.

INDEX WORDS: Archaea, methanogens, *Methanococcus maripaludis*, Tn-seq, PolD, DNA replication, aromatic amino acids, coenzyme M, auxotrophs

GENOME-WIDE SURVEY OF GENE FUNCTIONALITY FOR THE METHANOGENIC ARCHAEON *METHANOCOCCUS MARIPALUDIS*

by

Felipe Andres Sarmiento Boban

B.S., Universidad de Santiago de Chile, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

GENOME-WIDE SURVEY OF GENE FUNCTIONALITY FOR THE METHANOGENIC ARCHAEON *METHANOCOCCUS MARIPALUDIS*

by

FELIPE ANDRES SARMIENTO BOBAN

Major Professor: William B. Whitman

Committee: Anna C. Karls
Robert J. Maier
Jan Mrázek
Juergen Wiegel

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2013

DEDICATION

To my wife, Patricia Jorquera for her love, patience and support. To my family: my parents, Ljubica Boban and Ricardo Sarmiento, my sister, Macarena Sarmiento, and my grandmothers Violeta Barria and Lia Leiva.

To my late grandfather Juan Boban. I was not there to say goodbye, but I dedicate my dissertation to you.

I love you all!

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor and friend Dr. William B. Whitman for the tremendous amount of time and energy that he has invested in my training. Tireless and always accessible he has helped me develop into a good scientist, teaching me the necessary skills to move forward, critical thinking, keen observation, and perseverance. His passion for science, great ideas and encouragement are inspirational and were instrumental in aiding me to perform my research. I would also like to acknowledge Dr. Jan Mrázek for his prompt and constant support at any time through my research. I also thank the rest of my doctoral committee: Dr. Anna C. Karls, Dr. Robert J. Maier, and Dr. Juergen Wiegel for their invaluable support and ideas over the course of my research. I also thank Dr. Yuchen Liu for teaching me many of the techniques in the laboratory, tireless support and useful discussions. Furthermore, I thank many past and present members of the Whitman lab, including: Boguslaw Lupa, Magdalena Sieprawska-Lupa, Kamlesh Jangid, Christopher Reisch, Iris Porat, Sarahi Garcia, Nick Galloway, Courtney Ellison, Hannah Bullock, Warren Crabb, Feng Long, Yixuan Zhu and Lv Zhe for their company, friendship, support and encouragement. I also thank my classmates, Laura Cuff, Bryndan Durham, Claire Edwards, Matt Hawkins, Bradley Tolar, Rupal Prabuh and Megan Patch. Finally, I would like to thank the rest of the Microbiology Department, friends and everyone else who make me feel welcome in US and helped me through these years.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1. Archaea

Originally all known living organisms were classified either as prokaryotes or eukaryotes. Then a five kingdoms classification system was proposed, where prokaryotes were still regarded as a single group distinct from the four eukaryotic kingdoms (128). The pioneer work of Carl Woese and collaborators stunned the scientific community with the reclassification of the prokaryotes into two groups, the Archaebacteria and Bacteria (131). Throughout the analysis of prokaryotic rRNA sequences, Woese et al. discovered that phylogeneticaly these sequences can be divided into two lineages different enough to classify them as separate domains. Thus, living organisms were classified into three domains of life, Archaea, Bacteria and Eukarya (132). Thus classification has been validated over years of research in microbiology and comparative genomics. Archaea represent a diverse group of single-cell microorganisms, which exhibit many unique features. These microorganisms normally thrive under extreme conditions of temperature, salinity, pH or pressure. However, some also coexist in moderate environments with Bacteria and Eukarya (18, 27). Archaea are adapted to very diverse types of metabolism, which range from anaerobic and aerobic respiration to fermentation and from photo- and chemo-autotrophy to heterotrophy (101). In addition, Archaea include the only known microorganisms capable of methanogenesis, the anaerobic production of methane, one of the earliest microbial processes identified so far in the geochemical record (14), and which demonstrate the ancient roots of the Archaea domain. In spite of this outstanding diversity, members of the Archaea present specific physiological characteristics that unite them, such as the composition and structure of the

membrane lipids (37). All archaeal membrane phospholipids are synthesized from isoprenoid ethers and glycerol-1-phosphate (G1P). Conversely, membranes of bacteria and eukarya are formed from fatty acid esters linked to the stereoisomer glycerol-3-phosphate (G3P) (36).

Archaea, for the most part, resemble to Bacteria in size and shape, but they present important differences at a genetic, molecular and metabolic level. Remarkable differences are present in the machinery and functionality of the information processing systems. For example, the presence of histones in several members of the Archaea (112), the similarity of proteins involved in transcription and translation, and the structure of the ribosome are more closely related to those of Eukarya (68). The most striking similarities between Archaea and Eukarya are observed in DNA replication, one of the most conserved processes in living organisms, where the components of the replicative apparatus and the overall process closely resemble a simplified version of the eukaryotic DNA replication system (6, 30). However, a few characteristics of the archaeal information processing system are shared only with Bacteria, such as the circular shaped chromosome, the similar average size of the genome, the presence of polycistronic transcription units and Shine-Dalgarno sequences in the mRNA instead of Kozak sequences (36, 76). These observations suggest that archaeal DNA replication is a process played by eukaryotic-like proteins in a bacterial context (35). In addition, there are specific features which belong only to the domain Archaea, such as the DNA polymerase D and other specifics genes, which add another level of complexity and uniqueness to these microorganisms. The mosaic nature of this process in living organisms represents only a part of the complex structural and functional mixture of the three domains of life. Several evolutionary hypotheses have been proposed to account for this mystery (32, 61). The results of comparative genomics and ultrastructural studies have narrowed these hypotheses to two possible alternative scenarios. The archezoan scenario

holds that an archaeal ancestor developed a nucleus and evolved into a primitive archezoan which later acquired an α-protobacterium forming the mitochondria and evolved into eukaryotes. The symbiogenesis scenario holds that an archaeal ancestor plus an α-protobacterium formed a chimeric cell which evolved into the eukaryotic cell (61).

Through the SSU rRNA study of cultivated Archaea this domain is taxonomically divided in five phyla. Crenarchaeota and Euryarchaeota are the best characterized, and this taxonomic division is strongly supported by comparative genomics, as several genes with key roles in chromosome structure and DNA replication are present in one phylum but absent in the other. Crenarchaeotes also exhibit a vast physiological diversity, including aerobes and anaerobes, fermenters, chemoheterotrophs, and chemolithotrophs (18). A vast majority of the crenarchaeotes are thermophiles, with some few exceptions (60). Euryarchaeotes exhibit an even larger diversity, with several different extremophiles among their ranks in addition to mesophilic microorganisms. Interestingly, every methanogen known so far belongs to this domain, where they are separated into six different orders (see below). Additionally, through environmental sequencing, 16s rRNA analysis, and cultivation efforts, three new phyla have been recently proposed: Thaumarchaeota, Korarchaeota and Aigarchaeota (12, 31, 95). A sixth phyla has been proposed previously, Nanoarchaeota, harboring only one microorganism *Nanoarchaeum equitans* with the smallest cellular genome ever sequenced (43). However, a recent report suggests that *N. equitans* may correspond to a fast-evolving euryarchaeal linage rather than the representative of an early archaeal phylum (13).

2. Methanogens

Methanogens are strictly anaerobic archaea which are distinguished by their capacity to produce methane as the end product of their anaerobic respiration and deriving most of the required energy from this process. The first reports of methanogenesis goes back to the Roman Empire,

but the credit for methane discover is given to Alessandro Volta who in 1776 examined samples of a "flammable air" from lake sediments (133). For almost a century, the genesis of methane was attributed only to geochemical processes, but Bechamp in 1868 provided the first clues for a possible role of microorganisms (133). It took almost another century to isolate the first pure cultures of methanogens: *Methanobacterium formicicum* and *Methanosarcina barkeri* (56). Later on, methanogens played a major role in the recognition of the domain Archaea. In 1976, Carl Woese was studying the rRNA of several prokaryotes, among them the methanogen *Methanothermobacter thermoautotrophicus* (formelly known as *Methanobacterium thermoautotrophicum*). The rRNA sequence of this methanogen was different enough to introduce the first clues about a new type of prokaryotes (130).

Methanogens are now known to comprise a large and diverse group, with many isolated strains. Indeed, the genomes of approximately 60 different strains are fully sequenced, on draft, or in process to be sequenced. Taxonomic classification places them as representatives of the domain Archaea belonging exclusively to the phylum Euryarcheota, where they are presently divided in six well-established orders (Methanobacteriales, Methanocellales, Methanococcales, Methanomicrobiales, Methanosarcinales, and Methanopyrales) and 31 genera (75). Phylogenic classification of methanogens based on 16S rRNA sequences revealed that they can be divided into two groups: group I includes the order Methanopyrales, Methanobacteriales, Methanococcales. Group II includes the orders Methanosarcinales, Methanomicrobiales and Methanocellales (5). Further studies of the set of homologous enzymes for the hydrogenotrophic pathway of methane biosynthesis and enzymes for the biosynthesis of the unusual coenzymes involved in this process revealed that those genes are ubiquitous in methanogens and share a common history. The study concluded that hydrogenotrophic methanogenesis appeared only

once during evolution, suggesting an ancient monophyletic origin of methanogens (5). This ancestral form was possibly preserved in two independent lineages (groups I and II) and it was likely lost in the non-methanogenic lineages which are interspersed in the phylogenetic tree of methanogens. These lineages include (Archaeoglobales, Halobacteriales and Thermoplasmatales) (5). Other hypotheses which explain this odd branching of methanogens are (i) methanogenesis related genes were acquired by horizontal gene transfer or (ii) methanogens share a common ancestor with *Archaeoglobus* (75). However, a new study, which used common orthologous genes from eight methanogen species and applied *Thermococcales* species as an outgroup, proposed a new classification scheme where the two groups are maintained but places Methanopyrales and methanomicrobiales together with methanosarcinales in group II rather than with Methanococcales (78).

Methanogens are widely distributed in diverse anaerobic environments. They have been cultivated mainly from temperate habitats, but they are also common in habitats with extreme salinity, temperature and pH. In the absence of more favorable electron acceptors such as sulfate, oxidized metals, and nitrite, methanogens play a key ecological role, catalyzing the last step of the anaerobic mineralization of organic molecules (Figure 1-1). In addition, methanogens maintain an extremely low partial pressure of H_2 and remove the excess of other fermentation products, keeping anaerobic fermentative pathways energetically favorable (39).

Methanogenesis is a type of anaerobic respiration which uses a limited number of substrates. Most methanogens are hydrogenotrophs, which reduce CO_2 to CH_4 using hydrogen or formate as electron donors. However, some representatives of the order Methanosarcinales are methylotrophs or acetotrophs, which uses methyl compounds or acetate as substrates, respectively. Although these pathways start differently, the last step is driven by the same

enzyme, methyl-coenzyme M reductase, which is ubiquitous in all methanogens. In addition, specific coenzymes play major roles in methanogenesis functioning as C-1 moiety carriers, which are unusual outside this group of microorganisms. Although, methanogens obtain energy from methane production, this is an energetically poor life-style. Indeed, cells derive less than one ATP equivalent from each molecule of methane produced (1).

Methanogenesis plays a significant role in the global carbon cycle, processing about 1.7% of the carbon fixed every year by photosynthesis (39). Methane, which is a potent greenhouse gas, is directly released to the atmosphere or oxidized by aerobic and anaerobic methane-oxidizers. The current annual global emission of methane to the atmosphera is 500-600 Tg and about 80% of that is microbially produced (39).

3. General overview of the DNA replication process

DNA is a major building block of life, and all known living organisms are adapted for its use. The precise and timely duplication of DNA is achieved by DNA replication, a complex process which involves dozens of specialized proteins. The process itself is conserved in the three domains of life with some modifications, but key proteins vary among them. Both strands of the double strand of DNA (dsDNA) are copied at the same time by the same unidirectional replication machine, but because of the antiparallel nature of the DNA, the leading strand is copied continuously and the lagging strand is copied discontinuously as Okazaki fragments (35). DNA replication starts when specific proteins recognize and bind the origin of replication, forming a protein-DNA complex, This complex recruits a helicase to unwind the double strand DNA (dsDNA) by using ATP. The single strand DNA (ssDNA) is protected by ssDNA-binding proteins (SSB), and the main core of the replication machinery, composed of DNA primase and DNA polymerase, is recruited. A sliding-clamp protein surrounds the DNA and binds to the

polymerase to increase processivity. The discontinuous replication in the lagging strand is complemented by the actions of RNase, DNA ligase and Flap endonuclease.

4. DNA replication components

Origin of replication. The number of origins of replication in a genome is correlated with phylogeny and varies among the different domains of life (4). Members of the Bacteria usually possess only one replication origin, but eukaryotic chromosomes contain multiple replication origins. This difference was generally accepted as a clear divisor in DNA replication of eukaryotes and prokaryotes. However, members of the Archaea display either one or multiple origins of replication. The origins of replication in archaea are commonly AT-rich fragments which possess conserved sequences called origin of recognition boxes (ORB) and are well conserved across many archaeal species. Also, smaller versions of the ORBs, called mini-ORBs, have been identified (108). Members of the Crenarchaeota phylum display multiple origins of replication. For example, species belonging to the Sulfolobales order contain three origins of replications (77, 108), which are believed to serve purposes other than simply shortening the time required for replication, such as counteract DNA damage (2). In addition, two and four origins of replication have been found in members of the Desulfurococcales (*Aeropyrium pernix*) and Thermproteales (*Pyrobaculum calidifontis*), respectively (104, 107). Members of the Euryarchaeota phylum seem to possess only one origin of replication. For example, species belonging to the Thermococcales (*Pyrococcus abyssi*) and Archaeoglobales (*Archaeoglobus fulgidus*) orders contain only one origin of replication (79, 86, 90). However, members of the order Halobacteriales (*Halobacterium* sp. NRC-1 and *Haloferax volcanii*) possess numerous origins of replication (93, 138). To date, no methanogen strating DNA replication origins has been experimentally demonstrated, with the exception of *Methanothermobacter thermautotrophicus* (80). However, bioinformatics analysis by GC-skew in *Methanosarcina*

acetivorans (34) and Z-curve analysis for *Methanocaldococcus jannaschii* and *Methanosarcina mazei* suggest a simple origin of replication. However, a similar analysis of *M. maripaludis* S2 was inconclusive (137). The fact that the order *Halobacteriales* is the only euryarchaeote that displays multiple origin of replication, suggest that there may have been a lineage-specific duplication of the origin of replication in this group.

Origin recognition, DNA unwinding and primer synthesis. The origin of replication is recognized by specific proteins that bind and melt the DNA and assist in the loading of the replicative helicases. In bacteria this protein corresponds to DnaA, which binds to the DnaA boxes (82). In eukaryotes, a protein complex known as ORC (origin recognition complex) composed of 6 different proteins (Orc1-6) binds at the replication origin and recruits other proteins (6). In archaea the candidates for replication initiation are the homologs of the eukaryotic Orc1 and Cdc6 (cell division cycle 6) encoded by almost all archaeal genomes (84). Their genes are commonly located adjacent to the origin of replication. These proteins have been studied in detail in representatives of the Euryarchaeota (*Pyrococcus*) and Crenarchaeota (*Sulfolobus*) and have been shown to bind to the ORB region with extreme specificity (86, 108). In addition, purified Cdc6-1 from *S. solfataricus* binds to ORB elements from other crenarachaeotes and euryarchaeotes *in vitro*, suggesting that these proteins recognize sequences conserved across the archaeal domain (108). The crystal structure of these proteins has been solved in *Pyrobaculum aerophilum* (73) and *Aeropyrum pernix* (118), which contributed to the overall understanding of the activity of this replication initiation proteins (reviewed in (48)). The number of Cdc6/Orc genes found in archaeal genomes varies from one in several euryarchaeotes (*Pyrococcus*, *Thermococcus*, *Picrophilus*) and two crenarchaeotes (*Crenarchaeum* and *Pyrobaculum*) to seventeen in the euryarchaeote *Haloarcula marismortui* (6). Interestingly,

methanogen representatives of the orders *Methanococcales* and *Methanopyrales* do not possess homologs for the Cdc6/Orc protein (6, 40), but representatives of the orders *Methanobacteriales, Methanosarcinales, Methanomicrobiales and Methanocellales* possess between two to three copies of this gene, suggesting an important variation in the replication initiation process. The mechanics of replication initiation and the players involved in representatives of the orders *Methanopyrales* and *Methanococcales* remain as an intriguing enigma of the archaeal replication process.

After recognition proteins bind to the origin of replication, they recruit a helicase to unwind the double helix of DNA. In bacteria the replicative helicase is DnaB. In eukaryotes the most accepted candidate for the replicative DNA helicase is the MCM complex (minichromosome maintenance complex), which is a heterohexamer that is activated when associated with other replicative proteins, forming the CMG complex (9, 46, 63, 121). Most archaeal genomes studied so far encode at least one MCM homologue. Its helicase activity has been demonstrated *in vitro* for several archaea, includng *Methanothermobacter thermoautotrophicus* ΔH, where the MCM proteins arrange in a double homohexamer configuration (53, 114). Also, the *in vivo* interaction between the MCM complex and the Cdc6/Orc proteins has been demonstrated in other archaeas (26, 51, 116). Interestingly, some archaea possess more than one homologue of the MCM protein. A recent study identified thirteen species of archaea with more than one *mcm* gene (62). The number of MCM protein homologues is especially high in the order Methanococcales, where different representatives possess from 2 to 8 copies (125, 126). For instance, *Methanococcus maripaludis* S2 possess four homologs of the MCM protein (40). Through a shotgun proteomic study, peptides for three of them have been detected (134), suggesting that multiple MCMs are functional in *M. maripaludis*.

However, our results through a genome–wide survey of gene functionality in *M. maripaludis* demonstrated that only one of these genes, MMP0030, was essential for growth (113). Similar results have been found in *Thermococcus kodakarensis*, in which three MCM homologs are present, but only one is essential (103). In addition, coexpression of recombinant MCMs from *M. maripaludis* S2 allowed co-purification of all four proteins (125), suggesting that *M. maripaludis* may form heteromeric complex in which MMP0030 predominates. A recent study has proposed that two of the MCM homologues that are conserved among representatives of the order *Methanococcales* are a consequence of an ancient duplication that occurred previous to the divergence of the different methanococcales genera (125). It has also been proposed that the increased number of MCM homologues in the order *Methanococcales* is a direct consequence of mobile elements (62). Indeed, it has been hypothesized that mobile elements may have taken advantage of the ancient duplication of the MCM genes to take over the replication system by forming cellular MCM heterocomplex (62). In any case, the large number of MCM homologues in the order *Methanococcales* maybe product of an intricate and complex evolutionary history. Whether or not it is related to the absence of the replication initiation protein Cdc6/Orc is an interesting possibility (62).

In eukaryotes the MCM complex is not active on its own and requires the association of two accessory factors, the heterotetrameric GINS complex (from the Japanese go-ishi-ni-san meaning 5-1-2-3, after the four subunits Sld5 and Psf1-3) and the CDC45 protein. This complex, called CMG (Cdc45/MCM/GINS), is thought to be the active replicative helicase (46). Homologues for GINS subunits have been identified in archaea by bioinformatics analysis. One homolog is most similar with the eukaryotic proteins Psf2 and Psf3 (GINS23) and is common in the crenarchaeotes. Another homolog is most similar to the eukaryotic proteins Sld5 and Psf1

(GINS15) and is found largely in the euryarchaeotes (81). Representatives of the crenarchaeota (*S. solfataricus*) and the euryarchaeota (*P. furiosus* and *T. kodakarensis*) possess both homologs and form heterotetramers similar to the ones found in eukaryotes with a ratio 2:2 (83, 102, 136). Although, the GINS complex is expected to be essential for DNA replication initiation, no stimulation of the MCM helicase activity was obtained in *S. solfataricus* when adding GINS complex *in vitro*, although the proteins did bind each other (83). In contrast, the helicase activity of *P. furiosus* is clearly stimulated by the GINS complex (136). The crystal structure of the GINS complex of *T. kodakarensis* was recently determined. The backbone structure and the assembly are similar to the human complex, but some notable differences are present (102). Interestingly, many other euryarchaeotes possess only one homolog to GINS15. One of those, *Thermoplasma acidophilum* forms a homotetramer complex (48). Moreover, i*n vitro* the *T. acidophilum* MCM helicase activity was not affected when the GINS complex was added (96). These results suggest that other proteins may be involved in the formation of a stable helicase in many archaea. Not much is known about the GINS complex in methanoarchaea. Like other Methanococcales, *M. maripaludis* S2 possesses a single copy of the gene encoding for GINS15, which is essential for growth (113, 126). In addition, no copies of GINS23 are present, which has only been detected in crenarchaeotes and the euryarchaeote order Thermococcales (8).

The replication related Cdc45 protein is ubiquitous in eukaryotes, but its exact role has not yet been elucidated. Interestingly, homologues of the Cdc45 protein are not present in archaea. However, a bioinformatics analysis revealed that the eukaryotic Cdc45 and the prokaryotic RecJ, which is a conserved 5'-3'exonuclease in most bacteria and archaea, possess a common ancestry (111). These results suggest that both proteins are orthologues derived from the last common ancestor of cellular life and they share a homologous DHH domain. Currently,

it is not known if the archaeal homologues of RecJ are involved in DNA replication initiation, but a few studies have shed some light in the role of this protein. In *S. solfataricus*, a homologue of the DNA binding domain of RecJ was purified together with a GINS protein (83). Similarly, in *T. kodakarensis*, a homolog of RecJ (TK1252p) was co-purified with proteins of the GINS complex (70) and formed an stable *in vitro* association with the GINS complex (69). Among methanococcales, every sequenced genome possesses at least one homolog to the RecJ protein. Two RecJ homologs found in *M. jannaschii* (MJ0977 and 0831) partially complemented a *recJ* mutation in *E. coli*. The recombinant MJ0977 also possessed high levels of thermostable single strand DNase activity similar to RecJ (105). *In M. maripaludis* S2 four proteins possess the DHH domain with similarity to the *M. jannaschii* RecJ homologs (MMP1682, 0547, 1078 and 1314). However none of them was essential for growth (113). Thus, it is possible that RecJ activity is redundant and encoded by multiple genes. Alternatively, another protein, which has a low similarity with *M. jannashii* RecJ homolog (MMP0285) but possess the DHH domain, was essential for growth (113). This protein possess other domains related to transport, which make it an unlike candidate for RecJ. However, these results are not definitive and more experimental data has to be obtained to assign a function to this protein.

As soon as the DNA is unwound, single strand DNA is protected from nucleases and chemical modification by single-stranded DNA binding proteins (48). These proteins are present in the three domain of life and are called SSB in bacteria, where they arrange as homotetramers or homodimers (24, 129), and RPA (replication protein A) in eukaryotes, where they form a stable heterotrimer composed of 70, 32 and 14 KD proteins (45). Although, all single-stranded binding proteins contain different combinations of the oligonucleotide/oligosaccharide- binding domain or OB fold (89), the sequence similarity is low among RPA and the bacterial SSBs (48).

In archaea, different single-stranded binding proteins types have been observed. In crenarchaeotes, single-stranded binding proteins are only well characterized in *S. solfataricus*. They contain only one OB fold and can adopt a heterotetramer conformation (38). However, the protein also appears to exist as a monomer (124). The crenarchaeal single-stranded binding proteins, with one OB fold, appears to be more similar to the bacterial proteins, but its structure is more similar to the eukaryotic RPAs (48, 55). In contrast, RPAs have been studied from different representatives of the euryarchaeotes, revealing an interesting diversity. *M. jannaschii* and *M. thermoautotrophicus* possess a single subunit RPA, which shares amino acid similarity with the eukaryotic RPA70 and possesses four or five OB folds (52, 54) *P. furiosus* possesses three different RPA subunits, which forms a stable heterotrimer that is involved in homologous DNA recombination *in vitro* (58). *Methanosarcina acetivorans* also possesses three RPA proteins (RPA1-3), which possess four, two and two OB folds, respectively. However, they do not interact, and they function as homodimers (106). The genome of *M. maripaludis* S2 possesses three RPA homologues (MMP0122, 0616 and 1032) (126). However, only MMP0616 and 1032 were essential for growth (113), suggesting a possible different complex configuration for RPA proteins in this microorganisms. It has been hypothesized that the diversity in OB folds in the archaeal RPAs is a direct consequence of homologous recombination (71).

DNA synthesis starts with the production of a RNA primer. This primer is synthesized by a DNA-dependent RNA polymerase or primase. DnaG, a single subunit protein, is the primase for bacteria. In eukaryotes, a two subunit primase, consisting of a small catalytic subunit (PriS) and a large subunit (PriL), is found in a complex with DNA polymerase α and polymerase B subunit (84). In archaea, the first primase found was the eukaryotic-like small subunit PriS in *M. jannaschii* (28). Later, homologs for the PriS and PriL subunits were described in several

Pyrococcus species (10, 49, 67, 74, 85). The eukaryotic-like DNA primase has also been characterized in the crenarchaeote *S. solfataricus* (65, 66), which it has been shown to interact with MCM through GINS23 (83). A unique polymerization activity across discontinuous DNA templates has been characterized in the PriSL of *S. solfataricus*, suggesting that this primase may be involved in double-stranded break repair in Archaea (42). So far, the eukaryotic-like primases found in Archaea exhibit similar properties: PriS functions as the catalytic subunit and PriL modulates its activity (48). Interestingly, the archaeal primase small subunit resembles DNA polymerases from the Pol X family in sequence and structure, suggesting a similarity in their catalytic mechanism (42). In addition, the eukaryotic-like primase has the unique ability to synthesize DNA and RNA (6).

Homologs of the bacterial DnaG primase are also found in archaea. However, *in vitro* studies of the *P. furiosus* enzyme failed to detect primer synthesis activity. Instead, the DnaG participates in RNA degradation as part of the exosome complex (48). The *T. kodakarensis* enzyme co-purified with proteins of the exososme (70). In contrast, the *S. solfataricus* DnaG homolog has limited primer synthesis activity, and a dual primase system with PriSL has been proposed to function during DNA replication (139). Hu et al. hypothesized an interesting theory which suggests that LUCA employed a dual-primase system (DnaG and PriSL) which also served roles in RNA degradation and the nonhomologous end joining (NHEJ) pathway. The system was inherited by Archaea. However, DnaG became the only primase in Bacteria, but was lost in Eukarya. Meanwhile, PriSL became the only primase in Eukarya and evolved into the DNA Pol X family involved in NHEJ. In Bacteria, PriSL evolved into the Pol domain of LigD involved in NHEJ (42).

M. maripaludis S2 also possesses both eukaryotic- and bacterial-like primases. A recent genome-wide survey of this methanoarchaea revealed that genes encoding both subunits of the eukaryotic like primase, PriS and PriL (MMP0009 and MMP0071) were essential for growth, which is consistent with a role in replication. In contrast the DnaG primase (MMP1286) was nonessential (113). These results suggest that in *M. maripaludis* and probably other euryarchaeotes, a dual primase system is not in place.

Elongation (DNA polymerase). The primer is further extended by a DNA polymerase. Seven families of DNA polymerases have been described so far by amino acid sequence similarity (A, B, C, D, E, X, and Y) (48). In bacteria, *E. coli* possesses five DNA polymerases, Pol I-V, where the three first belong to the families A, B, C respectively, and the last two belong to the family Y. The major replicative DNA polymerase in *E. coli* is PolIII, which synthesize DNA with a high processivity (135). In eukaryotes, a great diversity of DNA polymerases is found, but DNA replication requires three DNA polymerases which belong to the family B (Polα, Polδ, and Polε) (44). In archaea, fewer types of DNA polymerases are present, but they have an interesting evolutionary division. Different representatives of the DNA polymerase B-family are ubiquitous among archaea, but DNA polymerase family D is present in every phyla studied so far with the exception of Crenarchaea. In addition, a third family, DNA polymerase Y has been identified in *S. solfataricus* (11).

Representatives of the DNA polymerase family B has been identified in all member of Archaea. The crenarchaeota possess both PolBI and PolBII. The recombinant *S. solfataicus*, *Pyrodictium occultum* and *A. pernix* enzymes have been characterized (15). In contrast, euryarchaeota only contains PolBI. The recombinant *P. furiosus* enzyme has been characterized (122). The PolBI enzymes have similar amino acid sequence and overall structure with a potent

3’-5’exonuclease proofreading activity (15). A unique property of the archaeal family PolB is the ability to stall DNA polymerization in the presence of uracil or hypoxanthine (21, 22). The presence of uracil in DNA is harmful. Commonly, cytosine deamination converts G:C base pairs into the pro-mutagenic G:U mismatches which result in 50% of the offspring containing a A:T transition mutant after replication (72). Interestingly, in some crenarchaeotes and euryarchaeotes, one of the B family DNA polymerase paralogs possess disrupted versions of the sequence motifs that essential for the catalytic functions, suggesting these enzymes possess a structural role (109).

DNA polymerase family D (PolD) is a novel enzyme that was originally discovered in *P.furiosus* (47). It has been further identified in all euryarchaeotes (16). For a long time, this enzyme was considered a euryarchaeote-specific polymerase, but the three newly discovered archaeal phyla also possess genes for PolD (12, 31, 95). Polymerase D is a heterodimer with a small subunit (DP1) and a large subunit (DP2). It has been proposed that the large subunit harbor the polymerase activity, although its sequence is very different from the other DNA polymerases. The small subunit possesses high similarity with the noncatalytic B-subunits of eukaryotic DNA polymerases α, δ, and ε (3). Studies of the *M. jannaschii* enzyme demonstrated that the small subunit possesses a strong 3’-5’exonuclease activity, suggesting that it may be involved in proofreading activity (6, 50). In addition, efficient polymerase and proofreading activity has been detected from a purified PolD of *P. furiosus*, and the residues Asp-1122 and Asp-1124 are essential for polymerization reaction in *P. horikoshii* (115, 123).

The family B DNA polymerase is expected to be essential for replication in crenarchaeotes. However, in euryarchaeotes it is not known which enzyme is involved in replication. A recent study in *Halobacterium* sp. NRC-1 showed that both PolB and PolD are essential for viability, and it has been proposed that they could be working together at the

replication fork synthesizing the leading and lagging strand (7, 110). However, a recent report demonstrated that both subunits of PolD were essential, but PolB was nonessential for the growth of *M. maripaludis* S2, suggesting that PolB does not play a critical role in replication (113). Likewise, in *T. kodakarensis* PolD can be co-isolated with different proteins of the archaeal replication fork, but PolB was mainly co-isolated with proteins of unknown function (70). A deletion of the PolB in *T. kodakarensis* also had no detectable effect on viability. Interestingly, PolB mutants have increased sensitivity to UV radiation, suggesting that PolB may be involved in DNA reparation (23). Mutants of the PolD genes could not be isolated, suggesting that PolD is essential. Finally, it has been demonstrated that the PolD from *P. abyssi* is able to perform RNA primer extension and RNA strand displacement, but PolB cannot, suggesting that PolD could be required for elongation on both strands (41).

These results indicate that PolD is the essential replicative polymerase in the phylum Euryarchaea, which suggest a fundamental change in the replication mechanism between crenarchaeotes and euryarchaeotes, and an unanticipated variability in archaeal DNA replication. The observed differences in replication among Archaea phyla could be a direct consequence of a division at an early stage of evolution, before the replication system was fully formed. Thus, PolD and PolB were present in the common ancestor of all Archaea and later lost in the Crenarchaea branch. Alternatively, PolD may have evolved relatively late, after the division of Euryarchaea and Crenarchaea.

Other accesory proteins. Purified DNA polymerases possess low processivity; however, the addition of an accessory factor, the sliding clamp, gives DNA polymerases the required processivity to replicate genomic DNA. This factor, whose structure resembles a doughnut, interacts with DNA and the polymerase to stabilize it during replication. Its structure is

conserved in the three domains of life. In bacteria, the sliding clamp or β-clamp is a homodimeric ring (59). In eukaryotes the sliding clamp or proliferating cell nuclear antigen (PCNA) is a homotrimeric ring (48). In archaea, the majority of crenarchaeotes have multiple PCNA homologs which form heterotrimeric rings (25, 29). In contrast, most euryarchaeotes possess a single PCNA homolog (6). Interactions of DNA polymerase with the sliding clamp are mediated through a common motif, called the PCNA Interacting Protein (PIP) box (127). The three dimensional structures of PCNA-PolB and the PolB-PCNA-DNA complex from *P. furiosus* have been solved, shedding light of the interactions between these molecules (88, 92). In addition, a recent report demonstrated that the interaction between PCNA, PolD and PolB requires two and one PIP motifs, respectively (17). *T. kodakarensis*, is the only well documented example of an euryarchaeote with two homologues for PCNA, TK0535 and TK0582. Recent work demonstrated that both proteins form stable homotrimeric rings that interact with *T. kodakarensis* PolB *in vitro* (64). It has been proposed that one of the PCNA genes was acquired by lateral gene transfer (33). *M. maripaludis* S2 also possesses two PCNA homologues (MMP1126 and 1711) (126) Both genes were essential for the growth, suggesting that they would play an important role in replication (113). The gene MMP1711 possesses high similarity to both *T. kodakarensis* genes. However, MMP1126 possesses only low similarity to them. Moreover, it has an S-adenosylmethionine-dependent methyltransferases domain, which is unexpected for a PCNA.

For PCNA to assemble around DNA, a specific loading factor, replication factor C (RFC), is required. The eukaryotic RFC is a heteropentameric complex comprising five different subunits (RFC1-5) (117). In contrast, in archaea the RFC consist of two different proteins, a small subunit RFCS and a large subunit RFCL. They also form a heteropentameric complex in a

4:1 ratio (RFCS:RFCL). Structural and biochemical studies of RFC have been conducted in several euryarchaeotes, such as *Archaeoglobus fulgidus* and *Pyrococcus* specie (6). An interesting case is observed in *Methanosarcina acetivorans*. This RFC possesses three subunits (RFC1, 2, L) found in a ratio 3:1:1 (19). The organization and spatial distribution exhibit a similarity to the *E. coli* minimal γ-complex, but the function of the subunits is probably not conserved (20).

During DNA replication, the lagging strand is discontinuously synthesized by extending the RNA primers or Okazaki fragments. In archaea, during replication, Okazaki fragments were first demonstrated in *P. abyssi* and *Sulfolobus acidocaldarius* (87). During replication, these RNA primers are replaced with DNA. The removal of the primers is performed by the enzyme RNase H, which is ubiquitous in the three domains of life. According to sequence similarity, in prokaryotes (Archaea and Bacteria) RNases H are classified into three groups: RNase HI, HII, and HIII. In eukaryotes, they are classified as RNase H1 and H2 (57). However, by phylogenetic analysis using amino acid sequences, RNases H have been further classified into 2 groups: Type 1 (prokaryotic RNase HI, eukaryotic RNase H1, and viral RNase H) and Type 2 (prokaryotic RNase HII and HIII, and eukaryotic RNase H2) (98). These two types of RNase H possess different specific activities, metal ion preferences and cleavage sites (97). Originally it was believed that archaea only possess type II RNases H (98), but several type I archaeal RNase H enzymes have since been discovered (99, 100). Interestingly, in eukaryotes, RNase H1 and H2 tend to coexist. However, different combinations of the three prokaryotic RNases H (I, II and III) are found, except for the combination of RNases HI and HIII. This mutually exclusive evolution seems to be related to functional redundancy (57).

A third enzyme involved in primer replacement in DNA replication is the Flap endonuclease I (FEN-1), which recognizes double-stranded DNA with an unannealed 5′-flap, and cleavages it. Eukaryotic homologs of FEN-1 are found in many archaeal members (48).

M. maripaludis S2 possesses genes for the prokaryotic RNase HI and HII, and for FEN-1, but none are essential, suggesting that they may be redundant with each other (40, 113). Possibly, both RNases H persist and are evolutionary stable in the genome, because the genes perform the same function but one is less efficient than the other (94).

After the Okazaki fragments are replaced by DNA in the lagging strand, the nick between the newly synthesized and the elongated DNA is repaired by a DNA ligase. This enzyme use a nucleotide cofactor to catalyze the formation of the phosphodiester bond in three well characterized steps (120). The enzyme is common to all three domains, but can be grouped into two families based on cofactor specificity (ATP or NAD^+). The DNA ligases from several crenarchaeaoteotes and euryarchaeotes have been further characterized (48). Many archaeal DNA ligases possess dual cofactor specificity (ATP/NAD^+ or ATP/ADP), but every archaeal DNA ligase characterized so far uses ATP. A thermophilic DNA ligase from *M. thermoautotrophicus* was characterized *in vitro* which use ATP as sole cofactor (91). A recently characterized DNA ligase from the chrenarchaeotes *Sulbococcus zilligii* displayed specificity for three cofactors (ATP/NAD^+/GTP) (119). Major structural work in this enzyme has been achieved in *P. furiosus* and *S. solfataricus*, which has been detailed reviewed (48). In *M. maripaludis* S2 one gene that code for a DNA ligase has been identified, which is completely essential for the growth of this archaeon (40, 113).

5. Conclusions

DNA replication in Archaea possess a dual nature, where the machinery is structurally and functionally similar to the eukaryotic replication system, but it is executed in a bacterial mode

within a bacterial context (32, 90). Additionally, unique archaeal features demonstrate the complexity of this process and that it is not just a simplified version of the eukaryotic system. For example, the archaeal specific DNA polymerase D is conserved across the Archaea domain with the exception of the Crenarchaea phyla. The recent discovery that it is the essential replicative polymerase in two different euryarchaeal species demonstrates an unanticipated variability in archaeal DNA replication. Interestingly, the absence of PolD in crenarchaeotes is not the only difference in DNA replication. These differences include the absence of histones in crenarchaeotes (112), the presence of multiple origins of replication in Crenarchaeotes and a single one in most euryarchaeotes, the absence of GINS23 in most euryarchaeotes, and the presence of one homologue of PCNA in euryarchaeotes compared to the multiple homologs in crenarchaeotes. These distinctive characteristics between the phyla highlight the complexity of archaeal DNA replication and its complex evolutionary history (Table 1-1).

Among the Euryarchaea, the methanococcales possess several differences in the DNA replication system that make them unique. For example, they lack homologs for the Orc/Cdc6 proteins, suggesting a unique mechanism for initiation of replication. In addition, *Methanococcales* possess a large number of MCM protein homologs. It has been proposed that these distinctive features are connected, and because of the absence of the Cdc6/Orc proteins, MCM proteins may be interacting with other unknown initiation enzymes, which derive in the observed complex phylogeny of the multiple MCM homologs. Presumably, these differences also account for the inability to recognize the origin of replication in these microorganisms. This unique scenario for initiation of DNA replication may be a direct consequence of various processes such as duplication, mobile elements and viruses interactions (62), which affected the

last common ancestor of Methanococcales in an early evolutionary stage and after diversification into the different known families and species, suggesting a very complex evolutionary history.

References

1. **Amend, J. P., and E. L. Shock.** 2001. Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and bacteria. FEMS Microbiol Rev **25:**175-243.
2. **Andersson, A. F., E. A. Pelve, S. Lindeberg, M. Lundgren, P. Nilsson, and R. Bernander.** 2010. Replication-biased genome organisation in the crenarchaeon *Sulfolobus*. BMC Genomics **11:**454.
3. **Aravind, L., D. D. Leipe, and E. V. Koonin.** 1998. Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Nucleic Acids Res **26:**4205-13.
4. **Aves, S. J.** 2009. DNA replication initiation. Methods Mol Biol **521:**3-17.
5. **Bapteste, E., C. Brochier, and Y. Boucher.** 2005. Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. Archaea **1:**353-63.
6. **Barry, E. R., and S. D. Bell.** 2006. DNA replication in the archaea. Microbiol Mol Biol Rev **70:**876-87.
7. **Berquist, B. R., P. DasSarma, and S. DasSarma.** 2007. Essential and non-essential DNA replication genes in the model halophilic Archaeon, *Halobacterium* sp. NRC-1. BMC Genet **8:**31.
8. **Berthon, J., D. Cortez, and P. Forterre.** 2008. Genomic context analysis in Archaea suggests previously unrecognized links between DNA replication and translation. Genome Biol **9:**R71.

9. **Bochman, M. L., and A. Schwacha.** 2009. The Mcm complex: unwinding the mechanism of a replicative helicase. Microbiol Mol Biol Rev **73:**652-83.

10. **Bocquier, A. A., L. Liu, I. K. Cann, K. Komori, D. Kohda, and Y. Ishino.** 2001. Archaeal primase: bridging the gap between RNA and DNA polymerases. Curr Biol **11:**452-6.

11. **Boudsocq, F., S. Iwai, F. Hanaoka, and R. Woodgate.** 2001. Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta. Nucleic Acids Res **29:**4607-16.

12. **Brochier-Armanet, C., B. Boussau, S. Gribaldo, and P. Forterre.** 2008. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the *Thaumarchaeota*. Nat Rev Microbiol **6:**245-52.

13. **Brochier, C., S. Gribaldo, Y. Zivanovic, F. Confalonieri, and P. Forterre.** 2005. Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to *Thermococcales*? Genome Biol **6:**R42.

14. **Canfield, D. E.** 2006. Biochemistry: gas with an ancient history. Nature **440:**426-7.

15. **Cann, I. K., and Y. Ishino.** 1999. Archaeal DNA replication: identifying the pieces to solve a puzzle. Genetics **152:**1249-67.

16. **Cann, I. K., K. Komori, H. Toh, S. Kanai, and Y. Ishino.** 1998. A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. Proc Natl Acad Sci U S A **95:**14250-5.

17. **Castrec, B., C. Rouillon, G. Henneke, D. Flament, J. Querellou, and J. P. Raffin.** 2009. Binding to PCNA in Euryarchaeal DNA Replication requires two PIP motifs for DNA polymerase D and one PIP motif for DNA polymerase B. J Mol Biol **394:**209-18.

18. **Chaban, B., S. Y. Ng, and K. F. Jarrell.** 2006. Archaeal habitats--from the extreme to the ordinary. Can J Microbiol **52:**73-116.

19. **Chen, Y. H., S. A. Kocherginskaya, Y. Lin, B. Sriratana, A. M. Lagunas, J. B. Robbins, R. I. Mackie, and I. K. Cann.** 2005. Biochemical and mutational analyses of a unique clamp loader complex in the archaeon *Methanosarcina acetivorans*. J Biol Chem **280:**41852-63.

20. **Chen, Y. H., Y. Lin, A. Yoshinaga, B. Chhotani, J. L. Lorenzini, A. A. Crofts, S. Mei, R. I. Mackie, Y. Ishino, and I. K. Cann.** 2009. Molecular analyses of a three-subunit euryarchaeal clamp loader complex from *Methanosarcina acetivorans*. J Bacteriol **191:**6539-49.

21. **Connolly, B. A.** 2009. Recognition of deaminated bases by archaeal family-B DNA polymerases. Biochem Soc Trans **37:**65-8.

22. **Connolly, B. A., M. J. Fogg, G. Shuttleworth, and B. T. Wilson.** 2003. Uracil recognition by archaeal family B DNA polymerases. Biochem Soc Trans **31:**699-702.

23. **Cubonova, L., T. Richardson, B. W. Burkhart, Z. Kelman, B. A. Connolly, J. N. Reeve, and T. J. Santangelo.** 2013. Archaeal DNA polymerase D but not DNA polymerase B is required for genome replication in *Thermococcus kodakarensis*. J Bacteriol. doi:10.1128/JB.02037-12

24. **Dabrowski, S., M. Olszewski, R. Piatek, and J. Kur.** 2002. Novel thermostable ssDNA-binding proteins from *Thermus thermophilus* and *T. aquaticus*-expression and purification. Protein Expr Purif **26:**131-8.

25. **Daimon, K., Y. Kawarabayasi, H. Kikuchi, Y. Sako, and Y. Ishino.** 2002. Three proliferating cell nuclear antigen-like proteins found in the hyperthermophilic archaeon Aeropyrum pernix: interactions with the two DNA polymerases. J Bacteriol **184:**687-94.

26. **De Felice, M., L. Esposito, B. Pucci, M. De Falco, M. Rossi, and F. M. Pisani.** 2004. A CDC6-like factor from the archaea Sulfolobus solfataricus promotes binding of the mini-chromosome maintenance complex to DNA. J Biol Chem **279:**43008-12.

27. **DeLong, E. F., and N. R. Pace.** 2001. Environmental diversity of bacteria and archaea. Syst Biol **50:**470-8.

28. **Desogus, G., S. Onesti, P. Brick, M. Rossi, and F. M. Pisani.** 1999. Identification and characterization of a DNA primase from the hyperthermophilic archaeon *Methanococcus jannaschii*. Nucleic Acids Res **27:**4444-50.

29. **Dionne, I., R. K. Nookala, S. P. Jackson, A. J. Doherty, and S. D. Bell.** 2003. A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. Mol Cell **11:**275-82.

30. **Edgell, D. R., and W. F. Doolittle.** 1997. Archaea and the origin(s) of DNA replication proteins. Cell **89:**995-8.

31. **Elkins, J. G., M. Podar, D. E. Graham, K. S. Makarova, Y. Wolf, L. Randau, B. P. Hedlund, C. Brochier-Armanet, V. Kunin, I. Anderson, A. Lapidus, E. Goltsman, K. Barry, E. V. Koonin, P. Hugenholtz, N. Kyrpides, G. Wanner, P. Richardson, M. Keller, and K. O. Stetter.** 2008. A korarchaeal genome reveals insights into the evolution of the Archaea. Proc Natl Acad Sci U S A **105:**8102-7.

32. **Forterre, P., J. Filée, and H. Myllykallio.** 2004. Origin and evolution of DNA and DNA replication machineries, p. 145-168. *In* L. Ribas (ed.), The Genetic Code and the Origin of Life. Landes Bioscience, Georgetown, TX.

33. **Fukui, T., H. Atomi, T. Kanai, R. Matsumi, S. Fujiwara, and T. Imanaka.** 2005. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. Genome Res **15:**352-63.

34. **Galagan, J. E., C. Nusbaum, A. Roy, M. G. Endrizzi, P. Macdonald, W. FitzHugh, S. Calvo, R. Engels, S. Smirnov, D. Atnoor, A. Brown, N. Allen, J. Naylor, N. Stange-Thomann, K. DeArellano, R. Johnson, L. Linton, P. McEwan, K. McKernan, J. Talamas, A. Tirrell, W. Ye, A. Zimmer, R. D. Barber, I. Cann, D. E. Graham, D. A. Grahame, A. M. Guss, R. Hedderich, C. Ingram-Smith, H. C. Kuettner, J. A. Krzycki, J. A. Leigh, W. Li, J. Liu, B. Mukhopadhyay, J. N. Reeve, K. Smith, T. A. Springer, L. A. Umayam, O. White, R. H. White, E. Conway de Macario, J. G. Ferry, K. F. Jarrell, H. Jing, A. J. Macario, I. Paulsen, M. Pritchett, K. R. Sowers, R. V. Swanson, S. H. Zinder, E. Lander, W. W. Metcalf, and B. Birren.** 2002. The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. Genome Res **12:**532-42.

35. **Grabowski, B., and Z. Kelman.** 2003. Archeal DNA replication: eukaryal proteins in a bacterial context. Annu Rev Microbiol **57:**487-516.

36. **Gribaldo, S., and C. Brochier-Armanet.** 2006. The origin and evolution of Archaea: a state of the art. Philos Trans R Soc Lond B Biol Sci **361:**1007-22.

37. **Hanford, M. J., and T. L. Peeples.** 2002. Archaeal tetraether lipids: unique structures and applications. Appl Biochem Biotechnol **97:**45-62.

38. **Haseltine, C. A., and S. C. Kowalczykowski.** 2002. A distinctive single-strand DNA-binding protein from the Archaeon *Sulfolobus solfataricus*. Mol Microbiol **43:**1505-15.

39. **Hedderich, R., and W. B. Whitman.** 2006. Physiology and biochemistry of the methane-producing Archaea, p. 1050-1079. *In* M. Dworkin, S. Falkow, E. Rosenberg, K. Schleifer, and E. Stackebrandt (ed.), The Prokaryotes, Third ed, vol. 2. Springer-Verlag, New York.

40. **Hendrickson, E. L., R. Kaul, Y. Zhou, D. Bovee, P. Chapman, J. Chung, E. Conway de Macario, J. A. Dodsworth, W. Gillett, D. E. Graham, M. Hackett, A. K. Haydock, A. Kang, M. L. Land, R. Levy, T. J. Lie, T. A. Major, B. C. Moore, I. Porat, A. Palmeiri, G. Rouse, C. Saenphimmachak, D. Soll, S. Van Dien, T. Wang, W. B. Whitman, Q. Xia, Y. Zhang, F. W. Larimer, M. V. Olson, and J. A. Leigh.** 2004. Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*. J Bacteriol **186:**6956-69.

41. **Henneke, G., D. Flament, U. Hubscher, J. Querellou, and J. P. Raffin.** 2005. The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. J Mol Biol **350:**53-64.

42. **Hu, J., L. Guo, K. Wu, B. Liu, S. Lang, and L. Huang.** 2012. Template-dependent polymerization across discontinuous templates by the heterodimeric primase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. Nucleic Acids Res **40:**3470-83.

43. **Huber, H., M. J. Hohn, R. Rachel, T. Fuchs, V. C. Wimmer, and K. O. Stetter.** 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature **417:**63-7.

44. **Hubscher, U., G. Maga, and S. Spadari.** 2002. Eukaryotic DNA polymerases. Annu Rev Biochem **71:**133-63.

45. **Iftode, C., Y. Daniely, and J. A. Borowiec.** 1999. Replication protein A (RPA): the eukaryotic SSB. Crit Rev Biochem Mol Biol **34:**141-80.

46. **Ilves, I., T. Petojevic, J. J. Pesavento, and M. R. Botchan.** 2010. Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. Mol Cell **37:**247-58.

47. **Imamura, M., T. Uemori, I. Kato, and Y. Ishino.** 1995. A non-alpha-like DNA polymerase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Biol Pharm Bull **18:**1647-52.

48. **Ishino, Y., and S. Ishino.** 2012. Rapid progress of DNA replication studies in Archaea, the third domain of life. Sci China Life Sci **55:**386-403.

49. **Ito, N., I. Matsui, and E. Matsui.** 2007. Molecular basis for the subunit assembly of the primase from an archaeon *Pyrococcus horikoshii*. FEBS J **274:**1340-51.

50. **Jokela, M., A. Eskelinen, H. Pospiech, J. Rouvinen, and J. E. Syvaoja.** 2004. Characterization of the 3' exonuclease subunit DP1 of *Methanococcus jannaschii* replicative DNA polymerase D. Nucleic Acids Res **32:**2430-40.

51. **Kasiviswanathan, R., J. H. Shin, and Z. Kelman.** 2005. Interactions between the archaeal Cdc6 and MCM proteins modulate their biochemical properties. Nucleic Acids Res **33:**4940-50.

52. **Kelly, T. J., P. Simancek, and G. S. Brush.** 1998. Identification and characterization of a single-stranded DNA-binding protein from the archaeon *Methanococcus jannaschii*. Proc Natl Acad Sci U S A **95:**14634-9.

53. **Kelman, Z., J. K. Lee, and J. Hurwitz.** 1999. The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* DeltaH contains DNA helicase activity. Proc Natl Acad Sci U S A **96:**14783-8.

54. **Kelman, Z., S. Pietrokovski, and J. Hurwitz.** 1999. Isolation and characterization of a split B-type DNA polymerase from the archaeon *Methanobacterium thermoautotrophicum* deltaH. J Biol Chem **274:**28751-61.

55. **Kerr, I. D., R. I. Wadsworth, L. Cubeddu, W. Blankenfeldt, J. H. Naismith, and M. F. White.** 2003. Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of *Sulfolobus* SSB protein. EMBO J **22:**2561-70.

56. **Kluyver, A. J., and C. G. Schnellen.** 1947. On the fermentation of carbon monoxide by pure cultures of methane bacteria. Arch Biochem **14:**57-70.

57. **Kochiwa, H., M. Tomita, and A. Kanai.** 2007. Evolution of ribonuclease H genes in prokaryotes to avoid inheritance of redundant genes. BMC Evol Biol **7:**128.

58. **Komori, K., and Y. Ishino.** 2001. Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. J Biol Chem **276:**25654-60.

59. **Kong, X. P., R. Onrust, M. O'Donnell, and J. Kuriyan.** 1992. Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp. Cell **69:**425-37.

60. **Konneke, M., A. E. Bernhard, J. R. de la Torre, C. B. Walker, J. B. Waterbury, and D. A. Stahl.** 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature **437:**543-6.

61. **Koonin, E. V.** 2010. The origin and early evolution of eukaryotes in the light of phylogenomics. Genome Biol **11:**209.

62. **Krupovic, M., S. Gribaldo, D. H. Bamford, and P. Forterre.** 2010. The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. Mol Biol Evol **27:**2716-32.

63. **Labib, K., and J. F. Diffley.** 2001. Is the MCM2-7 complex the eukaryotic DNA replication fork helicase? Curr Opin Genet Dev **11:**64-70.

64. **Ladner, J. E., M. Pan, J. Hurwitz, and Z. Kelman.** 2011. Crystal structures of two active proliferating cell nuclear antigens (PCNAs) encoded by *Thermococcus kodakaraensis*. Proc Natl Acad Sci U S A **108:**2711-6.

65. **Lao-Sirieix, S. H., and S. D. Bell.** 2004. The heterodimeric primase of the hyperthermophilic archaeon *Sulfolobus solfataricus* possesses DNA and RNA primase, polymerase and 3'-terminal nucleotidyl transferase activities. J Mol Biol **344:**1251-63.

66. **Lao-Sirieix, S. H., L. Pellegrini, and S. D. Bell.** 2005. The promiscuous primase. Trends Genet **21:**568-72.

67. **Le Breton, M., G. Henneke, C. Norais, D. Flament, H. Myllykallio, J. Querellou, and J. P. Raffin.** 2007. The heterodimeric primase from the euryarchaeon *Pyrococcus abyssi*: a multifunctional enzyme for initiation and repair? J Mol Biol **374:**1172-85.

68. **Lecompte, O., R. Ripp, J. C. Thierry, D. Moras, and O. Poch.** 2002. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. Nucleic Acids Res **30:**5382-90.

69. **Li, Z., M. Pan, T. J. Santangelo, W. Chemnitz, W. Yuan, J. L. Edwards, J. Hurwitz, J. N. Reeve, and Z. Kelman.** 2011. A novel DNA nuclease is stimulated by association with the GINS complex. Nucleic Acids Res **39:**6114-23.

70. **Li, Z., T. J. Santangelo, L. Cubonova, J. N. Reeve, and Z. Kelman.** 2010. Affinity purification of an archaeal DNA replication protein network. MBio **1**.

71. **Lin, Y., L. J. Lin, P. Sriratana, K. Coleman, T. Ha, M. Spies, and I. K. Cann.** 2008. Engineering of functional replication protein a homologs based on insights into the evolution of oligonucleotide/oligosaccharide-binding folds. J Bacteriol **190:**5766-80.

72. **Lindahl, T.** 1993. Instability and decay of the primary structure of DNA. Nature **362:**709-15.

73. **Liu, J., C. L. Smith, D. DeRyckere, K. DeAngelis, G. S. Martin, and J. M. Berger.** 2000. Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. Mol Cell **6:**637-48.

74. **Liu, L., K. Komori, S. Ishino, A. A. Bocquier, I. K. Cann, D. Kohda, and Y. Ishino.** 2001. The archaeal DNA primase: biochemical characterization of the p41-p46 complex from *Pyrococcus furiosus*. J Biol Chem **276:**45484-90.

75. **Liu, Y.** 2010. Taxonomy of methanogens, p. 550-558. *In* K. N. Timmis (ed.), Handbook of hydrocarbon and lipid microbiology, vol. 1. Springer-Verlag Berlin Heidelberg, New York.

76. **Londei, P.** 2005. Evolution of translational initiation: new insights from the archaea. FEMS Microbiol Rev **29:**185-200.

77. **Lundgren, M., A. Andersson, L. Chen, P. Nilsson, and R. Bernander.** 2004. Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. Proc Natl Acad Sci U S A **101:**7046-51.

78. **Luo, H., Z. Sun, W. Arndt, J. Shi, R. Friedman, and J. Tang.** 2009. Gene order phylogeny and the evolution of methanogens. PLoS One **4:**e6069.

79. **Maisnier-Patin, S., L. Malandrin, N. K. Birkeland, and R. Bernander.** 2002. Chromosome replication patterns in the hyperthermophilic euryarchaea *Archaeoglobus fulgidus* and *Methanocaldococcus (Methanococcus) jannaschii*. Mol Microbiol **45:**1443-50.

80. **Majernik, A. I., and J. P. Chong.** 2008. A conserved mechanism for replication origin recognition and binding in archaea. Biochem J **409:**511-8.

81. **Makarova, K. S., Y. I. Wolf, S. L. Mekhedov, B. G. Mirkin, and E. V. Koonin.** 2005. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. Nucleic Acids Res **33:**4626-38.

82. **Marians, K. J.** 1992. Prokaryotic DNA replication. Annu Rev Biochem **61:**673-719.

83. **Marinsek, N., E. R. Barry, K. S. Makarova, I. Dionne, E. V. Koonin, and S. D. Bell.** 2006. GINS, a central nexus in the archaeal DNA replication fork. EMBO Rep **7:**539-45.

84. **Marsh, V. L., and S. D. Bell.** 2007. DNA replication and the cell cycle, p. 159-169. *In* R. A. Garrett and H.-P. Klenk (ed.), Archaea: evolution, physiology and molecular biology. Blackwell Publishing Ltd.

85. **Matsui, E., M. Nishio, H. Yokoyama, K. Harata, S. Darnis, and I. Matsui.** 2003. Distinct domain functions regulating de novo DNA synthesis of thermostable DNA primase from hyperthermophile *Pyrococcus horikoshii*. Biochemistry **42:**14968-76.

86. **Matsunaga, F., P. Forterre, Y. Ishino, and H. Myllykallio.** 2001. In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. Proc Natl Acad Sci U S A **98:**11152-7.

87. **Matsunaga, F., C. Norais, P. Forterre, and H. Myllykallio.** 2003. Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. EMBO Rep **4:**154-8.

88. **Mayanagi, K., S. Kiyonari, H. Nishida, M. Saito, D. Kohda, Y. Ishino, T. Shirai, and K. Morikawa.** 2011. Architecture of the DNA polymerase B-proliferating cell nuclear antigen (PCNA)-DNA ternary complex. Proc Natl Acad Sci U S A **108:**1845-9.

89. **Murzin, A. G.** 1993. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J **12:**861-7.

90. **Myllykallio, H., P. Lopez, P. Lopez-Garcia, R. Heilig, W. Saurin, Y. Zivanovic, H. Philippe, and P. Forterre.** 2000. Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. Science **288:**2212-5.

91. **Nakatani, M., S. Ezaki, H. Atomi, and T. Imanaka.** 2000. A DNA ligase from a hyperthermophilic archaeon with unique cofactor specificity. J Bacteriol **182:**6424-33.

92. **Nishida, H., K. Mayanagi, S. Kiyonari, Y. Sato, T. Oyama, Y. Ishino, and K. Morikawa.** 2009. Structural determinant for switching between the polymerase and exonuclease modes in the PCNA-replicative DNA polymerase complex. Proc Natl Acad Sci U S A **106:**20693-8.

93. **Norais, C., M. Hawkins, A. L. Hartman, J. A. Eisen, H. Myllykallio, and T. Allers.** 2007. Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. PLoS Genet **3:**e77.

94. **Nowak, M. A., M. C. Boerlijst, J. Cooke, and J. M. Smith.** 1997. Evolution of genetic redundancy. Nature **388:**167-71.

95. **Nunoura, T., Y. Takaki, J. Kakuta, S. Nishi, J. Sugahara, H. Kazama, G. J. Chee, M. Hattori, A. Kanai, H. Atomi, K. Takai, and H. Takami.** 2011. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. Nucleic Acids Res **39:**3204-23.

96. **Ogino, H., S. Ishino, K. Mayanagi, G. T. Haugland, N. K. Birkeland, A. Yamagishi, and Y. Ishino.** 2011. The GINS complex from the thermophilic archaeon, *Thermoplasma acidophilum* may function as a homotetramer in DNA replication. Extremophiles **15:**529-39.

97. **Ohtani, N., M. Haruki, M. Morikawa, R. J. Crouch, M. Itaya, and S. Kanaya.** 1999. Identification of the genes encoding Mn2+-dependent RNase HII and Mg2+-dependent RNase HIII from *Bacillus subtilis*: classification of RNases H into three families. Biochemistry **38:**605-18.

98. **Ohtani, N., M. Haruki, M. Morikawa, and S. Kanaya.** 1999. Molecular diversities of RNases H. J Biosci Bioeng **88:**12-9.

99. **Ohtani, N., H. Yanagawa, M. Tomita, and M. Itaya.** 2004. Cleavage of double-stranded RNA by RNase HI from a thermoacidophilic archaeon, *Sulfolobus tokodaii* 7. Nucleic Acids Res **32:**5809-19.

100. **Ohtani, N., H. Yanagawa, M. Tomita, and M. Itaya.** 2004. Identification of the first archaeal Type 1 RNase H gene from *Halobacterium* sp. NRC-1: archaeal RNase HI can cleave an RNA-DNA junction. Biochem J **381:**795-802.

101. **Oost, J. v. d., and B. Siebers.** 2007. The glycolytic pathways of Archaea: evolution by tinkering, p. 247-260. *In* R. A. Garret and H.-P. Klenk (ed.), Archaea: evolution, physiology and molecular biology. Blackwell Publishing Ltd.

102. **Oyama, T., S. Ishino, S. Fujino, H. Ogino, T. Shirai, K. Mayanagi, M. Saito, N. Nagasawa, Y. Ishino, and K. Morikawa.** 2011. Architectures of archaeal GINS complexes, essential DNA replication initiation factors. BMC Biol **9:**28.

103. **Pan, M., T. J. Santangelo, Z. Li, J. N. Reeve, and Z. Kelman.** 2011. Thermococcus kodakarensis encodes three MCM homologs but only one is essential. Nucleic Acids Res **39:**9671-80.

104. **Pelve, E. A., A. C. Lindas, A. Knoppel, A. Mira, and R. Bernander.** 2012. Four chromosome replication origins in the archaeon *Pyrobaculum calidifontis*. Mol Microbiol **85:**986-95.

105. **Rajman, L. A., and S. T. Lovett.** 2000. A thermostable single-strand DNase from *Methanococcus jannaschii* related to the RecJ recombination and repair exonuclease from Escherichia coli. J Bacteriol **182:**607-12.

106. **Robbins, J. B., M. C. Murphy, B. A. White, R. I. Mackie, T. Ha, and I. K. Cann.** 2004. Functional analysis of multiple single-stranded DNA-binding proteins from *Methanosarcina acetivorans* and their effects on DNA synthesis by DNA polymerase BI. J Biol Chem **279:**6315-26.

107. **Robinson, N. P., and S. D. Bell.** 2007. Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. Proc Natl Acad Sci U S A **104:**5806-11.

108. **Robinson, N. P., I. Dionne, M. Lundgren, V. L. Marsh, R. Bernander, and S. D. Bell.** 2004. Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. Cell **116:**25-38.

109. **Rogozin, I. B., K. S. Makarova, Y. I. Pavlov, and E. V. Koonin.** 2008. A highly conserved family of inactivated archaeal B family DNA polymerases. Biol Direct **3:**32.

110. **Rouillon, C., G. Henneke, D. Flament, J. Querellou, and J. P. Raffin.** 2007. DNA polymerase switching on homotrimeric PCNA at the replication fork of the euryarchaea *Pyrococcus abyssi*. J Mol Biol **369:**343-55.

111. **Sanchez-Pulido, L., and C. P. Ponting.** 2011. Cdc45: the missing RecJ ortholog in eukaryotes? Bioinformatics **27:**1885-8.

112. **Sandman, K., and J. N. Reeve.** 2006. Archaeal histones and the origin of the histone fold. Curr Opin Microbiol **9:**520-5.

113. **Sarmiento, F., Mrázek, J and Whitman, WB.** 2013. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. Proc Natl Acad Sci U S A. doi: 10.1073/pnas.1220225110.

114. **Shechter, D. F., C. Y. Ying, and J. Gautier.** 2000. The intrinsic DNA helicase activity of *Methanobacterium thermoautotrophicum* delta H minichromosome maintenance protein. J Biol Chem **275:**15049-59.

115. **Shen, Y., K. Musti, M. Hiramoto, H. Kikuchi, Y. Kawarabayashi, and I. Matsui.** 2001. Invariant Asp-1122 and Asp-1124 are essential residues for polymerization

catalysis of family D DNA polymerase from *Pyrococcus horikoshii*. J Biol Chem **276:**27376-83.

116. **Shin, J. H., B. Grabowski, R. Kasiviswanathan, S. D. Bell, and Z. Kelman.** 2003. Regulation of minichromosome maintenance helicase activity by Cdc6. J Biol Chem **278:**38059-67.

117. **Shiomi, Y., J. Usukura, Y. Masamura, K. Takeyasu, Y. Nakayama, C. Obuse, H. Yoshikawa, and T. Tsurimoto.** 2000. ATP-dependent structural change of the eukaryotic clamp-loader protein, replication factor C. Proc Natl Acad Sci U S A **97:**14127-32.

118. **Singleton, M. R., R. Morales, I. Grainge, N. Cook, M. N. Isupov, and D. B. Wigley.** 2004. Conformational changes induced by nucleotide binding in Cdc6/ORC from *Aeropyrum pernix*. J Mol Biol **343:**547-57.

119. **Sun, Y., M. S. Seo, J. H. Kim, Y. J. Kim, G. A. Kim, J. I. Lee, J. H. Lee, and S. T. Kwon.** 2008. Novel DNA ligase with broad nucleotide cofactor specificity from the hyperthermophilic crenarchaeon *Sulfophobococcus zilligii*: influence of ancestral DNA ligase on cofactor utilization. Environ Microbiol **10:**3212-24.

120. **Tomkinson, A. E., S. Vijayakumar, J. M. Pascal, and T. Ellenberger.** 2006. DNA ligases: structure, reaction mechanism, and function. Chem Rev **106:**687-99.

121. **Tye, B. K.** 1999. MCM proteins in DNA replication. Annu Rev Biochem **68:**649-86.

122. **Uemori, T., Y. Ishino, H. Toh, K. Asada, and I. Kato.** 1993. Organization and nucleotide sequence of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*. Nucleic Acids Res **21:**259-65.

123. **Uemori, T., Y. Sato, I. Kato, H. Doi, and Y. Ishino.** 1997. A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization. Genes Cells **2:**499-512.

124. **Wadsworth, R. I., and M. F. White.** 2001. Identification and properties of the crenarchaeal single-stranded DNA binding protein from *Sulfolobus solfataricus*. Nucleic Acids Res **29:**914-20.

125. **Walters, A. D., and J. P. Chong.** 2010. An archaeal order with multiple minichromosome maintenance genes. Microbiology **156:**1405-14.

126. **Walters, A. D., and J. P. Chong.** 2009. *Methanococcus maripaludis*: an archaeon with multiple functional MCM proteins? Biochem Soc Trans **37:**1-6.

127. **Warbrick, E.** 1998. PCNA binding through a conserved motif. Bioessays **20:**195-9.

128. **Whittaker, R. H.** 1969. New concepts of kingdoms or organisms. Evolutionary relations are better represented by new classifications than by the traditional two kingdoms. Science **163:**150-60.

129. **Williams, K. R., J. B. Murphy, and J. W. Chase.** 1984. Characterization of the structural and functional defect in the Escherichia coli single-stranded DNA binding protein encoded by the ssb-1 mutant gene. Expression of the ssb-1 gene under lambda pL regulation. J Biol Chem **259:**11804-11.

130. **Woese, C. R.** 2007. The birth of the Archaea: a personal retrospective, p. 1-15. *In* R. A. Garrett and H.-P. Klenk (ed.), Archaea: evolution, physiology and molecular biology. Blackwell Publishing Ltd.

131. **Woese, C. R., and G. E. Fox.** 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A **74:**5088-90.

132. **Woese, C. R., O. Kandler, and M. L. Wheelis.** 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A **87:**4576-9.

133. **Wolfe, R. S.** 1993. An historical overview of methanogenesis, p. 1-32. *In* J. G. Ferry (ed.), Methanogenesis: ecology, physiology, biochemistry & genetics. Chapman & Hall, New York.

134. **Xia, Q., E. L. Hendrickson, Y. Zhang, T. Wang, F. Taub, B. C. Moore, I. Porat, W. B. Whitman, M. Hackett, and J. A. Leigh.** 2006. Quantitative proteomics of the archaeon *Methanococcus maripaludis* validated by microarray analysis and real time PCR. Mol Cell Proteomics **5:**868-81.

135. **Yao, N. Y., R. E. Georgescu, J. Finkelstein, and M. E. O'Donnell.** 2009. Single-molecule analysis reveals that the lagging strand increases replisome processivity but slows replication fork progression. Proc Natl Acad Sci U S A **106:**13236-41.

136. **Yoshimochi, T., R. Fujikane, M. Kawanami, F. Matsunaga, and Y. Ishino.** 2008. The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. J Biol Chem **283:**1601-9.

137. **Zhang, R., and C. T. Zhang.** 2005. Identification of replication origins in archaeal genomes based on the Z-curve method. Archaea **1:**335-46.

138. **Zhang, R., and C. T. Zhang.** 2003. Multiple replication origins of the archaeon *Halobacterium* species NRC-1. Biochem Biophys Res Commun **302:**728-34.

139. **Zuo, Z., C. J. Rodgers, A. L. Mikheikin, and M. A. Trakselis.** 2010. Characterization of a functional DnaG-type primase in archaea: implications for a dual-primase system. J Mol Biol **397:**664-76.

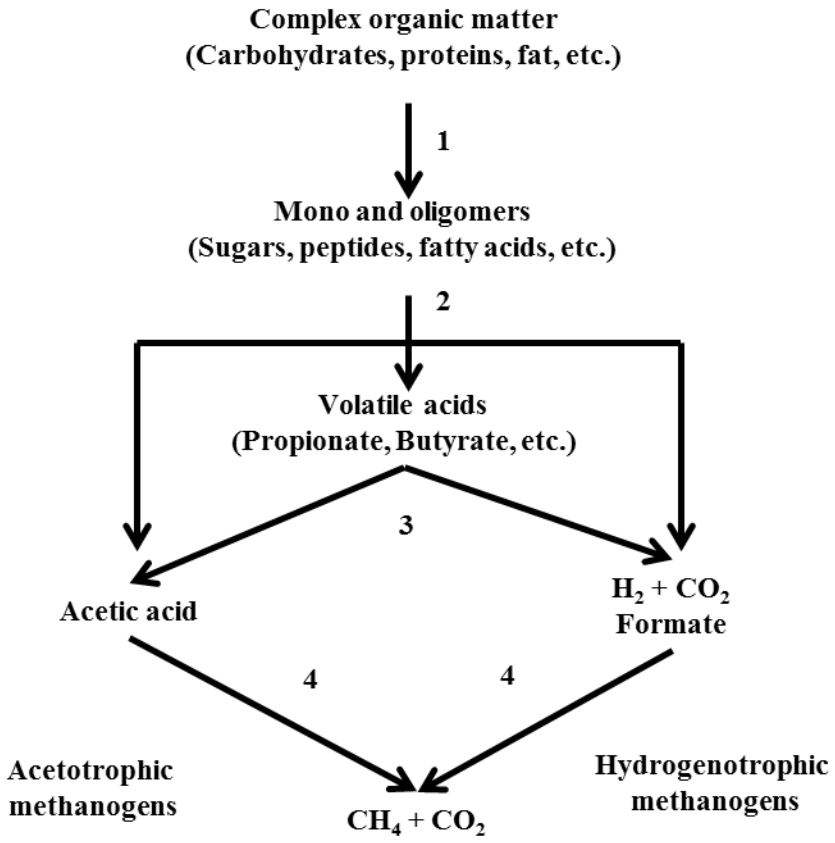

Figure 1-1. Anaerobic mineralization for the conversion of organic matter to methane. Numbers indicate the process that catalyzes the specific step. 1. hydrolysis, 2. fermentation, 3. acetogenesis, 4. methanogenesis.

Table 1-1. DNA replication proteins and features in Bacteria, Eukarya and the two major phyla of Archaea. Modified from (48).

DNA Replication Stage	Bacteria	Eukarya	Archaea	
			Crenarchaea	**Euryarchaea**
Origin of replication	Single	Multiple	Multiple	Single[a]
Origin recognition	DnaA	ORC complex (ORC 1-6)	Cdc6/Orc1	Cdc6/Orc1[b]
DNA unwinding (Helicase)	DnaB	MCM complex (MCM 2-7)	MCM complex	MCM complex
DNA unwinding (Accesory proteins)	DnaC	Cdc6	GINS23/GINS15	GINS15[c]
		Cdt1		
		GINS complex (Sld5, Psf1-3)	RecJ homolog?	RecJ homolog?
		Cdc45		
Primer synthesis	DnaG	Pol α/primase complex	DNA primase (PriSL)/DnaG[d]	DNA primase (PriSL)
DNA synthesis (Polymerase)	Pol III (Family C DNA polymerase)	Polα, Polδ, and Polε (Family B DNA polymerase)	Family B DNA polymerase	Family D DNA polymerase
DNA synthesis (Processivity factors)	γ-complex (Clamp loader)	RFC (Clamp loader)	RFC (Clamp loader)	RFC (Clamp loader)
	β-clamp (Clamp)	PCNA (Clamp)	PCNA (Clamp)	PCNA (Clamp)
Maturation (Okazaki fragment processing)	Pol I (Family A DNA polymerase)	Fen1/Dna2	Fen1	Fen1
	RnaseH	RnaseH	RnaseH	RnaseH
	DNA ligase	DNA ligase	DNA ligase	DNA ligase

[a] One exception, the order *Halobacteriales.*
[b] Not known for members of the Euryarchaeota orders *Methanococcales* and *Methanopyrales*
[c] GINS23 has been founded only in the order *Thermococcales* in Euryarchaea
[d] *S. sulfataricus* did show primase activity *in vitro*

CHAPTER 2

GENETIC SYSTEMS FOR HYDROGENOTROPHIC METHANOGENS[1]

[1]**Sarmiento, F., Leigh J.A. and W.B. Whitman.** 2011. *Methods Enzymol*. 494:43-73. Reprinted here with permission of the publisher.

Abstract:

Methanogens are obligate anaerobic Archaea that produce energy from the biosynthesis of methane. These lithotrophic microorganisms are widely distributed in oxygen-free environments and participate actively in the carbon cycle. Indeed, methanogenesis plays a major role in the last step of the anoxic degradation of organic substances, transforming acetate, CO_2 and H_2 to methane. The vast majority of the known methanogens are classified as hydrogenotrophic because they use principally H_2 as the electron donor to drive the reduction of CO_2. Unlike many other cultured Archaea, many methanogens thrive in neutral pH, low salinity, and temperate environments. This has been a great advantage in cultivating these organisms in laboratory conditions and in the development of genetic tools. Moreover, the hydrogenotroph *Methanococcus maripaludis* is currently a model organism among Archaea, not only for its utility in genetic but for biochemical and physiological studies. Over time, a broad spectrum of genetic tools and techniques has been developed for methanococci, such as site-directed mutagenesis, selectable markers, transformation methods and reporter genes. These tools have contributed greatly to the overall understanding of this group of microorganisms and the processes that govern its life style. In this chapter, we describe in detail the available genetic tools for the hydrogenotrophic methanogens.

1. Introduction

Methanogens are a diverse group of microorganisms that are distinguished by their ability to obtain most of their metabolic energy from the biosynthesis of methane or methanogenesis.

These microorganisms share specific characteristics that make them unique. First, they are Archaea, belonging to the phylum *Euryarchaeota*. Second, they are obligate anaerobes, being found only in oxygen-free environments. Finally, they are obligate methane producers. Indeed, to date no methanogens have been identified that can growth under physiological conditions without producing methane (15).

Methanogens are widely distributed and are found in different anaerobic habitats on Earth. Those habitats range from temperate environments, such as fresh water sediments, to environments with extreme salinity, temperature and pH, such as hot springs. In these anaerobic habitats, different microorganisms degrade organic matter to produce H_2, CO_2, and acetate. In the absence of sulfate, oxidized metals, and nitrite, methanogens consume these substrates, contributing in two different ways to the anaerobic food chain. First, methanogens catalyze the last step of the anoxic degradation of organic substances, producing methane, which is released to the atmosphere. Second, by maintaining an extremely low partial pressure of H_2, they keep the fermentative pathways energetically favorable (15).

Taxonomic classification of methanogens places them in the phylum *Euryarchaeota* and divides them into six well-established orders (*Methanobacteriales*, *Methanocellales*, *Methanococcales*, *Methanomicrobiales*, *Methanosarcinales*, and *Methanopyrales*) and 31 genera. This taxonomy is support by comparative 16S rRNA gene sequence analysis and some distinct phenotypic properties, such as lipid composition, substrate range, and cell envelope structure (24, 38).

In spite of this diversity in taxonomy and physiology, methanogens can only utilize a restricted number of substrates. Methanogenesis is limited to three main types of substrates: CO_2, methyl-group containing compounds, and acetate. Most organic substances cannot be utilized by methanogens directly and must be converted by other microorganisms to the substrates for methanogenesis.

Most methanogens that belong to the orders *Methanobacteriales, Methanococcales, Methanomicrobiales, Methanocellales,* and *Methanopyrales* are hydrogenotroph and can reduce CO_2 to CH_4 using H_2 as the electron donor. In addition, a few representatives of the order *Methanosarcinales* are facultative hydrogenotrophs. Because these organisms are primary methylotrophs or acetotrophs, they are not discussed further in this chapter. The reduction of CO_2 is driven in four steps, some of which involve the unusual or unique coenzymes that function as C-1 moiety carriers. The process begins with the binding of CO_2 to the coenzyme methanofuran (MFR) and its subsequent reduction to the formyl level. Then, the formyl group is transferred to the coenzyme H_4MPT and dehydrated to methenyl-H_4MPT. The carbon group is subsequently reduced from methenyl-H_4MPT to methylene-H_4MPT and methyl- H_4MPT. Consequently, the methyl group is transferred to the thiol of coenzyme M (CoM). In the last step, methyl coenzyme M reductase (MCR) catalyzes the reduction of methyl-CoM to methane using coenzyme B (CoB) as an electron donor. The oxidation of CoB yields the heterodisulfide with CoM, which must then be reduced to regenerate the thiols (26).

While, H_2 is the main electron donor for methanogenesis, many hydrogenotrophic methanogens can use formate, ethanol or some secondary alcohols as electron donors. In the first case, four molecules of formate are oxidized to CO_2, and one molecule of CO_2 is reduced to methane. For the alcohols, ethanol is oxidized to acetate, and the secondary alcohols are oxidized

to ketones to drive the reduction of CO_2. However, methanogens grow poorly using alcohols as electron donors (15).

Among the hydrogenotrophic methanogens, *M. maripaludis* has become one of the most useful model organisms to develop new molecular tools. This microorganism, which belongs to the order *Methanococcales*, presents several advantages for genetic studies. 1) It is a mesophile that grows best at 37 C. 2) It has a relatively fast growth rate, with a doubling time of 2 hours. 3) It is a facultative autotroph that takes up different organic substrates from the medium. 4) Plating techniques have been optimized to reach efficiency near 100%. 5) Cells lyse in low-ionic-strength buffers or in very low concentration of detergents, facilitating the isolation of DNA and other cellular components (44).

A broad collection of molecular biology tools, including transformation, *in frame* deletions, transposon mutagenesis and auxotrophic selection have been developed. These have contributed greatly on our understanding of the genetics of methanogens. In this chapter, we present and describe the principal advances that have been made in genetic systems for hydrogenotrophic methanogens, taking as the main reference point the order *Methanococcales*.

2. Genome sequences

In 1996, the genome of *Methanocaldococcus jannaschii* was completely sequenced (11)**.** This milestone represented the first sequenced genome of a representative of the Archaea as well as a methanogen. Since then 25 genomes of hydrogenotrophic methanogens have been sequenced, belonging to five of the six corresponding taxonomic orders. Based on genome characteristics, the currently available genomes can be clustered in two groups. The first group consists of the genomes of the orders *Methanococcales*, *Methanobacteriales* and *Methanopyrales*, which generally have small genomes that range between 1.3 Mb and 1.9 Mb and a low G+C content

between 27 and 34 mol%. Exceptions include only *Methanobrevibacter ruminantum M1*, which has a genome size of 2.9 Mb, and *Methanothermobacter thermauthotrophicus* Δ*H* and *Methanopyrus kandleri AV19*, which have mol% GC of 49.54% and 61.6%, respectively. The second group consists of the orders *Methanomicrobiales* and *Methanocellales*. These hydrogenotrophic methanogens possess larger genomes that range between 1.8 Mb and 3.5 Mb and higher genome GC contents between 45 and 62 mol%.

A complete list of the hydrogenotrophic methanogens that have fully sequenced genomes with their principal features are presented in Table 2-1.

3. Growth and storage of methanogen cultures

3.1. Liquid media preparation

Methanogens are obligate anaerobic microorganisms and require two main conditions to be successfully grown: low oxygen partial pressure and a low redox potential of -0.33 V. A three-step system to prepare prereduced media that meets these conditions was developed 60 years ago (17). First, media was boiled to expel dissolved oxygen. Second, to impede oxygenation of the medium, it was kept under an anaerobic gas phase. Finally, a reducing agent and a redox-indicator was added to keep a low redox potential and to indicate the oxidative state of the medium, respectively. This Hungate culturing technique and its later modifications, such as the introduction of the modified Wolin-Miller tube and the development of commercially available anaerobic glove boxes, greatly accelerate the research on anaerobic microorganisms including methanogens (3).

Base on this technique, *M. maripaludis* is routinely cultured in a basal medium (McN) that contains the minimal requirements for this methanogen to growth (Tables 2-2 and 2-3). However, a complex medium can be prepared by adding some nutrients, such as Casamino acids,

vitamins, yeast extract and acetate, which enhance growth. Normally, the preferred source of carbon and energy is H_2/CO_2, which is added as a mixed gas at a pressure of 275 kPa. However, recently *M. maripaludis* has been cultivated in a similar medium but in the presence of formate instead of H_2/CO_2 and with glycylglycine buffer. In this condition, *M. maripaludis* has a lower growth yield but with a similar growth rate (Boguslaw Lupa, personal communication). From a practical and safety perspective, formate presents an improvement to the cultivation method of *M. maripaludis* because it avoids working with high pressure of a flammable gas.

Preparation of 100 ml of McN, McCV or McF liquid media: In a 500 ml round-bottom flask, mix the components indicated in table 2-2, stopper loosely, and boil the solution for 5 to 10 seconds under a stream of N_2/CO_2 (80:20 v/v). Allow the solution to cool under the stream of N_2/CO_2 and add 0.05 g of L-cysteine hydrochloride hydrate or sodium 2-mercaptoethanosulfonate to reduce the medium. Tightly stopper and transfer the flask into the anaerobic chamber. Dispense the solution into Balch tubes or serum bottles. Seal the tubes with 20 mm blue septum stoppers (Bellco Glass, Inc.) and secure each stopper with an aluminum seal. Remove the tubes from the chamber and exchange gases through a three-cycle procedure of H_2/CO_2 and vacuum. Pressurize each tube to 137 kPa of H_2/CO_2 and sterilize the tubes by autoclaving at 121 C for 20 min.

For the preparation of McF, a few modifications should be introduced. First, the medium composition is similar to the McN and McCV but NaCl is replaced by sodium formate and glycylglycine is added (Table 2-2). In addition, instead of H_2/CO_2 the gas in the tubes is exchanged with N_2/CO_2 at 35-70 kPa. Since cells growing in the presence of formate do not require high concentrations of H_2/CO_2, the tubes can also be sealed with stoppers that have less resistance to pressure, such as the gray butyl rubber septum stoppers (Bellco Glass, Inc.).

Before inoculation of all media, add anaerobically 0.1 ml of 2.5 % Na_2S*9H_2O to each 5 ml of medium using anaerobic procedures. In the case of McN and McCV, after inoculation, pressurize each tube to 275 kPa. Incubate at 37 C overnight.

Other representatives of the order *Methanococcales*, such as *Methanococcus voltae* and *Methanococcus vannielii* are cultured following the same procedure using media prepared with the modifications that are indicated in Table 2-2.

Side notes:

- There is always a risk of an explosion when autoclaving media in seal glass tubes and bottles. For this reason, avoid chipped or heavily scratched glassware, autoclave tubes and bottles in protective wire baskets, and wear safety glasses.
- Metabolism of formate yields CO_2 and methane that is accumulated inside the tube. Approximately, 1 mmol of formate generate 25 ml of gas. Therefore, in a 30-ml tube with 5-ml of medium, the gas presence could increase by 100 kPa or 15 psi. Depending upon the pH, some of the CO_2 will be diddolved in the medium. Nevertheless, the glassware used must be able to withstand this increase in pressure without exploding.

3.2. Solid media preparation

Different methods have been described to grow methanogens on solid medium. In this chapter, the two most widely used methods will be described: the Petri dish and the serum bottle (45).

Preparation of 100 ml of McN or McCV solid media in petri dishes: In a 500 ml round-bottom flask, mix the components indicated in table 2-2 but reduce the $NaHCO_3$ to 0.2 g and add 1% of agar. Boil the solution under a stream of N_2/CO_2 (80:20 v/v). Heat slowly and mix gently to avoid burning the agar and foaming. Under a stream of N_2/CO_2, allow the solution to cool and add 0.05 g of L-cysteine hydrochloride hydrate or sodium 2-mercaptoethanosulfonate to

reduce the medium. Secure the stopper with a suitable clamp, place the bottles in metal safety cages and autoclave them for 20 min at 121 C. Allow the medium to cool, and transfer the flask into the anaerobic chamber. Dispense 20 ml of medium in 100 x 15 mm plates, and allow the plates to dry in the chamber atmosphere (20% CO_2, 3-5% H_2, balance N_2) for two to three days. One day before use, incubate the plates in a canister or pressure tank under an atmosphere of H_2S (see below).

Inside the chamber, transfer 0.5 ml of a cell suspension to each plate, spread it by gently moving the plate in circles or using a sterile bent glass rod. To prevent killing the cells at high dilutions, the glass rods should be immersed in reduced medium before use. Tape the plates together, invert and place them into the pressure tank. Pressurize the tank using H_2/CO_2 (80:20 v/v) to 100 kPa. The H_2/CO_2 gas mixture neutralizes the alkaline sulfide solution, releasing volatile H_2S. Incubate the canister at 37 C for 5-7 days, maintaining the pressure at 100 kPa by the addition of H_2/CO_2. Finally, release the pressure of the canister in a fume hood and transfer inside the chamber to pick isolated colonies.

Side notes:

- Plastic petri plates are made anaerobic by storage in the anaerobic chamber for at least one day before use. To remove air from the sleeves of plates, two corners are cut prior to transferring the plates through the anaerobic air lock.
- To generate an atmosphere of H_2S inside the canister, insert a paper towel inside a Balch tube containing 10 ml of 20 % solution of Na_2S*9H_2O and allow the paper towel absorb the solution. Then place the tube inside the canister in contact with the plates one day before inoculation. Allow the tube to remain in the canister for the entire incubation period.

- Before opening the canister inside the anaerobic chamber, the sulfide is removed to avoid poisoning the catalyst. First, release the pressure in the fume hood. Once inside the chamber, flush the canister three cycles with 100 kPa N_2 and partial vacuum of - 60 kPa.
- Water vapor in the anaerobic chamber inactives the catalyst. To remove the water vapor produced while pouring the hot agar medium a beaker with calcium chloride is keep inside the chamber. It absorbs the water vapor. This desiccant should be replaced regularly and specifically before plating.

Preparation of 100 ml of McN or McCV solid medium by the serum bottle plate method: Prepare McN or McCV medium as described above, but in only 50 ml of distilled water. In a separate flask, prepare a 2% agar solution with 50 ml of distilled water and heat until the agar is completely melted. Combine both solutions, loosely stopper in a round-bottom flask and boil the solution under a stream of N_2/CO_2 (80:20 v/v). Under the stream of N_2/CO_2, allow the solution to cool and add 0.05 g of L-cysteine hydrochloride hydrate or sodium 2-mercaptoethanosulfonate. However, before autoclaving, transfer the hot medium to the anaerobic chamber and dispense 20 ml of the solution to 70-ml serum bottles. Seal the serum bottles with 20 mm blue stoppers (Bellco Glass, Inc.) and secure each stopper with an aluminum seal. Remove the bottles from the chamber and exchange gases through a three-cycle procedure of H_2/CO_2 and vacuum. Transfer the bottles to a wire safety basket and sterilize by autoclaving at 121 C. Agar bottles can be stored inside the anaerobic chamber for later use.

Prior to plating, melt the agar by placing the bottles in boiling water or autoclaving in a 5 minutes cycle at 121 C. After they cool to 48-55 C, add 0.4 ml of 2.5 % solution of a Na_2S*9H_2O, mix the components, and allow the bottle to cool on its side to solidify the medium.

Anaerobically transfer 0.5 ml of culture suspension to the bottles, and incubate them on their sides with the agar side up at 37 C for 3-5 days.

Side notes:

- The addition of 137 kPa of H_2/CO_2 to the medium changes the pH of the solution and prevents the precipitation of some salts, such as phosphate salts.
- The maximum capacity of a 500 ml round-bottom flask is 300 ml of liquid solution. Over this limit can result in the explosion of the bottles during autoclaving.
- For any anaerobic work, all plastic and glass material (plates, tips, syringes, bottles and tubes) should be brought into the anaerobic chamber 24 hours before use to allow oxygen to diffuse from the plastic.

3.3. Glycerol stock culture preparation

Facile storage of stock cultures is essential for maintaining collections of mutants and other strains useful for genetic studies. The original protocol for long-term storage of methanogens in glycerol stocks was designed for *M. maripaludis* (47) and has been successfully implemented for all the *Methanococcus* species. However, this method has not proven suitable for all other methanogens. Here is described the most updated version of this protocol.

Preparation of 100 ml of 60% glycerol stock solution in McCV media: One day before the preparation of the glycerol stock solutions, take 3-ml serum bottles and 13 mm stoppers inside the anaerobic chamber.

In a 160-ml serum bottle, prepare 60 ml of anaerobic 100% glycerol by sparging it with N_2 for 3-4 hours or incubating it in the anaerobic chamber 2-3 days. In a 500 ml round-bottom flask prepare 100 ml of McCV medium as indicated above. Add 40 ml of McCV medium to the anaerobic glycerol into the anaerobic chamber and mix them thoroughly in the serum bottle to

obtain a final concentration of 60% (vol/vol) glycerol. Aliquot 1 ml of the solution to the 3-ml serum bottles, and seal the bottles using 13 mm stoppers and aluminum seals. Sterilize by autoclaving at 121 C for 20 min. After cooling transfer the bottles to the anaerobic chamber until use.

Stock cultures, are prepared inside the anaerobic chamber by transferring 1 ml of culture to each 3-ml serum bottle. After thorough mixing, the bottles are stored in the -80 C freezer. To inoculate a cell culture from a frozen stock, transfer the stock to a cold bath at – 20 C. Scratch the surface of the frozen sample with a syringe needle and then flush the syringe with McCV medium in an anaerobic culture tube.

In the cases of other hydrogenotrophic methanogens, such as representatives of the order *Methanobacteriales*, storage of frozen suspensions requires anoxic and reducing conditions. Hippe and Miller describe two different protocols for long-term storage of cultures that has been successfully implemented (16, 30).

4. Genetic tools

4.1. Genetic markers

Because of the absence of peptidoglycan in their cell walls and their unique ribosome structure, methanogens are naturally resistant to many of the common antibiotics. Thus, the quest for functional genetic markers for methanogens was a difficult task. Furthermore, some antibiotics that inhibit methanogens base their modus operandi on the toxic effect of their side groups or act as cell wall detergents, such as chloramphenicol and tetracycline, respectively (4); Bock and Kandler, 1985). Thus, common resistance genes are not effective. To solve this problem, Possot tested the resistance of *M. voltae* to 12 different antibiotics and found that puromycin was an ideal candidate for use as a genetic marker (36). Puromycin transacetylase provides resistance to

this antibiotic in many microorganisms, and the structural gene from *Streptomyces alboniger* was cloned under the control of the *M. voltae* methyl coenzyme M reductase promoter, generating the *pac* cassette that confers puromycin resistance to methanogens (14). Puromycin is still widely used in different representatives of methanogens because it is efficient, stable and reliable, usually effective at a final concentration of 2.5 μg/ml in both liquid medium and plates.

The other antibiotic that is widely used in methanogens is neomycin (8, 46). Neomycin resistance was first reported in *M. maripaludis* when the aminoglycoside phosphotransferase genes APH3'I and APH3'II were cloned under the control of the *M. voltae* methyl reductase promoter and transformed into *M. maripaludis* (2). This antibiotic is usually used at a final concentration of 500 μg/ml in plates and 1mg/ml for liquid medium.

Puromycin and neomycin preparation: Both solutions are prepared following the same protocol at a 200 x concentration. Dissolve puromycin or neomycin to a concentration of 0.5 or 100 mg/ml, respectively, with distilled water. Transfer the solution to a serum bottle and sparge with N_2 for 1 hour. Inside the anaerobic chamber, the solution is filter-sterilized by passage through a 0.2 μm filter. Aliquots of 10-ml are transferred to sterile Balch tubes, pressurized to 137 kPa using H_2/CO_2, and stored at -20 C.

4.2 Shuttle vectors

Diverse plasmids of hydrogenotrophic methanogens have been fully described. These plasmids include pME2001 and pME2200 from *Methanobacterium thermoautotrophicum* (9, 40), pFVI and pFZI from *Methanobacterium formicicum* (32), and two extrachromosomal elements from *Methanococcus jannaschii* (11). Tumbula *et al* described the complete sequence of a cryptic 8,285-bp low-copy number plasmid from *M. maripaludis* (pURB500) that provided good insights about the minimal replication regions for methanogen plasmids (42). The plasmid

pURB500 contains two large non-coding regions named ORFLESS1 and ORFLESS2, which possess an impressive number of direct and imperfect inverted repeats and potential stem-loop structures.

The first replicating shuttle vector (pDLT44) for methanogens was constructed based in the ligation of the cryptic plasmid pURB500 and an *E. coli* pUC18 plasmid that contained the methanococcal puromycin resistance (*pac*) cassette, named pMEB.2 (14)**.** This plasmid replicates in different strains of *M. maripaludis* and is stable for storage (under ampicillin selection) in *E. coli*, but it does not replicate in *M. voltae* (42). An expression shuttle vector (pWLG30+lacZ) was then developed for *M. maripaludis* from pDLT44 (13). This plasmid contains *lacZ* under the control of a methanococcal promoter, which can be used for screening in both *E. coli* and methanococci.

The two shuttle expression vectors most frequently used in *M. maripaludis* are plasmids pMEV1 and pMEV2 (Figures 2-1 and 2-2). They are derivatives of pWLG30+lacZ plasmid. Both contain most of pURB500 including, the origin of replication in *M. maripaludis*, as well as an origin of replication and ampicillin resistance marker for *E. coli*. Furthermore, both plasmids possess the *lacZ* gene under the control of the strong *M. voltae* histone promoter P_{hmvA} (1). To clone and express a gene under this promoter, replace the *lacZ* gene using the upstream restriction site *Nsi*I and the downstream restriction site *Xba*I. Finally, both plasmids only differ on the resistance marker for *M. maripaludis*, pMEV1 carries the *pac* cassette for puromycin resistance, and pMEV2 carries neomycin resistance.

Side notes:

- To clone a gene under the control of the P_{hmvA} promoter, design the 5' primer, to amplify the target gene, *in frame* with the *Nsi*I site. This enzyme cuts in the restriction sequence

ATGCA'T. The ATG for transcription initiation is the A'T. If the first ATG is used for transcription, the ORF will be too close to the ribosome binding site and will not be expressed.

- For the 3' primer, include an *Xba*I site for cloning in the pMEV vectors.
- To insert a PCR product into the pMEV vectors, triple digest the pMEV vector with *Nsi*I, *Bgl*II, and *Xba*I and use the vector without gel purification. The triple digestion with *Bgl*II reduces the cloning background due to the inefficient digestion by *Nsi*I. Thus, *Bgl*II digestion prevents self-ligation of the incomplete digestion product.

4.3 Integration plasmid and gene replacement mutagenesis

The first integration plasmids for methanogens (Mip1 and Mip2) were constructed based on the *E. coli* pUC18 plasmid and pMEB.2, which carries the *pac* cassette. These plasmids replicate in *E. coli* and integrate into the genome of *M. voltae*. In order to allow for homologous recombination into the *M. voltae* genome, the plasmids possessed the *his*A gene from this methanogen (14).

The plasmid pIJ03 is used commonly in *M. maripaludis* for single or double recombination into the genome (39). This suicide plasmid lacks an origin of replication for methanococci. The *pac* cassette for selection is flanked by two different multi-cloning sites to facilitate cloning of PCR products of any DNA sequence (Figure 2-3). Single homologous recombination of this plasmid leads to the incorporation of the whole plasmid into the genome. On the other hand, double homologous recombination leads to a deletion of the target gene by replacement with the *pac* cassette. Integration of the plasmids is selected by resistance to puromycin.

Several mutants have been constructed using this technique. For example: Stathopoulos *et al* used the plasmid pIJ03-*cysS* to disrupt the gene *cys*S that encode cisteinyl-tRNA synthetase, proving that this gene is not essential for viability of *M. maripaludis* (39). Lin *et al* cloned regions immediately upstream and downstream of the genes *porE* and *porF*, respectively, and used the plasmid pIJ03+CR to replace genes *porE* and *porF* and demonstrate their importance in the anabolic pyruvate oxidoreductase of *M. maripaludis* (23). Recently, Liu *et al* developed *M. maripaludis mmp1527*::*pac* mutant, a lysine auxotroph, using the plasmid pIJA03-MMP1527 (25). This work demonstrated that methanococci use the diaminopimelate aminotransferase (*dapL*) pathway for lysine biosynthesis.

4.3 Reporter genes

Two fusion reporter genes have been described to monitor gene expression in hydrogenotrophic methanogens, *uidA* and *lacZ*. These reporters can be placed under the control of different promoters to assay their strength. However, it must be remembered that reporter genes differ from the original genes that they are replacing, which could affect their expression levels.

Beneke *et al* developed a reporter gene system for *M. voltae* based on *uidA*, which encodes for β-glucuronidase (5). This reporter gene was placed under the control of the intergenic region of the genes encoding the divergently transcribed selenium-free [NiFe]-hydrogenases. The complete expression cassette was introduced into the integration vector Mip1, which carried the *pac* cassette and the gene *hisA* for homologous recombination into the genome. After transformation to *M. voltae*, β-glucuronidase expression became dependent on the depletion of selenium in the growth medium.

β-glucuronidase activity assay:

1- Cells are harvested by centrifugation, and the pellet is resuspended in test buffer (20 mM potassium phosphate pH 7, 100 mM β-mercaptoethanol), which lyses the cells.

2- Centrifuge the cell lysate for 15 min at 10,000 xg at room temperature and use the supernatant to test enzyme activity.

3- Mix 2 ml of reaction buffer with 1 mM 4-nitrophenyl-β-1,4-glucuronide and initiate the reaction by adding up to 100 µl of cell extract. The reaction is performed at 25 C and is stopped with 100 µl of 1 M sodium carbonate.

4- The reaction product (4-nitrophenol) is detected by its absorption at 405 nm (molar extinction $\varepsilon = 18{,}500\ M^{-1}\ cm^{-1}$).

Cohen-Kupiec *et al* developed a series of promoter-*lacZ* fusions to demonstrate a repressor binding site in the *nifH* promoter that regulates gene expression of the nitrogenase reductase component of the nitrogenase complex (12). The fusions were constructed using a promoterless *lacZYA* operon, which was cloned after a 1.2-kb DNA fragment containing the *nifH* promoter region from *M. maripaludis*. The resulting cassette was cloned into an integration plasmid, which carried the *pac* cassette for puromycin resistance, and introduced into *M. maripaludis* by transformation. Similar constructs were developed to study the regulatory response of *M. maripaludis* to alanine (22). In addition, an X-Gal colony screen was devised and used to identify super-repressor mutants of the transcriptional regulator NrpR (21).

β-galactosidase activity assay: is carried out using the original procedure (29).

X-Gal colony screen:

1- After anaerobic growth of the colonies (see "Solid Media Preparation"), plates are exposed for 30 min to air or until the resazurin in the medium is oxidized and has turned pink in color.

2- Two ml of X-Gal (25 mg/ml diluted in 100% dimethylformamide) are mixed with 3 ml of growth media and sprayed over the plates to completely cover the agar surface.

3- After 3-5 seconds, remove the excess solution and let the plates dry.

4- Color development is apparent after 30 – 60 min. Viable cells could be recovered from colonies exposed to air for as long as two hours.

5. Transformation methods

Over the years, different methods for transformation in methanogens have been tested. The first successfully approaches were developed for *M. voltae*. (6) discovered a natural transformation method that yielded between 8 and 9 transformants per μg of plasmid DNA. Four years later, the transformation frequency was increased 50- to 80- fold using chromosomal DNA through an anaerobic electroporation method, but this method was still inefficient for plasmid DNA (28). An important modification in the transformation methods used protoplasts of *M. voltae* (33). Protoplasting increased the natural transformation efficiency of plasmid DNA to 705 transformants/μg of DNA and the electroporation-mediated transformation efficiency to 177 transformants/μg (34). While these methods greatly improved the natural transformation efficiency, none of them reached the high number of transformants common with bacteria. However, in 1994, a polyethylene glycol (PEG)-mediated transformation method was developed for *M. maripaludis* (43). This optimized procedure gave a frequency of transformation of $1.8x10^5$ transfomants/μg of plasmid DNA, using 0.8 μg of plasmid and $3x10^9$ cells, representing an

increment of 4 orders of magnitude to the natural transformation method. Finally in 2001, a liposome delivery transformation method that was originally developed for *Methanosarcine acetivorans* (27) was adapted for use in *M. voltae* (41).

In this chapter, the methodology for the natural- and electroporation-mediated transformation of protoplasts of *M. voltae* and the PEG- mediated transformation method of *M. maripaludis* are described. For more details about the liposome delivery transformation method, please refer to the chapter on methylotrophic methanogens.

5.1. Transformation of *M. voltae* by protoplast regeneration

In many Archaea, the S-layer forms an external barrier outside of the cytoplasmic membrane and prevents the uptake of exogenous DNA. In methanococci, the S-layer is a paracrystalline array built by repetitions of a 76-kDa protein (19). The present transformation method is based on the disruption of the S-layer and the formation of protoplasts to allow an easier entrance of DNA into the cell.

Anaerobic protoplasting buffer (APB): Prepare solutions A, B and C in 50-ml serum bottles, degassing with N_2 gas for 1 hour to remove O_2 and autoclave.

- Solution A: 0.1 M Tris-HCl (pH 7.3) containing resazurin at 1 mg/liter.
- Solution B: 2 M sucrose in 0.1 M Tris-HCl containing resazurin at 1 mg/liter.
- Solution C: 1 M NaCl in 0.1 M Tris-HCl containing resazurin at 1 mg/liter.

Under anaerobic conditions, add 4 ml of solution B and 1 ml of solution C to 15 ml of solution A in a 50-ml serum bottle. Inject 0.2 ml of H_2S gas and shake the solution until the resazurin turns colorless. In a fume hood, flush the vial headspace for 5 minutes with N_2. Leave a slight overpressure and store the solution inside the chamber at room temperature.

Protoplast formation and regeneration on agar medium: Cultivate *M. voltae* in Balch tubes with 5 ml of McN + vitamins solution and the suggested medium modification for this microorganism (Table 2-2) or in BD medium (34). Incubate at 35 C to mid exponential growth phase (A_{660} of 0.5-0.8). Pressurize the culture tubes to 137 kPa with H_2/CO_2 and centrifuge them at 4000 rpm for 20 min at room temperature. Gently invert each tube and expel the supernatant using a PrecisionGlide^{TM} Vacutainer needle (Becton, Dickinson and Company) that possesses two needles. Keeping the same Vacutainer needle inserted in the tube, invert the tube again and insert the other needle into a pre-pressurize tube with 5 ml of APB. The pressure in the APB tube forces the buffer through the needle to the side of the culture tube. Resuspend cell pellets into APB at no more than five times the initial cell concentration. Immediately, pressurize the culture tubes to 137 kPa using H_2/CO_2, centrifuge at 4000 rpm for 20 min at room temperature, and expel the supernatant fraction that possesses the lysed cell content, membranes and cell wall material. Transfer the tubes inside the chamber, gently resuspend in anaerobic and fresh McCV medium supplemented with 1% of bovine serum albumin and dispense 0.2 ml aliquots of the appropriate dilution onto surface of McCV agar (or BD agar) in plates. Spread the inoculum by gently swirling the plates or using a sterile bent glass rod. Culture the plates in a canister pressurized with H_2/CO_2 (80:20 v/v) to 100 kPa at 30 C, as described above.

Colonies develop after 7-10 days of incubation.

Side notes:

- Resuspension of whole cells of *M. voltae* into APB results in approximately 50% lysis, but 99% of the remaining cells are converted into protoplasts.

- Leaving protoplasts in the APB over extended periods of time results in additional cell lysis. Once the protoplast are formed (around 1 minute), resuspend them in growth medium supplemented with BSA as soon as possible.
- Protoplasting procedure can be applied to other *Methanococcus* species with different degrees of success, particularly to those species that tend to lyse in hypotonic solutions. For example *M. maripaludis* has a 75% conversion to protoplasts.
- During the 7-10 days of incubation of the plates, add H_2/CO_2 to maintain the canister pressure at 100 kPa.

Natural transformation of *M. voltae* protoplasts method:

1- Pressurize Balch tubes containing protoplasts in APB (see "Protoplasts formation and regeneration in agar medium") for 10 seconds with 137 kPa of H_2/CO_2 and centrifuge (4000 rpm) for 20 min at room temperature. Resuspend to 0.5-1.5 x 10^9 protoplasts per ml in McCV medium-BSA.

2- Anaerobically add 15 µg of plasmid DNA in TE buffer to 3.6 ml of the protoplast suspension. To avoid oxygen, it is recommended to transfer DNA inside the chamber and allow it to exchange gases for 2 hours prior to mixing with protoplasts.

3- Incubate without shaking for 2 hours at 30 C.

4- Pressurize the tube with 137 kPa of H_2/CO_2, centrifuge (4000 rpm) for 20 min at room temperature, discard the supernatant and resuspend the pellet in 6 ml of McCV medium-BSA.

5- Incubate the cell suspension at 30 C in a water bath with gentle agitation. Pressurize the tube with H_2/CO_2 to 275 kPa after 2, 4 and 22 hours of incubation.

6- After 24 hours of incubation, centrifuge for 20 minutes at room temperature, discard the supernatant, and resuspend the pellet in 0.6 ml of McCV-BSA.

7- Inside the chamber, prepare serial dilutions of the transformed protoplast suspension. Plate on agar McCV supplemented with 5 μg/ml puromycin. For plating and incubation procedures refer to the protoplast regeneration protocol.

Side notes:

- Prior to and after the 24-hour incubation, withdraw 0.1 ml aliquot of protoplasts to determine the cell number by direct counting.
- From 10^9 protoplasts and 15 μg of plasmid DNA, expect to obtain 705 $\pm$ 4 transformants per μg of DNA.
- The frequency of spontaneous mutations to puromycin resistance in *M. voltae* is less than 10^{-7} cell $^{-1}$ (36).
- Cells are collected by centrifugation directly in the Balch tubes in a J2-21 centrifuge (Beckman), using a JA-14 rotor and specific adaptors for these tubes. These glass tubes break under high centrifugation. For that reason, it is recommended not to exceed a rotational speed of 4000 rpm.

Anaerobic electroporation buffer (AEB) preparation: The AEB was originally proposed by Micheletti *et al* (28). Here the same buffer is described with one modification, the solution is reduced using H_2S instead of Ti(III) citrate. In a 50-ml serum vial, prepare 10 ml aliquots of AEB (0.1 M HEPES (pH 6.5), 0.4 M sucrose, 0.05 M NaCl, 0.05 M KCl, 0.05 M $MgCl_2$ and resazurin (1 mg/l) under N_2 and autoclave for 15 minutes at 121 C. Inject 0.5 ml of H_2S gas and shake the contents. When resazurin turns colorless, flush the vial headspace with N_2. *M. voltae* is stable in AEB for at least 15 min.

Transformation by electroporation of *M. voltae* protoplasts:

1- Wash protoplasts once in APB (see "Protoplasts formation and regeneration in agar medium") and resuspend to 400x their initial concentration in AEB under a H_2/CO_2 gas phase.

2- Add 50 µl of the protoplast suspension (2-5 x 10^9 protoplasts) in AEB to 5 µg of plasmid DNA contained in a small 5-ml serum bottle kept under a continuous flow of N_2/CO_2. Add AEB to a final volume of 300 µl and seal the serum bottle under a headspace of N_2/CO_2 (137 kPa). Place the serum bottle and an electroporation cuvette in ice for 10 minutes.

3- Simultaneously, flush the serum bottle and the electroporation cuvette with N_2/CO_2. Transfer the complete contents of the serum bottle to the electroporation cuvette. Cap the cuvette (under N_2/CO_2 gas phase) and electroporate with a 2.9 msec time constant (400 V, 125 µF, 0.2-cm electrode gap and 5-ohm resistor in-line).

4- Remove the cap and immediately flush the cuvette with N_2/CO_2.

5- Using a sterile 1-ml glass syringe, add 0.9 ml of McCV-BSA to the cuvette. Add the complete contents of the cuvette to a Balch tube with 9 ml of McCV-BSA and pressurize with 137 kPa H_2/CO_2.

6- Centrifuge at 4000 rpm for 20 minutes at room temperature and expel the supernatant. Resuspend the pellet in 2 ml of McCV-BSA and pressurize the tube with 137 kPa of H_2/CO_2.

7- Incubate the tube at 30 C for 24 hours and pressurize with H_2/CO_2 to 275 kPa after 2, 4 and 22 hours of incubation.

8- Centrifuge the tube at 4000 rpm for 20 min at room temperature and resuspend the pellet in 0.2 ml of McCV-BSA.

9- Dilute with McCV-BSA and plate in McCV-BSA agar in the presence and absence of puromycin as described in "Natural transformation of *M. voltae* protoplasts method".

Side notes:

- Electroporation cuvettes should have a headspace of N_2/CO_2 instead of H_2/CO_2 to avoid an explosion hazard.
- Prior to and after incubation for 24 hours, withdraw 0.1 ml aliquot of protoplasts to determine the cell number by direct counting.
- From 10^9 protoplasts and 5 μg of plasmid DNA, expect to obtain 177 $\pm$ 74 transformants per μg of DNA.

5.2. Polyethylene glycol (PEG)-mediated transformation of *M. maripaludis*

Preparation of transformation buffer (TB) and Transformation buffer + polyethylene glycol (TB+PEG): In a 100-ml beaker mix the components for TB (50 mM Tris Base, 0.35 M sucrose, 0.38 M NaCl, 0.00001% resazurin, and 1 mM $MgCl_2$. For TB+PEG, also add 40% wt/vol PEG8000). Adjust pH to 7.5 using HCl and transfer to a 50-ml serum bottle.

In a small serum bottle prepare a 50x cysteine/DTT solution (2.5% cysteine-HCl, 50 mM DTT). After adjusting the pH to 7.5 with Tris base, take into the anaerobic chamber.

Exchange gases by sparging with N_2 for 1 or 3 hours for TB and TB+PEG, respectively. Transfer the serum bottles inside the chamber and add 1 ml of 50x cysteine/DTT solution to 50 ml TB or TB+PEG. Incubate the solutions unclosed over-night or until the resazurin turns colorless. Filter-sterilize TB (or TB+PEG) using disposable 0.2 μm filters and transfer 5 ml

aliquots into sterile Balch tubes. Seal the tubes with a serum stoppers and secure with aluminum seals. Pressurize them for 10 seconds with 137 kPa of H_2/CO_2 (80:20 vol/vol).

Side notes:

- PEG takes several hours to dissolve. Mixing of the TB+PEG solution over-night is recommended.
- It is recommended to filter-sterilize TB+PEG using a vacuum-driven filtration system. The TB+PEG solution may take several hours to filter.

PEG-mediated transformation of *M. maripaludis*:

1- In a Balch tube, grow a 5 ml *M. maripaludis* culture in McCV medium to an A_{600} of 0.7-1.0.

2- Pressurize the Balch tube for 10 seconds with 137 kPa of H_2/CO_2 and centrifuge at 2300 rpm for 10 minutes at room temperature.

3- Gently invert each tube and discard the supernatant. Using a Vacutainer needle, add 5 ml of TB buffer (see "Protoplast formation and regeneration on agar medium").

4- Pressurize the Balch tube with 137 kPa of H_2/CO_2 and centrifuge at 4000 rpm for 20 minutes at room temperature.

5- Discard the supernatant and add anaerobically 0.375 ml of TB. Resuspend the pellet by swirling.

6- Transfer into the chamber, add 0.8-1.5 μg of plasmid DNA, mix and take out from the chamber.

7- Flush for 1 minute with 100% N_2 and pressurize for 10 seconds using the same gas.

8- Carefully add 0.225 ml of TB-PEG with a sterile syringe. Allow the TB-PEG to fall directly onto the cell suspension without touching the sides of the tube. Mix thoroughly by swirling.

9- Incubate without shaking at 37 C for 1 hour.

10- Prepare two Balch tubes with 5 ml of McCV medium plus 100 µl of 2.5% Na_2S*9H_2O. Pressurize them with 137 kPa of H_2/CO_2.

11- Using a Vacutainer needle, add 5 ml of McCV medium to the transformants. Resuspend the pellet by swirling.

12- Pressurize the serum tube for 10 seconds with 137 kPa of H_2/CO_2 and centrifuge at 4000 rpm for 20 minutes at room temperature.

13- Using a Vacutainer needle, discard the supernatant of the transformation tube and add 5 ml of McCV medium. Resuspend the pellet.

14- Flush the serum tube with H_2/CO_2 and pressurize it at 275 kPa for 10 seconds. Incubate with shaking at 37 C overnight.

15- Introduce the sample inside the chamber and make serial dilutions with McCV medium. Plate in McCV agar in the presence and absence of puromycin as described (see "Preparation of McN or McCV solid media by the petri dishes plate").

Side notes:

- Check absorbance (600 nm) of the transformation tubes during the procedure to ensure viability.

6. Other genetic techniques

6.1 Markerless mutagenesis

Efficient markerless mutation requires a positive and negative selection. Positive selection is used to create a merodiploid from the homologous recombination of the recombinant plasmid into the genome. Negative selection is used to select for the removal of the plasmid by a second homologous recombination event. Normally, both selections are provided in the plasmid constructed for mutagenesis. (31) developed a markerless mutation procedure for *M. maripaludis* with a positive selection for neomycin resistance and negative selection for sensitivity to the base analog 8-azahypoxanthine. This system was based on a markerless mutation system designed for *M. acetivorans* (37).

In this section, we describe a modification of the method for markerless mutagenesis in *M. maripaludis*, which was developed to be used in strain Mm901. This strain was derived from strain S2 by deletion of the gene *upt*, which encodes for uracil phosphoribosyltransferase and confers 6-azauracil sensitivity. The plasmid pCRUPTNEO carries the positive selection cassette (Neo^R), the gene for negative selection (*upt*) under the control of the promoter P_{hmv}, and a multi-cloning site for construction of the deletion (Figure 2-4). The complete process for markerless mutagenesis for the strain Mm901 of *M. maripaludis* is schematically represented in figure 2-5. The original method for markerless mutations, which was based in the susceptibility of *M. maripaludis* to the base analog 8-azahypoxanthine (31), required the use of the strain Mm900. This strain was derived from strain S2 of *M. maripaludis* by deletion of the gene *hpt*, which encodes for the enzyme hypoxanthine phosphoribosyltransferase. Recently, it was discovered that this mutantion has a polar effect on hydrogen metabolism. For this reason, this alternative method was developed.

Markerless mutagénesis for *M. maripaludis* method:

1- Amplify the gene of interest with more than 500 bp of flanking DNA to allow for homologous recombination and clone the resulting fragment in a desired plasmid.

2- Make an *in frame* deletion of the gene of interest, amplifying by PCR the edges of the gene and the flanking DNA without the internal portion of the gene and clone in the plasmid pCRUPTNEO at any of the 2 multi-cloning sites (Figure 2-4).

3- Transform the plasmid in *M. maripaludis* strain Mm901 and select for neomycin resistance (see "Transformation methods").

4- Pick one colony and grow it overnight in 5 ml of McCas medium (see below) in the presence of neomycin.

5- From this culture, inoculate 0.05 ml into 5 ml of McCas medium without neomycin.

6- Take aliquots of 0.1 ml and plate in McCas agar with 6-azauracil (0.25 μg/ml).

7- Pick several colonies and inoculate them in McCas medium. Screen by Southern blotting to distinguish mutant colonies from wild type.

Side notes:

- McCas media refer to McCV medium but without yeast extract. Yeast extract contains nucleobases and reduce the sensitivity to base analogs.
- 6-azauracil is added to agar medium before pouring plates from a stock of 10 mg/ml in 0.25 M NaOH.

6.2 Random mutagenesis with ethylmethanesulfonate

Ethylmethanesulfonate (EMS, $C_3H_8O_3S$) is a volatile organic solvent that is mutagenic and carcinogenic. It produces random mutations in DNA and RNA by nucleotide substitution, specifically by guanine alkylation. This dangerous compound has been successfully used to

randomly mutagenize *Methanobacterium ivanovii* (18) and methanococci (20)**.** The method used in this latter work is described here.

Ethylmethanesulfonate random mutagenesis:

1- Prepare a 0.2 M EMS stock solution in distilled water. Incubate it unstoppered for 24 hours in the anaerobic chamber. Inside the chamber, filter-sterilize the solution by passage through a 0.2 µm filter.

2- Grow methanococci in 5 ml of the appropriate medium under 200 kPa H_2/CO_2 (80:20 v/v) at 37 C (see "Liquid media preparation") until an A_{600} of 0.5 (4-5 x 10^8 cells/ml) is obtained.

3- Mutagenize the cells by adding 0.1 ml of 0.2 M EMS to 5 ml of culture and pressurize it with H_2/CO_2 to 200 kPa. Incubate for 1 hour at 37 C.

4- Centrifuge cells in the culture tube at 1500g for 20 min at room temperature and gently invert the tube. Discard the supernatant using a Vacutainer needle and resuspend the cell pellet in 5 ml of fresh medium without EMS (see "Protoplast formation and regeneration on agar medium").

5- Repeat step 4 two additional times to remove residual EMS.

6- Inoculate 1 ml of the suspension into 5 ml of fresh medium. Incubate the cells under H_2/CO_2 (275 kPa) at 37 C. These cells can then be used for desired selection or screening procedure (see "Solid media preparation").

Side notes:

- EMS is a mutagen and carcinogen, use extreme caution when handling. It is harmful if swallowed, inhaled or absorbed through the skin.
- After 1 hour of treatment with EMS, the viable cell number is reduced to 25%.

- For effective mutagenesis, a killing curve should be prepared when the sensitivity to EMS is not known.

6.3 Selection for auxotrophic mutants

Certain analogs of purines and pyrimidines are lethal for microorganisms when they are incorporated by growing cells via the salvage pathways. *M. voltae* and *M. maripaludis* are susceptible to base analogs (10, 20). Indeed, a strategy to enrich for auxotrophic mutants based in this characteristic was developed for *M. maripaludis* (see below). In this section a technique is described for the isolation of auxotrophs mutants using base analogs. This method was first developed for the isolation of acetate auxotrophs after EMS mutagenesis (20). In the protocol below, McCV is a rich permissive medium and McN medium lacks the nutrient essential for growth of the auxotrophy.

Selection for auxotrophic mutants:

1- Transfer 1 ml of the mutagenized culture to a fresh tube of 5ml McCV medium, pressurize with H_2/CO_2 (275 kPa) and let it grow until the A_{660} has reached approximately 0.4 at 37 C with shaking.

2- Transfer 0.2 ml of culture to a fresh tube of 5-ml McN medium tube that contains 5 mg of each of 6-azauracil and 8-azahypoxanthin. Pressurize the tube with H_2/CO_2 (275 kPa) and incubate for 48 hours at 37 C with shaking.

3- Harvest cells by centrifugation and wash twice with fresh McN medium. Resuspend the pellet in 5-ml McCV medium and incubate to mid exponential phase.

4- Repeat step 3 and 4 twice.

5- Plate the culture on McCV agar plates (see “Solid media preparation”). Replica plate colonies (using sterile toothpicks) onto McN and McCV agar plates. After growth, pick

colonies that grow on McCV but not in McN plates with a sterile syringe. Inoculate into a tube with 5-ml of McCV medium. This last step should be done inside the chamber.

Side notes:

- Composition of McN medium varies depending of the auxotroph of interest.
- Base analogs have low solubility in McCV medium. Two analogs are used to prevent selection of spontaneous resistance mutants.

6.4 Transposon approaches in *Methanococcus*

Over the years there have been two reported trials to establish transposon mutagenesis systems in hydrogenotrophic methanogens. In 1995, an *in vitro* transposon insertion mutagenesis technique was developed to study the *nif* operon of *M. maripaludis*. In that study, a 15.6-kb *nif* region from *M. maripaludis* was cloned in a λ vector and was used as a target for a transposon insertion using a Mudpur transposon that encoded the puromycin transacetylase gene for puromycin resistance. *M. maripaludis* was transformed using different mutagenized versions of the vector DNA, where each was found to replace the wild type DNA (7)**.** This methodology represents the first successful attempt of *in vitro* transposon mutagenesis in *M. maripaludis*.

In 2009 an *in vitro* random transposon system was reported for a rapid knockout of the tryptophan operon in *M. maripaludis*. In that study, a plasmid pKJ331-KAN, bearing a transposon derived from the Tn5 transposable element and the kanamycin and puromycin resistance markers, was used in the mutagenesis of a plasmid that contained the tryptophan operon of *M. maripaludis* in an *in vitro* reaction with the hyperactive Tn5 transposase (*tnp*). The resulting plasmids were transformed into *E. coli* using the kanamycin selection. After selection enrichment and purification from *E. coli*, they were transformed into *M. maripaludis* using the puromycin selection (35). Using this newly developed method, different tryptophan auxotrophs

were obtained, which demonstrated the function of various genes on the biosynthesis of tryptophan.

Acknowledgements:

We thank Boguslaw Lupa for help in preparation of the figures. This work was supported in part by a grant from NIH.

References:

1. **Agha-Amiri, K., and A. Klein.** 1993. Nucleotide sequence of a gene encoding a histone-like protein in the archaeon *Methanococcus voltae*. Nucleic Acids Res **21:**1491.
2. **Argyle, J. L., D. L. Tumbula, and J. A. Leigh.** 1995. Neomycin resistance as a selectable marker in *Methanococcus maripaludis*. Appl. Environ. Microbiol **62:**4233-4237.
3. **Balch, W. E., and R. S. Wolfe.** 1976. New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressureized atmosphere. Appl Environ Microbiol **32:**781-91.
4. **Beckler, G. S., L. A. Hook, and J. N. Reeve.** 1984. Chloramphenicol acetyltransferase should not provide methanogens with resistance to chloramphenicol. Appl Environ Microbiol **47:**868-9.
5. **Beneke, S., H. Bestgen, and A. Klein.** 1995. Use of the *Escherichia coli* uidA gene as a reporter in *Methanococcus voltae* for the analysis of the regulatory function of the intergenic region between the operons encoding selenium-free hydrogenases. Mol Gen Genet **248:**225-8.

6. **Bertani, G., and L. Baresi.** 1987. Genetic transformation in the methanogen *Methanococcus voltae* PS. J Bacteriol **169:**2730-8.

7. **Blank, C. E., P. S. Kessler, and J. A. Leigh.** 1995. Genetics in methanogens: transposon insertion mutagenesis of a *Methanococcus maripaludis* nifH gene. J Bacteriol **177:**5773-7.

8. **Bock, A., and O. Kandler.** 1985. Antibiotic sensitivity of archaebacteria, p. 525-544. *In* C. Woese, R. S. Wolfe, J. R. Sokatch, and L. N. Ornston (ed.), The Bacteria, vol. 3. Academic Press, New York.

9. **Bokranz, M., A. Klein, and L. Meile.** 1990. Complete nucleotide sequence of plasmid pME2001 of *Methanobacterium thermoautotrophicum* (Marburg). Nucleic Acids Res **18:**363.

10. **Bowen, T. L., and W. B. Whitman.** 1987. Incorporation of Exogenous Purines and Pyrimidines by *Methanococcus voltae* and Isolation of Analog-Resistant Mutants. Appl Environ Microbiol **53:**1822-6.

11. **Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J. F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. Geoghagen, and J. C. Venter.** 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii.* Science **273:**1058-73.

12. **Cohen-Kupiec, R., C. Blank, and J. A. Leigh.** 1997. Transcriptional regulation in Archaea: in vivo demonstration of a repressor binding site in a methanogen. Proc Natl Acad Sci U S A **94:**1316-20.

13. **Gardner, W. L., and W. B. Whitman.** 1999. Expression vectors for *Methanococcus maripaludis*: overexpression of acetohydroxyacid synthase and beta-galactosidase. Genetics **152:**1439-47.

14. **Gernhardt, P., O. Possot, M. Foglino, L. Sibold, and A. Klein.** 1990. Construction of an integration vector for use in the archaebacterium *Methanococcus voltae* and expression of a eubacterial resistance gene. Mol Gen Genet **221:**273-9.

15. **Hedderich, J., and W. Whitman.** 2004. Physiology and Biochemistry of the Methane-Producing Archaea, p. 1050-1079, The Prokaryotes, vol. 2. Springer-Verlag, New York.

16. **Hippie, H.** 1984. Maintenance of methanogenic bacteria, p. 69-81, Mantenance of microorganisms: A Manual of Laboratory Methods. Academic Press, London.

17. **Hungate, R. E.** 1950. The anaerobic mesophilic cellulolytic bacteria. Bacteriol Rev **14:**1-49.

18. **Jain, M. K., and J. G. Zeikus.** 1987. Methods for Isolation of Auxotrophic Mutants of *Methanobacterium ivanovii* and Initial Characterization of Acetate Auxotrophs. Appl Environ Microbiol **53:**1387-90.

19. **Koval, S. F., and K. F. Jarrell.** 1987. Ultrastructure and biochemistry of the cell wall of *Methanococcus voltae*. J Bacteriol **169:**1298-306.

20. **Ladapo, J., and W. B. Whitman.** 1990. Method for isolation of auxotrophs in the methanogenic archaebacteria: role of the acetyl-CoA pathway of autotrophic CO2 fixation in *Methanococcus maripaludis*. Proc Natl Acad Sci U S A **87:**5598-602.

21. **Lie, T. J., and J. A. Leigh.** 2007. Genetic screen for regulatory mutations in *Methanococcus maripaludis* and its use in identification of induction-deficient mutants of the euryarchaeal repressor NrpR. Appl Environ Microbiol **73:**6595-600.

22. **Lie, T. J., and J. A. Leigh.** 2002. Regulatory response of *Methanococcus maripaludis* to alanine, an intermediate nitrogen source. J Bacteriol **184:**5301-6.

23. **Lin, W., and W. B. Whitman.** 2004. The importance of porE and porF in the anabolic pyruvate oxidoreductase of *Methanococcus maripaludis*. Arch Microbiol **181:**68-73.

24. **Liu, Y.** 2010. Taxonomy of methanogens, p. 550-558, Handbook of Hydrocarbon and Lipid Microbiology, vol. 1. Springer-Verlag Berlin Heidelberg, New York.

25. **Liu, Y., R. H. White, and W. B. Whitman.** Methanococci use the diaminopimelate aminotransferase (DapL) pathway for lysine biosynthesis. J Bacteriol **192:**3304-10.

26. **Liu, Y., and W. Whitman.** 2008. Metabolic, phylogenetic and ecological diversity of the methanogenic archaea, p. 171-189, Incredible Anaerobes. New York Acdemy of Sciences, New York.

27. **Metcalf, W. W., J. K. Zhang, E. Apolinario, K. R. Sowers, and R. S. Wolfe.** 1997. A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors. Proc Natl Acad Sci U S A **94:**2626-31.

28. **Micheletti, P. A., K. A. Sment, and J. Konisky.** 1991. Isolation of a coenzyme M-auxotrophic mutant and transformation by electroporation in *Methanococcus voltae*. J Bacteriol **173:**3414-8.

29. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, New York.

30. **Miller, T. L.** 1989. *Methanobrevibacter*, p. 2178-2183, Bergey's Manual of Systematic Bacteriology, vol. 3. Williams & Wilkins, Baltimore.

31. **Moore, B. C., and J. A. Leigh.** 2005. Markerless mutagenesis in *Methanococcus maripaludis* demonstrates roles for alanine dehydrogenase, alanine racemase, and alanine permease. J Bacteriol **187:**972-9.

32. **Nolling, J., F. J. van Eeden, R. I. Eggen, and W. M. de Vos.** 1992. Modular organization of related Archaeal plasmids encoding different restriction-modification systems in *Methanobacterium thermoformicicum*. Nucleic Acids Res **20:**6501-7.

33. **Patel, G. B., C. G. Choquet, J. H. Nash, and G. D. Sprott.** 1993. Formation and Regeneration of *Methanococcus voltae* Protoplasts. Appl Environ Microbiol **59:**27-33.

34. **Patel, G. B., J. H. Nash, B. J. Agnew, and G. D. Sprott.** 1994. Natural and Electroporation-Mediated Transformation of *Methanococcus voltae* Protoplasts. Appl Environ Microbiol **60:**903-7.

35. **Porat, I., and W. B. Whitman.** 2009. Tryptophan auxotrophs were obtained by random transposon insertions in the *Methanococcus maripaludis* tryptophan operon. FEMS Microbiol Lett **297:**250-4.

36. **Possot, O., P. Gernhardt, A. Klein, and L. Sibold.** 1988. Analysis of drug resistance in the archaebacterium *Methanococcus voltae* with respect to potential use in genetic engineering. Appl Environ Microbiol **54:**734-40.

37. **Pritchett, M. A., J. K. Zhang, and W. W. Metcalf.** 2004. Development of a markerless genetic exchange method for *Methanosarcina acetivorans* C2A and its use in construction of new genetic tools for methanogenic archaea. Appl Environ Microbiol **70:**1425-33.

38. **Sakai, S., H. Imachi, S. Hanada, A. Ohashi, H. Harada, and Y. Kamagata.** 2008. *Methanocella paludicola* gen. nov., sp. nov., a methane-producing archaeon, the first

isolate of the lineage 'Rice Cluster I', and proposal of the new archaeal order Methanocellales ord. nov. Int J Syst Evol Microbiol **58:**929-36.

39. **Stathopoulos, C., W. Kim, T. Li, I. Anderson, B. Deutsch, S. Palioura, W. Whitman, and D. Soll.** 2001. Cysteinyl-tRNA synthetase is not essential for viability of the archaeon *Methanococcus maripaludis.* Proc Natl Acad Sci U S A **98:**14292-7.

40. **Stettler, R., P. Pfister, and L. T.** 1994. Characterization of a plasmid carried by *Methanobacterium thermoautotrophicum* ZH3, a methanogens closely related to *Methanobacterium thermoautotrophicum* Marburg. Syst. Appl. Microbiol. **17:**484-491.

41. **Thomas, N. A., C. T. Pawson, and K. F. Jarrell.** 2001. Insertional inactivation of the flaH gene in the archaeon Methanococcus voltae results in non-flagellated cells. Mol Genet Genomics **265:**596-603.

42. **Tumbula, D. L., T. L. Bowen, and W. B. Whitman.** 1997. Characterization of pURB500 from the archaeon *Methanococcus maripaludis* and construction of a shuttle vector. J Bacteriol **179:**2976-86.

43. **Tumbula, D. L., R. A. Makula, and W. B. Whitman.** 1994. Transformation of *Methanococcus maripaludis* and identification of a PstI-like restriction system. FEMS Microbiol. Lett. **121:**309-314.

44. **Tumbula, D. L., and W. B. Whitman.** 1999. Genetics of *Methanococcus*: possibilities for functional genomics in Archaea. Mol Microbiol **33:**1-7.

45. **Uffen, R. L., and R. S. Wolfe.** 1970. Anaerobic growth of purple nonsulfur bacteria under dark conditions. J Bacteriol **104:**462-72.

46. **Weisberg, W. G., and R. S. Tanner.** 1982. Aminoglycoside sensitivity of archaebacteria. FEMS Microbiol. Lett. **14**.

47. **Whitman, W. B., J. Shieh, S. Sohn, D. S. Caras, and u. Premachandran.** 1986. Isolation and characterization of 22 mesophilic methanococci. Syst. Appl. Microbiol. **7:**235-240.

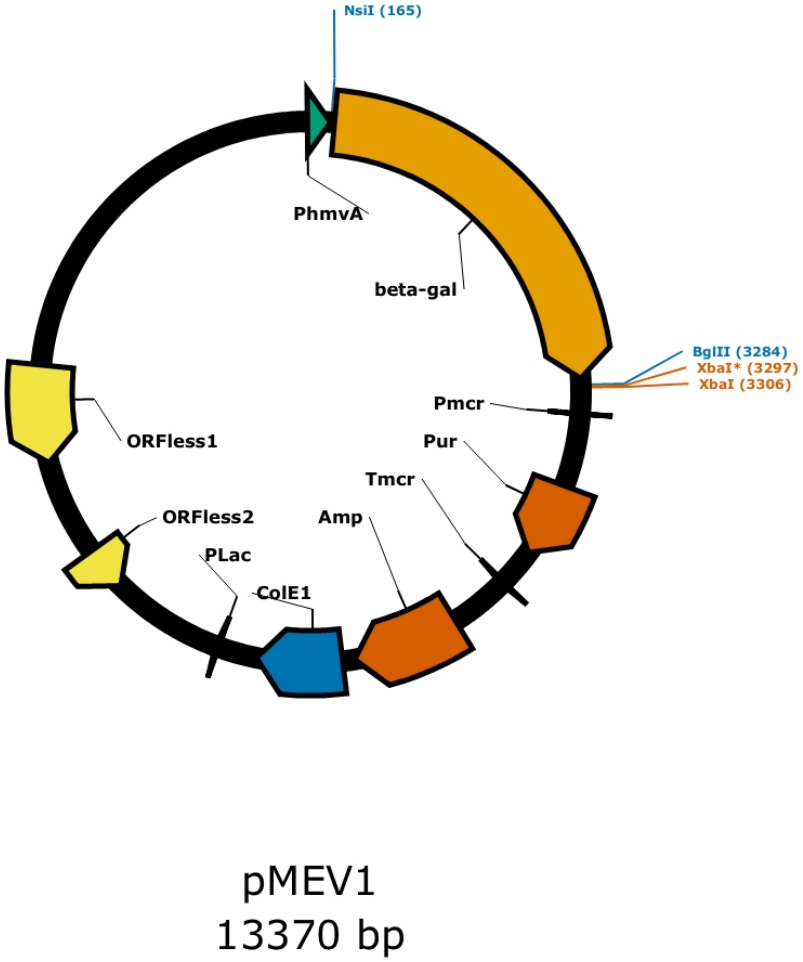

Figure 2-1. *Map of the methanococcal shuttle vector pMEV1*. The vector contains the puromycin resistance marker (Pur) for positive selection. P*hmva*, histone gene promoter from *M. voltae*; beta-gal, *lacZ* gene; P*mcr*, methyl coenzyme M reductase promoter from *M. voltae*; T*mcr*, methyl coenzyme M reductase terminator from *M. voltae*; Amp, ampicillin resistance cassette for *E. coli*; ColE1, origin of replication for *E. coli*; P*lac*, lac promoter; ORFless1 and ORFless2, possible origins of replication for *M. maripaludis*.

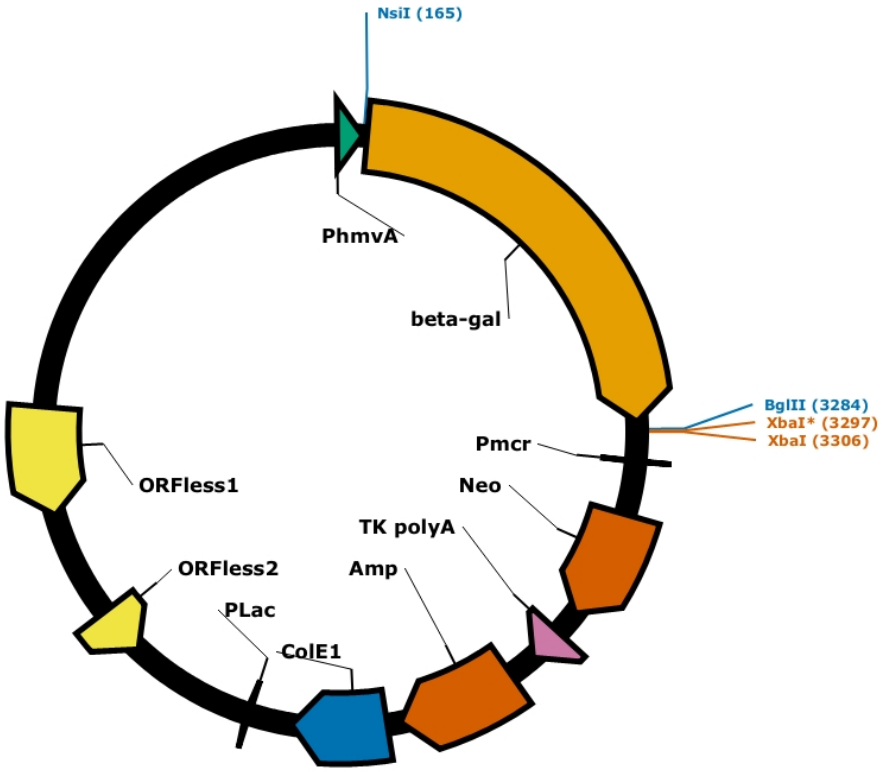

Figure 2-2. *Map of the methanococcal shuttle vector pMEV2.* The vector contains the neomycin resistance marker (Neo) for positive selection. P*hmva*, histone gene promoter from *M. voltae*; beta-gal, *lacZ* gene; P*mcr*, methyl coenzyme M reductase promoter from *M. voltae*; TK polyA, thymidine kinase polyadenylation signal sequence; Amp, ampicillin resistance cassette for *E. coli*; ColE1, origin of replication for *E. coli*; P*lac*, lac promoter; ORFless1 and ORFless2, possible origins of replication for *M. maripaludis*.

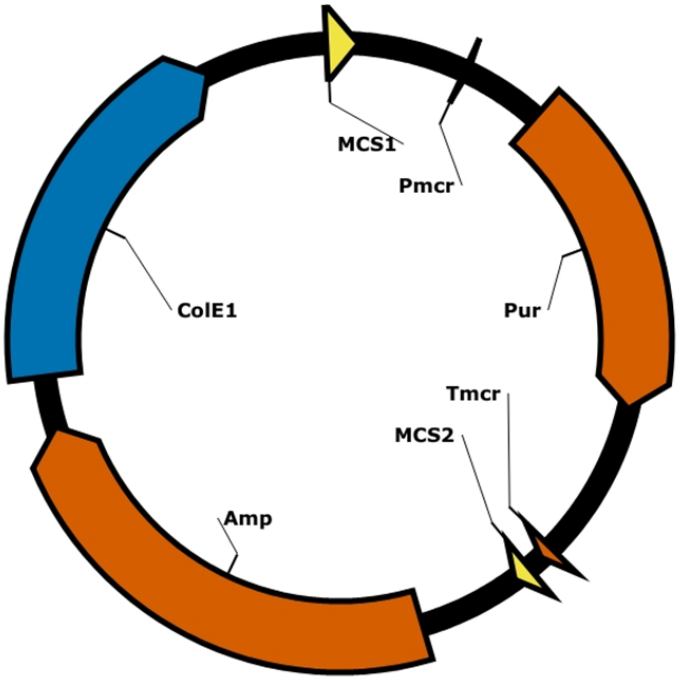

Figure 2-3. *Map of the methanococcal integration vector pIJA03* used for gene replacement. P*mcr*, methyl coenzyme M reductase promoter from *M. voltae*; Pur, puromycin transacetylase gene from *S. alboniger*; T*mcr*, methyl coenzyme M reductase terminator from *M. voltae*; MCS1, multi-cloning site 1(*Bam*HI, *Eco*RI, *Afl*III, *Mlu*I, *Nde*I, *Afl*II, *Bgl*II, *Xba*I, *Xba*I); MCS2, multi-cloning site 2 (*Eco*0109I, *Cla*I, *Afe*I, *Bmt*I, *Nhe*I, *Spe*I, *Kpn*I) Amp, ampicillin resistance cassette for *E. coli*; ColE1, origin of replication for *E. coli*.

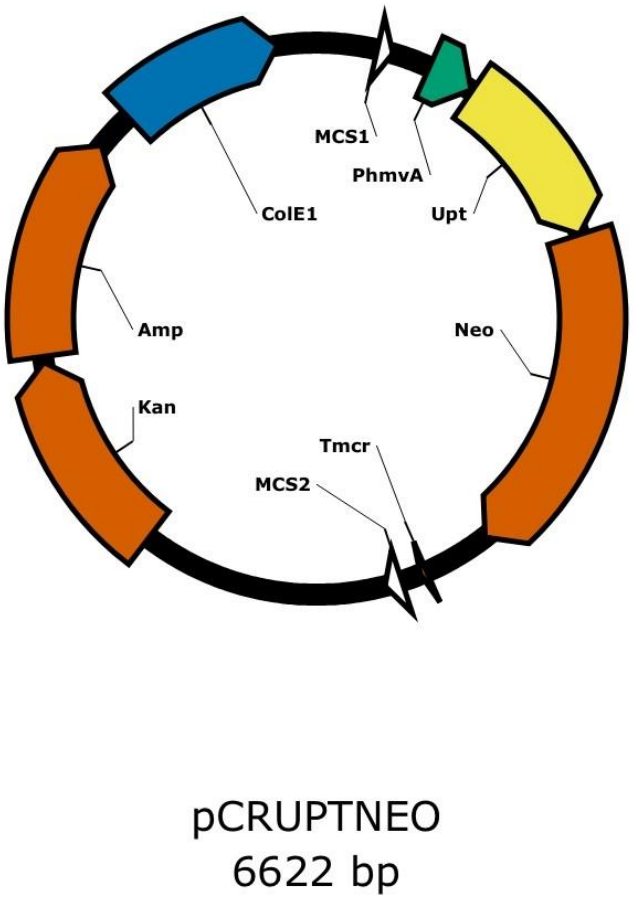

Figure 2-4. *Map of the methanococcal vector pCRUPTNEO* used for markerless mutagenesis. P*hmva*, histone gene promoter from *M. voltae*; Upt, uracil phosphoribosyltransferase gene; Neo, neomycin resistance cassette; T*mcr*, methyl coenzyme M reductase terminator from *M. voltae*; Kan, kanamycin resistance cassette for *E. coli*; Amp, ampicillin resistance cassette for *E. coli*; ColE1, origin of replication for *E. coli*; MCS1, multi-cloning site 1 (*Hind*III, *Acc*651, *Kpn*I, *Bam*HI, *Avr*II); MCS2, multi-cloning site 2 (*Apa*I, *Xba*I, *Not*I, *Afl*II).

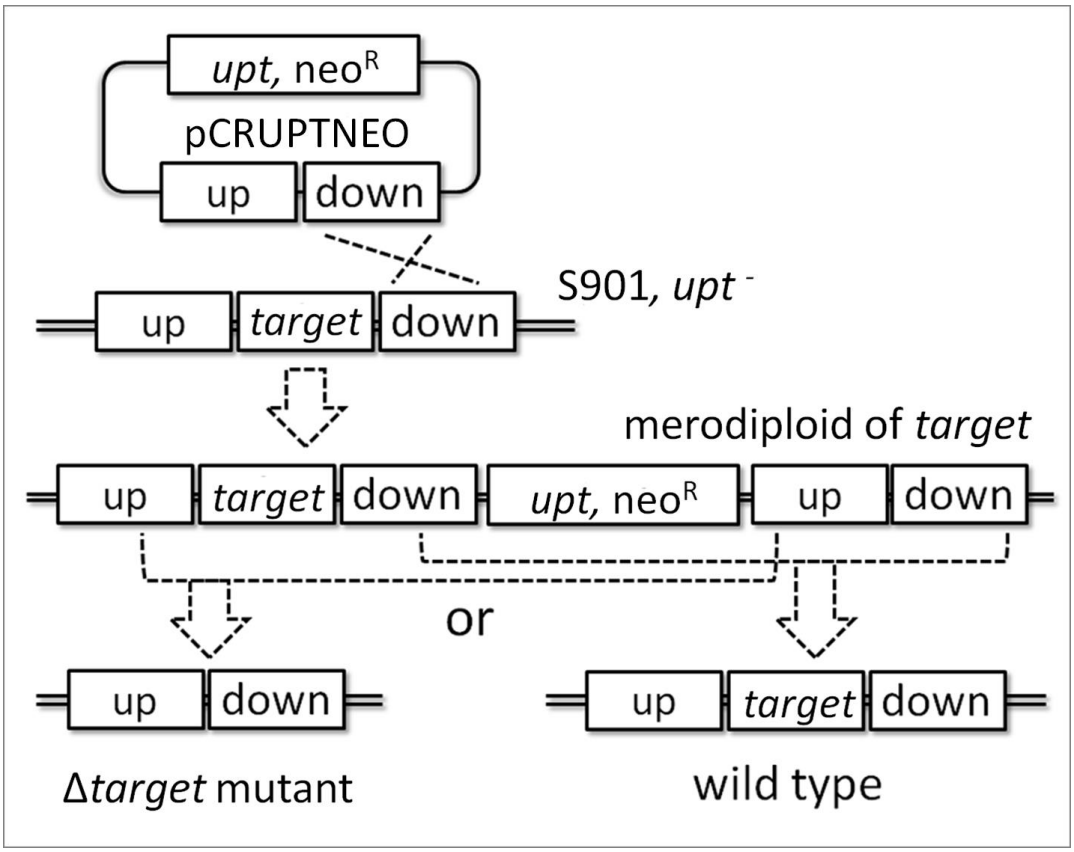

Figure 2-5. *Schematic representation of the markerless mutagenesis technique for M. maripaludis strain Mm901.* The first homologous recombination between the genomic DNA of the strain Mm901 of *M. maripaludis* and the cloned downstream (or upstream) DNA region in the plasmid pCRUPTNEO results in the formation of a merodiploid of the target gene. A second recombination can produce two different outcomes, a mutant strain for the target gene or restoration of the wild type strain. In the absence of selection, these should occur with equal frequency and are identified by screening with PCR or Southern blotting. *upt*, uracil phosphoribosyltransferase gene; neo^R, neomycin resistance cassette; *target*, gene of interest to be deleted.

Table 2-1. Characteristics of the sequenced genoms of Hydrogenotrophic methanogens

Hydrogenotrophic methanogen	Order	Genome size (bp)	Gene count	G+C percentage	NCBI Accession number
Methanosphaera stadtmanae DSM 3091	*Methanobacteriales*	1767403	1592	27.63%	NC_007681
Methanobrevibacter ruminantum M1	*Methanobacteriales*	2937203	2283	32%	NC_013790
Methanobrevibacter smithii ATCC 35061	*Methanobacteriales*	1853160	1837	31.03%	NC_009515
Methanobrevibacter smithii DSM 2375	*Methanobacteriales*	1704865	1747	31.28%	NZ_ABYW00000000
Methanobrevibacter smithii F1, DSM 2374	*Methanobacteriales*	1727775	1755	31.19%	NZ_ABYV00000000
Methanothermobacter thermautotrophicus Delta H	*Methanobacteriales*	1751377	1921	49.54%	NC_000916
Methanocella paludicola SANAE	*Methanocellales*	2957635	3064	54.92%	NC_013665
Methanocella sp. RC-I	*Methanocellales*	3179916	3170	54.60%	NC_009464
Methanocaldococcus fervens AG86	*Methanococcales*	1485061	1630	32.22%	NC_013156
Methanocaldococcus infernus ME	*Methanococcales*	1328194	1538	33.53%	NC_014122
Methanocaldococcus jannaschii DSM 2661	*Methanococcales*	1664970	1765	31.29%	NC_000909
Methanocaldococcus vulcanis M7	*Methanococcales*	1746329	1729	31%	NC_013407

Methanococcus aeolicus Nankai-3	*Methanococcales*	1569500	1552	30.04%	NC_009635
Methanococcus maripaludis C5	*Methanococcales*	1780761	1896	32.98%	NC_009135
Methanococcus maripaludis C6	*Methanococcales*	1744193	1888	33.42%	NC_009975
Methanococcus maripaludis C7	*Methanococcales*	1772694	1855	33.28%	NC_009637
Methanococcus maripaludis S2	*Methanococcales*	1661137	1772	33.10%	NC_005791
Methanococcus vannielii SB	*Methanococcales*	1720048	1752	31.33%	NC_009634
Methanococcus voltae A3	*Methanococcales*	1936387	1768	28.56%	NC_014222
Methanocorpusculum labreanum Z	*Methanomicrobiales*	1804962	1822	50.01%	NC_008942
Methanoculleus marisnigri JR1	*Methanomicrobiales*	2478101	2557	62.06%	NC_009051
Methanosphaerula palustris E1-9c	*Methanomicrobiales*	2922917	2866	55.35%	NC_011832
Methanospirillum hungatei JF-1	*Methanomicrobiales*	3544738	3304	45.15%	NC_007796
Methanopyrus kandleri AV19	*Methanopyrales*	1694969	1729	61.16%	NC_003551

Table 2-2. Media components.

Basal medium for *Methanococcus maripaludis* (McN)	
Composition per 100 mL	Amount
Glass-distilled water	50 ml
General salt soultion	50 ml
K_2HPO_4 (14 g/l)	1 ml
Trace mineral solution	1 ml
Iron stock solution	0.5 ml
Rezasurin (1 g/l)	0.1 ml
NaCl (293 g/l)	7.5 ml
$NaHCO_3$	0.5 g

Complex medium for *Methanococcus maripaludis* (McCV)	
McN	100 ml
Sodium acetate*3 H_2O (136 g/l)	1 ml
Yeast extract	0.2 g
Casamino acids	0.2 g
Trace vitamin solution	1 ml

Modification of Basal medium for *Methanococcus voltae*[a]	
Add Sodium acetate*3 H_2O (136 g/l)	1 ml
Add isoleucine	0.05 g
Add leucine	0.1 g
Add pantoyl lactone (1 mM)	1 ml

Modification of Basal medium for *Methanococcus vannielli*[a]	
Reduce NaCl (293 g/l)	1.8 ml
Modification of basal medium for growth on formate (McF)[a]	
Reduce glass-distilled water	30 ml
Omit NaCl (293 g/l)[b]	0 ml
Add sodium formate (5 M)	8 ml
Add sodium glycylglycine (1 M, pH 7)	20 ml

[a]All modifications are based in 100 ml volume of the solution

[b]NaCl is replaced by Sodium formate

Table 2-3. Composition of stock solutions for media.

General salt solution	
Composition per 1 L	Amount
KCl	0.67 g
$MgCl_2$*6 H_2O	5.50 g
$MgSO_4$*7 H_2O	6.90 g
NH_4Cl	1.00 g
$CaCl_2$*2 H_2O	0.28 g
Trace mineral solution	
Nitriloacetic acid	1.5 g
$MnSO_4$* H_2O	0.1 g
$Fe(NH_4)_2(SO_4)_2$*6H_2O	0.2 g
$CoCl_2$*5H_2O	0.1 g
$ZnSO_4$*7 H_2O	0.1 g
$CuSO_4$*5 H_2O	0.01 g
$NiCl_2$*6H_2O	0.025 g
Na_2SeO_3	0.2 g
Na_2MoO_4*2H_2O	0.1 g
Na_2WO_4*2H_2O	0.1 g
Iron stock solution	
$Fe(NH_4)_2(SO_4)$2*6H_2O	2 g
HCl	100 µl

Trace Vitamin solution	
Biotin	2 mg
Folic acid	2 mg
Pyridoxine hydrochloride	10 mg
Thiamine hydrochloride	5 mg
Riboflavin	5 mg
Nicotinic acid	5 mg
DL-calcium pantothenate	5 mg
Vitamin B_{12}	0.1 mg
p-aminobenzoic acid	5 mg
Lipoic acid	5 mg

CHAPTER 3

GENOME-SCALE ANALYSIS OF GENE FUNCTION IN THE HYDROGENOTROPHIC METHANOGENIC ARCHAEAON *METHANOCOCCUS MARIPALUDIS* [2]

[2]**Sarmiento, F., Mrázek, J. and W.B. Whitman.** 2013. *Proc Natl Acad Sci U S A.* 110(12): 4726-31. Reprinted here with permission of the publisher.

Abstract

A comprehensive whole-genome analysis of gene function by the Tn-seq methodology has been successfully implemented in a representative of the Archaea domain. Libraries of transposon mutants were generated for the hydrogenotrophic, methanogenic archaeon *Methanococcus maripaludis* S2 using a derivative of the Tn5 transposon. About 89,000 unique insertions were mapped to the genome, which allowed for the classification of 526 genes or about 30% of the genome as possibly essential or strongly advantageous for growth in rich medium. Many of these genes were homologous to eukaryotic genes that encode fundamental processes in replication, transcription and translation, providing direct evidence for their importance in Archaea. Some genes classified as possibly essential were unique to the archaeal or methanococcal lineages, such as that encoding DNA polymerase PolD. In contrast, the archaeal homolog to the gene encoding DNA polymerase B was not essential for growth, a conclusion which was confirmed by construction of an independent deletion mutation. Thus, PolD and not PolB is likely to play a fundamental role in DNA replication in methanococci. Similarly, 121 hypothetical ORFs were classified as possibly essential and are likely to play fundamental roles in methanococcal information processing or metabolism that are not established outside this group of prokaryotes.

Introduction

Methanogenic archaea are obligate anaerobic prokaryotes and widely distributed in O_2-free environments where electron acceptors other than CO_2 have been depleted. Methanogenesis is a highly specialized anaerobic respiration with a distinctive biochemistry composed of unusual coenzymes and catalysts whose roles are poorly understood (6). In anaerobic environments, methanogenesis plays a key role by catalyzing the terminal step of carbon mineralization and maintaining an extremely low partial pressure of H_2. Methane, the final product of this process, is also a significant greenhouse gas with about 80% of the atmospheric methane produced by these archaea (20).

Our understanding of the methanogenic archaea is far from complete. For instance, the methanogen *Methanococcus maripaludis* S2 possesses 1,779 genes, only a few of which have been characterized and an even smaller portion has been studied in detail. Close to 800 genes remain annotated as hypothetical proteins awaiting proper identification (7). Much of this uncertainty is shared with other archaea, where many of the fundamental life processes have not been investigated in the same detail as in the bacteria and eukaryotes. To address these issues, the function of methanococcal genes was evaluated by a saturation mutagenesis technique. Whole-genome libraries of Tn5 transposon mutants were constructed, and the individual mutations were mapped following enrichment of the transposon-chromosomal DNA junctions and Illumina sequencing (4). Because mutations in genes that are likely to be essential or strongly advantageous for growth are lethal or rapidly lost from the library, they may be identified by their low frequency in the libraries. This methodology has some limitations: gene and pathway redundancy may mask essential processes, polarity may cause genes downstream of an essential gene to falsely appear essential, truncated genes may retain some activity and falsely

appear nonessential, and small essential genes may escape detection. Moreover, it is difficult to distinguish genes that are truly essential, ie. absolutely required for growth under the specified conditions, from those that are advantageous for growth and rapidly diluted out from the libraries by faster growing mutants. Although definitive assignments of essentiality still require detailed analyses of each gene, the methodology generates hypotheses about the nature of specific genes and a great deal of insight into specific questions regarding methanogens as well as more general questions about the genetic, biochemistry and physiology of archaea.

Results and Discussion

Mutants mapping and criteria for gene classification

Methanococcus maripaludis strain S2 was randomly mutated by transformation with a derivative of the Tn5 transposon (Tn5<KAN-2-pac>) comprising a kanamycin resistance cassette for selection in *Escherichia coli* and a puromycin resistance cassette for selection in methanogenic archaea. Two high-density mutant collections of approximately 60,000 and 30,000 individual mutants were constructed. The locations of the insertions were mapped in the original libraries, which was formed after growth with puromycin selection for about 20 generations (T0), and after growth without puromycin selection for seven (T1) or fourteen generations (T2) in rich or minimal medium (Table S3-1). Across all libraries, about 92% of the sequence reads were mapped to the genome. While insertions were generally evenly distributed around the genome, a few locations possessed a high number of reads. Presumably, these reads represented hotspots for the transposon insertion, which is not uncommon for the Tn5 transposon (5), or artifacts of the sequencing process. For a complete list of reads mapped to the genome, refer to Dataset S1.

M. maripaludis is polyploid and contains 30-55 genomes per cell, depending upon the growth phase (9). To prevent accumulation of heterologous genes, gene conversion rapidly

homogenizes the genomes following insertions (9). Therefore, in our experimental design, insertions in nonessential genes are expected to rapidly replace the wild-type alleles during growth with puromycin selection (Figure S3-1). In fact, mutants homozygous for transposon insertions are readily isolated in the tryptophan operon under permissive conditions (19). Because the full replacement of the wild-type allele is lethal for essential genes, cells are expected to maintain both the wild-type and mutant genes during puromycin selection. When the puromycin selection is removed, the mutant alleles are expected to be rapidly lost, and wild-type alleles are expected to dominate (9). To test this model, the fraction of genomes with insertions was determined by real-time PCR for library 1. The relative copy numbers of the *pac* cassette served as a marker for the transposon, and the *cdhA* gene, which was present in a single copy on the genome, served as a marker for the genome. Following puromycin selection at T0, 0.71$\pm$0.20 ($\pm$ standard deviation) of the genomes possessed the *pac* cassette by this measure. In complex medium, the ratio dropped to 0.47$\pm$0.09 and 0.37$\pm$0.03 in the T1 and T2 libraries, respectively, presumably because gene conversion led to the loss of deleterious insertions in cells that were heterozygous and selection for wild-type cells. In contrast, in minimal medium the ratio initially dropped to 0.50$\pm$0.06 and then increased to 0.78$\pm$0.16 in the T1 and T2 libraries, respectively. The increase was attributed to selection for mutants with insertions in a specific gene with faster growth than the wild-type under these conditions (see below). Therefore, these results were consistent with the expectations for polyploidy cells. Interestingly, polyploidy offers a unique advantage in Tn-seq experiments because it allows direct demonstration of insertions in essential genes during antibiotic selection. In monoploid cells, selection against insertions in essential genes is inferred from the insertion density of nonessential genes and is never directly observed.

Genes essential to growth were expected to have fewer insertions than nonessential genes. A sliding window method was used to identify the number of insertions in genes (see Appendix A). In this approach, an essentiality index (EI) was calculated based upon the number of insertions within 800 bp windows. The EI was then calibrated empirically without making assumptions about the expected distribution of insertions. In the T0 libraries, insertions were found in nearly all genes, including many expected to be essential for growth, and the number of unique insertions per window resembled a Poisson distribution, with a maximum between 20 and 25 insertions per window (Figure 3-1A). This unimodal distribution indicated that there was little selection against insertions in most of the windows. In T1 and T2 cells grown in rich medium without puromycin selection, the number of unique insertions per window acquired a bimodal distribution, with a large increase in the number of windows with no or only few insertions (Figure 3-1A). Presumably, these windows encompassed possibly essential or strongly advantageous genes.

To test this hypothesis, the EI of 104 genes assumed to be essential based upon our general knowledge of the physiology and biochemistry of methanococci and 89 genes assumed to be nonessential were examined in rich medium (Figure. 3-1C, Dataset S2). The EI for the known essential genes in T2 were 0-3, indicating that few insertions remained following growth without puromycin. In contrast, the EI of the nonessential genes were in the range 7-81, with 99% nonessential genes having EI > 11. For that reason, an EI $\leq$ 3 was considered diagnostic of possibly essential genes, and an EI $\geq$ 11 was considered diagnostic of nonessential genes. Genes with an EI of 4-10 were unassigned and presumably comprised genes that were advantageous for growth but nonessential as well as small genes whose essentiality cannot be reliably determined by this method. For genes smaller than the window size of 800 bp, insertions at neighboring

genes might increase the EI even in the absence of insertions within the gene itself. Similar methods were used to calibrate the EI for library 2, and genes were classified as possibly essential when they satisfied the criteria in at least one library and were unassigned in the second library. Following growth in minimal medium, fewer unique insertions were found, and the range of EI was much lower (Figure 3-1B). This decrease resulted from a large increase in the number of reads in one specific gene (MMP1511) as discussed below. Calibration of the EI in these libraries yielded cut-offs of the EI for possibly essential and nonessential of ≤ 2 and ≥ 5, respectively.

A gene's index in library 1 was highly correlated with its index in library 2, and so the EI appeared to be reproducible (Figure 3-2A). In fact, following growth in rich medium, only genes for seven hypothetical proteins were classified possibly essential in library 2 and nonessential in library 1. In the absence of additional information, these genes were classified as unassigned. For essential and highly advantageous genes, the EI is also expected to depend upon the rate of gene conversion and the fitness of the resulting mutants as well as stochastic factors affecting the original number of insertions. The mechanism of gene conversion in the euryarchaeotes is not well understood. Although gene conversion is rapid and relatively independent of the fitness of the resulting homozygous mutant, the rate depends in part on the extent of the genetic difference between the genomes (12). Although the Tn5 transposon insertions are all identical, it is possible that the rate of gene conversion also depends upon the sequence surrounding the insertion site. Slowly growing mutants are also being selected against at the same time as gene conversion occurs. Therefore, differences in the observed EIs could result from differences in the gene conversion rates as well as the growth rates of the resulting homozygous mutants, and it is not possible to distinguish between truly essential and highly advantageous genes. In contrast, for

nonessential genes where the fitness of the resulting mutants is close to that of the wild-type, gene conversion should occur prior to the formation of the T0 library, and the EI should be constant in the subsequent libraries. Consistent with this prediction, the EI for many genes was highly correlated in the T0, T1 and T2 libraries, especially when the EI was >11 and the genes were presumably nonessential (Figure 3-2B and C).

To evaluate the effect of polarity on the functional categorization of genes, the EI of genes in the 209 operons previously identified in *M. maripaludis* S2 were examined (23). If the insertion of the Tn5 transposon interrupts transcript, genes upstream of a possibly essential gene will always appear possibly essential. However, this was not the case. Twelve of the operons possessed possibly essential genes with an EI $\leq$3 downstream of nonessential genes. The apparent lack of polarity is consistent with earlier observations of transcriptional read through from the *pac* cassette (18).

To further validate the essentiality index, this classification scheme was compared to the experimental evidence for the essentiality of 63 genes from published studies (Dataset S3). For 62 of these genes, the essentiality index correctly predicted whether or not the gene was required for growth under the conditions used here. For example, independent genetic evidence demonstrated that three genes of the eha hydrogenase (*ehaHIJ*), which encode the cation translocator of this enzyme complex, were essential (14). Their low essentiality indices in the current studies were consistent with this conclusion. However, for the gene *dapL*, which encodes a large bifunctional enzyme required for lysine biosynthesis, the EI remained in the range of 4-5 even following growth in minimal medium without lysine. These insertions were concentrated in regions outside of the catalytic and the pyridoxal 5'-phosphate binding sites of the gene, suggesting that a truncated protein with some functionality may have been formed. While these

results confirm the overall validity of this approach, they demonstrate the need for cautious interpretation of mutations within largely uncharacterized genes.

Possibly essential genes for growth in rich medium

In rich medium at T2, 526 genes were classified as possibly essential for growth. Many of the essential genes were required for fundamental biological processes, such as methanogenesis, replication, transcription and translation (Figure S3-2). Of the remaining genes, 834 genes were classified as nonessential and 419 could not be confidently assigned. A complete list of the EIs for the *M. maripaludis* genes under all the conditions examined is given in Dataset S4.

By these criteria, genes from many different functional categories were possibly essential for growth in rich medium (Table 3-1). In addition, over 45% of the protein encoding genes are annotated as hypothetical proteins, and 121 genes or 16% of them were possibly essential (Figure S3-2). In contrast, most of the genes assigned to cellular processes, signal transduction, or transport were nonessential. Some specific categories are analyzed in more detail in the Appendix A.

Replication. Because most mutations in the replicative system are lethal, replication in the Archaea has largely been studied in vitro, and the Tn-seq approach is especially informative. Of the 26 genes assigned to replication in methanococci, 16 were possibly essential. Like other euryarchaeotes, methanococcal genomes encode two replicative DNA polymerases (2), PolD and PolB. PolB is proposed to synthesize the leading strand and PolD to synthesize the lagging strand during replication (8). In addition, mutations in genes for both polymerases in *Halobacterium* were lethal (1). However, only the genes for PolD, MMP0008 and 0026, were possibly essential in methanococci (Figure 3-3A). To confirm this observation, a deletion mutant was constructed where the majority of the *polB* gene, encompassing most of the DNA

polymerase domain including the two essential aspartyl residues and all of the exonuclease domain, was replaced with the *pac* cassette (Figure 3-4 and Appendix A). Growth of this mutant was indistinguishable from the wild-type in rich and minimal medium (Figure 3-4B). These results suggested that PolD is the major replicative DNA polymerase in methanococci and PolB plays a secondary role or is redundant with another protein. In support of this conclusion, in *Thermococcus kodakarensis* PolD can be coisolated with PCNA and other proteins of the archaeal replication fork (13). Although PolB coisolated with the replication proteins MCM2 and RPA3, most of its associations were with proteins whose roles have yet to be defined. Methanococci also possess MMP1230, which possesses a nucleotidyltransferase domain common in DNA polymerase homologs. This gene is found in all methanogens and *Archaeoglobus* but missing in other related euryarchaeotes. Although essential, its distribution is more consistent with a role in tetrahydromethanopterin (H_4MPT) metabolism, which is shared by these archaea, than replication. In eukaryotes and archaea, minichromosome maintenance (MCM) proteins form homo- or heteromeric complexes that play fundamental roles in the initiation and progression of the replication fork. Four probable MCM homologs are found in *M. maripaludis*, but only 3 of them are expressed (21). Only one of these expressed MCM genes, MMP0030, was possibly essential (Figure 3-3B). Similar results have been found in *T. kodakarensis* where 3 MCM homologs are present, but only one is essential (17). In contrast, in eukaryotes all six of the MCM homologs are essential for the initiation of DNA synthesis (11) . These results suggest that *M. maripaludis* forms a homomeric MCM complex or a heteromer, where MMP0030 predominates. Similar MCM complexes have been observed in other methanogenic archaea, such as *M. thermautotrophicus* (3).

The genes which encode the large and small subunits of a homolog of the eukaryotic primase (MMP0009 and MMP0071) were both possibly essential. The *Pyrococcus furiosus* homolog to this primase has been previously characterized (15). In contrast, a homolog of the bacterial DnaG-type primase, MMP1286 was nonessential. A homolog of this protein in *T. kodakarensis* (TK1410) co-purified with the exosome complex (13), and the presence of both primase activities has been reported in *Sulfolobus solfataricus* (24). Other possibly essential genes related to replication were the processivity factors PCNA (MMP1126 and 1711), DNA toposiomerases (MMP0956, 0989 and 1437) and DNA ligase (MMP0970). Three homologs for replication protein A (RPA) are present in *M. maripaludis* S2 (21), but only two of them were possibly essential (MMP0616 and 1032). Lastly, the flap endonuclease gene (MMP1313, *fen1*) was nonessential.

Two other genes of *M. maripaludis* S2 which encode hypothetical proteins share homology with proteins that co-purified with the replication fork complex of *T. kodakarensis* (13). Although their functions are unknown in both archaea, MMP0668 (homolog to TK1313) and MMP1392 (homolog to TK0358) were both possibly essential and may play important roles in archaeal replication.

Hypothetical proteins. A total of 121 hypothetical proteins were possibly essential (Dataset S5). Of these, 56 are conserved throughout the order *Methanococcales*, 19 are conserved by all methanogenic archaea, four are conserved by the *Euryarcheota* phylum, and only one is conserved throughout *Archaea* (MMP0694). This gene possesses several domains involved in RNA metabolism, suggesting an involvement in post-transcriptional RNA modifications. The remaining 41 possibly essential hypothetical proteins are conserved at the

family, genus or species level, possibly playing major roles in the adaptations defining the life-style of *M. maripaludis*.

Possibly essential genes for growth in minimal medium

A total of 664 possibly essential genes were found in minimal medium at T2. The classification of most genes was the same as in complex medium. Many of the genes which become possibly essential in minimal medium did not have an obvious role in biosynthesis, including genes for two ferredoxins, MMP0389 and 1140, and 87 hypothetical proteins. Many genes that encode proteins involved in stress response also became possibly essential. These included MMP0684, which encodes the heat shock protein Hsp20; MMP0264, which encodes a mechanosensitive ion channel; and MMP0585, which encodes the universal stress protein or Usp. Finally two homologs of the SAM proteins, MMP0560 and 1221, became possibly essential in minimal medium.

In contrast to the other insertions, the abundance of reads in the gene MMP1511, which encodes an alanine/sodium symporter, increased from 0.6% of the total reads in the T0 libraries to 85% in the T2 libraries following growth in minimal medium. This gene is the last gene transcribed in an operon that includes alanine dehydrogenase and alanine racemase (16). However, the number of reads in these other genes did not increase. This result indicated that the inactivation of the symporter stimulated the growth rate in minimal medium. Indeed, the relative fitness of the mutant, calculated by the Malthusian parameter (22), was 1.5 times those of mutants of nonessential genes. Because this effect was not observed in rich medium, it appears to depend upon the absence of amino acids.

Conclusions

The Tn-seq technology was successfully implemented in *M. maripaludis*, generating the first comprehensive database of possibly essential and nonessential genes for an archaeon. While these results provide fresh insights into numerous metabolic and molecular pathways of these unique prokaryotes, it is important to note that the classifications of essentiality are hypotheses about the nature of specific genes. Definitive assignments of essentiality require detailed analyses of each gene, which is not possible in a global survey of the genome. Many of the genes for the fundamental processes in archaeal replication, transcription and translation, which have been identified largely based upon their similarity to eukaryotic homologs, proved possibly essential for growth. These results provide direct evidence for these roles and for a close relationship between archaea and eukaryotes. In contrast, the gene for the archaeal-specific DNA polymerase PolD was possibly essential, suggesting that it performs a fundamental role in replication. Thus, archaeal replication also possesses unique features. Interestingly, PolD is absent from the genomes of the crenarchaeotes, suggesting that PolB is the replicative DNA polymerase in this phylum. If true, this observation implies an unanticipated variability in archaeal replication. Similarly, many genes encoding hypothetical proteins proved to be possibly essential. Because many of these genes were only found in specific phylogenetic or physiological groups, they may enable important but unidentified functions unique to these archaea. For instance, many of these genes may be involved in coenzyme biosynthesis or other currently poorly described processes in methanogenesis. The further functional analysis of these genes will help to unveil many of the unsolved mysteries of the third domain of life.

Finally, it is remarkable that the number of possibly essential genes in this methanogenic archaeon is very similar to that found in bacteria. In *M. maripaludis*, 526 genes were classified as

possibly essential for growth on rich medium, which correspond to 30% of the total genes in the genome of *M. maripaludis*. Many of these genes encoded monomer biosynthesis and reflected the inability of these lithotrophs to utilize organic nutrients. For the heterotrophic bacteria tested, the number of essential genes ranges from 271 to 642 genes (10). These results imply that the total number of genes required for growth in laboratory media is fairly small. If only limited functionality is needed for growth under these conditions, only a few genes will appear possibly essential. Certainly, the number of genes required to encode the core informational processes is only a small fraction of the total. Alternatively, essentiality of function is often masked by the inherent redundancy of biological systems so that many mutations are compensated by alternative pathways.

Materials and methods

Additional details of the methods are described in Appendix A.

***In vivo* transposon mutagenesis**

Production of stable transposomes was achieved by incubation of 100 ng of the Tn5<KAN-2-pac> transposon with 2 µl of EZ-Tn5 transposase (1U/ µl, Epicentre) and transformed into *M. maripaludis* S2. After transformation cells were spread onto McCm agar plates supplemented with puromycin and incubated for 6 days in the presence of 100 kPa of H_2/CO_2 (80:20, v/v) at 37 °C. Puromycin-resistant colonies were washed off the plates and stored at -80 °C.

Mutant library passages

An aliquot of 300 µl from the mutant library frozen stock was diluted in 5 ml of McN medium to an absorbance (600 nm) of 0.5-0.6. This suspension was defined as T0. From this dilution, $4x10^7$ viable cells were inoculated into 20 ml of either McCm or McN medium. In both cases, ampicillin was added. In the first passage (T1), cells were grown to an absorbance (600 nm) of

0.5-0.6 at 37°C. For the second passage (T2), cells were transferred to fresh medium and grown to an absorbance of 0.5-0.6. Thus, each passage comprised about seven generations and a 100-fold amplification of the cell number. Genomic DNA was extracted from 5 ml of each passage using the ZR fungal/bacterial DNA miniprep (Zymo Research) and resuspended in TLE buffer (10mM Tris-HCl buffer [pH 8], 0.1mM EDTA).

High-throughput insertion tracking by deep sequencing (HITS or Tn-seq)

Five μg of genomic DNA were sheared to an average fragment size of 500 bp, and Illumina DNA libraries were prepared by ligating specific indexed linkers to the DNA fragments. Transposon-chromosome junctions were enriched using a biotinylated probe and PCR. Sequencing of the enriched DNA fragment library was performed at the Genome Services Laboratory at the Hudson Alpha Institute for Biotechnology, Huntsville, Alabama. Custom primer single-end sequencing (50 bp) was carried out on a HiSeq Flowcell v1.5 using a HiSeq2000 sequencer.

Acknowledgments

We thank Travis Glenn and Roger Nilsen for valuable technical assistance. This project has been funded by a contract from NIH. JM was supported by the National Science Foundation grant BDI-0950266.

References

1. **Berquist, B. R., P. DasSarma, and S. DasSarma.** 2007. Essential and non-essential DNA replication genes in the model halophilic Archaeon, *Halobacterium* sp. NRC-1. BMC Genet **8:**31.
2. **Cann, I. K., and Y. Ishino.** 1999. Archaeal DNA replication: identifying the pieces to solve a puzzle. Genetics **152:**1249-67.

3. **Fletcher, R. J., B. E. Bishop, R. P. Leon, R. A. Sclafani, C. M. Ogata, and X. S. Chen.** 2003. The structure and function of MCM from archaeal M. *Thermoautotrophicum*. Nat Struct Biol **10:**160-7.

4. **Gawronski, J. D., S. M. Wong, G. Giannoukos, D. V. Ward, and B. J. Akerley.** 2009. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. Proc Natl Acad Sci U S A **106:**16422-7.

5. **Goryshin, I. Y., J. A. Miller, Y. V. Kil, V. A. Lanzov, and W. S. Reznikoff.** 1998. Tn5/IS50 target recognition. Proc Natl Acad Sci U S A **95:**10716-21.

6. **Hedderich, R., and W. B. Whitman.** 2006. Physiology and biochemistry of the methane-producing Archaea, p. 1050-1079. *In* M. Dworkin, S. Falkow, E. Rosenberg, K. Schleifer, and E. Stackebrandt (ed.), The Prokaryotes, Third ed, vol. 2. Springer-Verlag, New York.

7. **Hendrickson, E. L., R. Kaul, Y. Zhou, D. Bovee, P. Chapman, J. Chung, E. Conway de Macario, J. A. Dodsworth, W. Gillett, D. E. Graham, M. Hackett, A. K. Haydock, A. Kang, M. L. Land, R. Levy, T. J. Lie, T. A. Major, B. C. Moore, I. Porat, A. Palmeiri, G. Rouse, C. Saenphimmachak, D. Soll, S. Van Dien, T. Wang, W. B. Whitman, Q. Xia, Y. Zhang, F. W. Larimer, M. V. Olson, and J. A. Leigh.** 2004. Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis.* J Bacteriol **186:**6956-69.

8. **Henneke, G., D. Flament, U. Hubscher, J. Querellou, and J. P. Raffin.** 2005. The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. J Mol Biol **350:**53-64.

9. **Hildenbrand, C., T. Stock, C. Lange, M. Rother, and J. Soppa.** 2011. Genome copy numbers and gene conversion in methanogenic archaea. J Bacteriol **193:**734-43.

10. **Juhas, M., L. Eberl, and J. I. Glass.** 2011. Essence of life: essential genes of minimal genomes. Trends Cell Biol **21:**562-8.

11. **Kearsey, S. E., and K. Labib.** 1998. MCM proteins: evolution, properties, and role in DNA replication. Biochim Biophys Acta **1398:**113-36.

12. **Lange, C., K. Zerulla, S. Breuert, and J. Soppa.** 2011. Gene conversion results in the equalization of genome copies in the polyploid haloarchaeon *Haloferax volcanii*. Mol Microbiol **80:**666-77.

13. **Li, Z., T. J. Santangelo, L. Cubonova, J. N. Reeve, and Z. Kelman.** 2010. Affinity purification of an archaeal DNA replication protein network. MBio **1**.

14. **Lie, T. J., K. C. Costa, B. Lupa, S. Korpole, W. B. Whitman, and J. A. Leigh.** 2012. Essential anaplerotic role for the energy-converting hydrogenase Eha in hydrogenotrophic methanogenesis. Proc Natl Acad Sci U S A **109:**15473-8.

15. **Liu, L., K. Komori, S. Ishino, A. A. Bocquier, I. K. Cann, D. Kohda, and Y. Ishino.** 2001. The archaeal DNA primase: biochemical characterization of the p41-p46 complex from Pyrococcus furiosus. J Biol Chem **276:**45484-90.

16. **Moore, B. C., and J. A. Leigh.** 2005. Markerless mutagenesis in Methanococcus maripaludis demonstrates roles for alanine dehydrogenase, alanine racemase, and alanine permease. J Bacteriol **187:**972-9.

17. **Pan, M., T. J. Santangelo, Z. Li, J. N. Reeve, and Z. Kelman.** 2011. *Thermococcus kodakarensis* encodes three MCM homologs but only one is essential. Nucleic Acids Res **39:**9671-80.

18. **Porat, I., W. Kim, E. L. Hendrickson, Q. Xia, Y. Zhang, T. Wang, F. Taub, B. C. Moore, I. J. Anderson, M. Hackett, J. A. Leigh, and W. B. Whitman.** 2006. Disruption of the operon encoding Ehb hydrogenase limits anabolic CO2 assimilation in the archaeon *Methanococcus maripaludis*. J Bacteriol **188:**1373-80.

19. **Porat, I., and W. B. Whitman.** 2009. Tryptophan auxotrophs were obtained by random transposon insertions in the Methanococcus maripaludis tryptophan operon. FEMS Microbiol Lett **297:**250-4.

20. **Reeburgh, W. S.** 2003. Global methane biogeochemistry, p. 65-89. *In* R. F. Keeling (ed.), Treatise on Geochemistry, vol. 4. Elsevier-Pergamon, Oxford, UK.

21. **Walters, A. D., and J. P. Chong.** 2009. *Methanococcus maripaludis*: an archaeon with multiple functional MCM proteins? Biochem Soc Trans **37:**1-6.

22. **Woods, R. J., J. E. Barrick, T. F. Cooper, U. Shrestha, M. R. Kauth, and R. E. Lenski.** 2011. Second-order selection for evolvability in a large *Escherichia coli* population. Science **331:**1433-6.

23. **Yoon, S. H., D. J. Reiss, J. C. Bare, D. Tenenbaum, M. Pan, J. Slagel, R. L. Moritz, S. Lim, M. Hackett, A. L. Menon, M. W. Adams, A. Barnebey, S. M. Yannone, J. A. Leigh, and N. S. Baliga.** 2011. Parallel evolution of transcriptome architecture during genome reorganization. Genome Res **21:**1892-904.

24. **Zuo, Z., C. J. Rodgers, A. L. Mikheikin, and M. A. Trakselis.** 2010. Characterization of a functional DnaG-type primase in archaea: implications for a dual-primase system. J Mol Biol **397:**664-76.

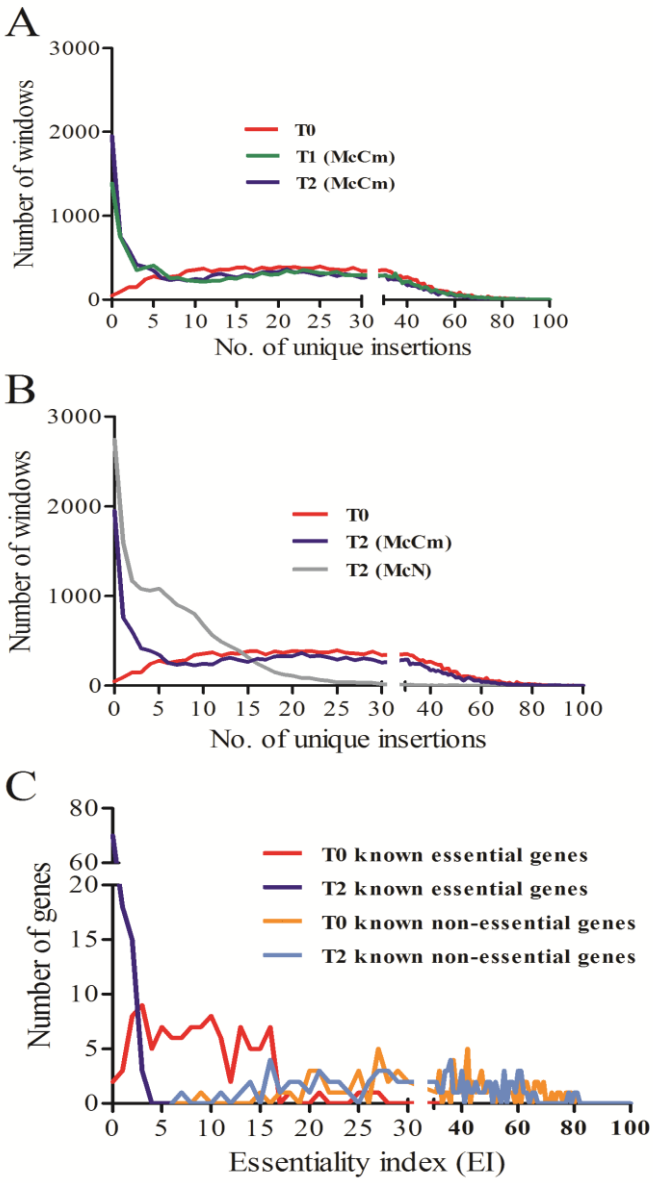

Figure 3-1. Distributions of the numbers of unique insertions in 800 bp windows in library 1. (A) Distribution following growth for seven (T1) and 14 generations (T2) in the absence of antibiotic in McCm. T0 is the initial library. (B) Distribution following growth in rich (McCm) or minimal (McN) medium in the absence of antibiotic. (C) Distribution of the essentiality index for 104 representative essential and 89 representative nonessential genes.

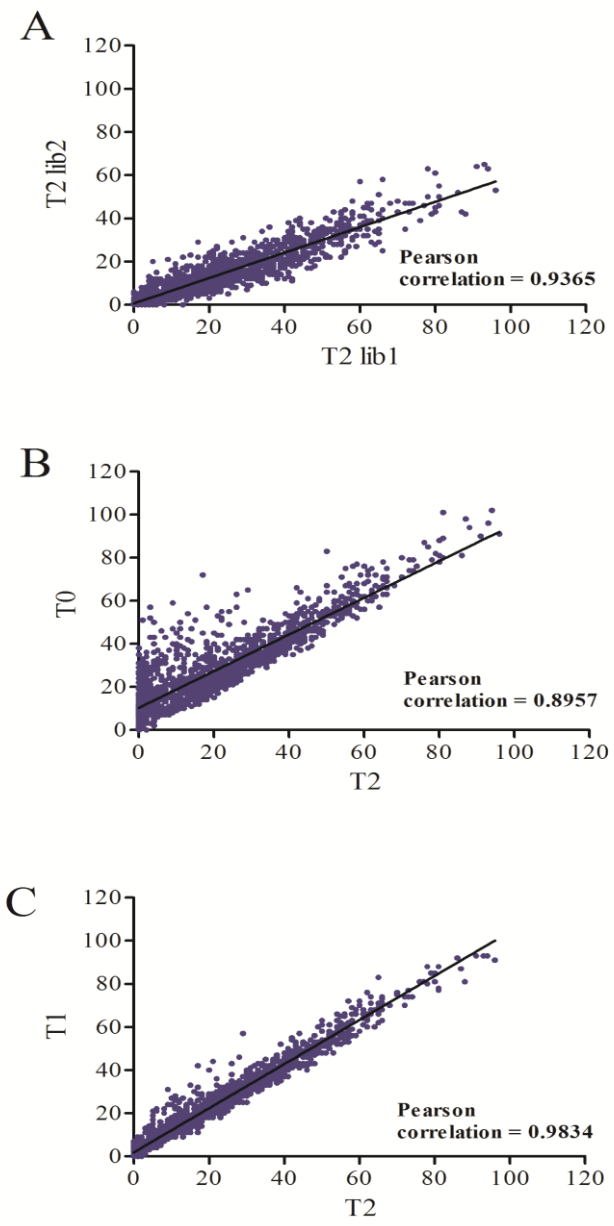

Figure 3-2. Reproducibility and stability of EI. (A) The correlation of a gene's index in library 1 to its index in library 2. (B) Correlation of the EI in the T0 and T2 of library 1. (C) Correlation of EI in the T1 and T2 of library 1.

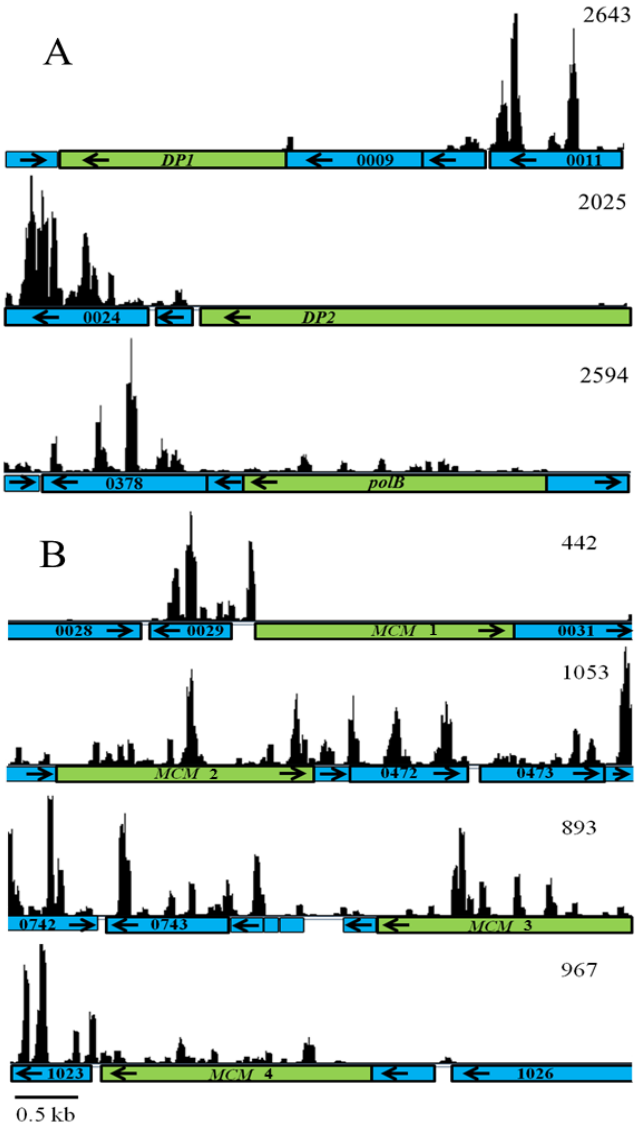

Figure 3-3. Distribution of reads across selected genes of *M. maripaludis* S2 genome. The y-axis is the number of reads within a window size of three. The numbers in the top right corners represent the maximum number of reads in each segment. (A) The DP1 and DP2 genes (green) encode the small and large subunits of the DNA polymerase type II (PolD), respectively. *polB* (green) encodes DNA polymerase B or PolB. (B) The four homologs for the MCM genes (green), which encode the minichromosome maintenance proteins. Surrounding genes are represented in blue. Numbers indicate the MMP identification, and arrows indicate the direction of transcription. Initial plots were generated using Artemis.

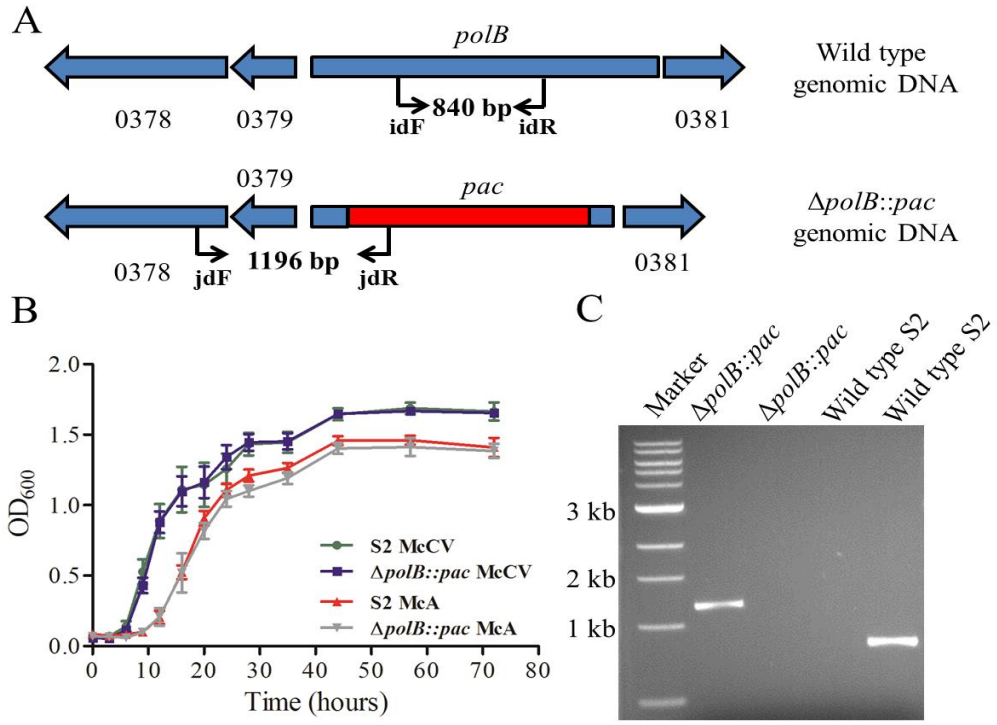

Figure 3-4. Characterization of the Δ*polB*::*pac* mutant strain. (A) Genetic maps of the gene *polB* in the wild-type strain and the replacement of almost the complete gene *polB* by the pac cassette in the Δ*polB*::*pac* mutant strain S123. Numbers indicate the MMP identification, and the black arrows indicate primers used for PCR amplification. (B) Growth curves of the wild-type and Δ*polB*::*pac* mutant strain in complex medium (McCV) and minimal medium + acetate (McA). The error bars indicate the standard deviation of four independent replicates. (C) Genotypic characterization of the Δ*polB*::*pac* mutant strain by PCR amplification. Lane 1, Standard 1kb ladder (New England Biolab); Lane 2 and 3, PCR amplifications of a junction fragment between genomic DNA and the *pac* cassette using primers jdF and jdR and an internal portion of the *polB* gene using primers idF and idR, respectively, for genomic DNA of Δ*polB*::*pac* mutant strain S123; Lane 3 and 4, same PCR amplifications of genomic DNA from the wild type strain *M. maripaludis* S2.

Table 3-1. Known protein-encoding genes related to fundamental biological processes in *M. maripaludis* S2.

Biological process	Sub-process[a]	Possibly essential[b]	Unassigned[b]	Nonessential[b]
Energy metabolism	Methanogenesis (32)	*mer, mch, fwdHFGDAC, mcrBDCGA, mtrEDCBAGH, or900, ftr, atwA*	*fwdB*	*hmd, fmdE1, mtd, fmdE2ACB1B2*
	ATP generation (9)	*atpHIKECFABD*		
	Membrane-bound hydrogenases (34)	*ehaABCDEFGHIJKLMNOT*	*ehbQ, ehbJ,* 1626, *ehbF*	*ehaPQRS, ehbCD, ehbN, ehbA, ehbOMLK, ehbG, ehbE*
	Cytoplasmic hydrogenases (17)	*frcD, vhuUAGD,*	*vhuB*	*frcBG, frcA, vhcDGAB, fruADGB*
Replication	DNA polymerases (3)	*DP1, DP2*		*polB*
	Replication factors (23)	*priL, mcm1, priS, rfcB, rfcA, rpa2, topA, lig, top6B, rpa1, pcnA, top6A, pcnA gins15*	*rnhA*	*mcm2, mcm3, mcm4, rpa3, dnaG, fen1, rnhB, smc1*
Transcription	RNA polymerase (12)	*rpoP, rpoL, rpoE', rpoD,rpoN, rpoK, rpoHB2B1A1A2*	*rpoF*	
	Transcription factors (8)	*tfe, tfb, tbp, nusA, nusG*	1015, *spt4*	*tfs*
Translation	tRNA synthetase and related proteins (26)	0212,0255, *tyrS, metS, alaS, thrS, valS,* 0688, *proS, leuS, serS, gltX, argS, pheT, lysS, ileS, pheS, trpS, hisS, aspS*		0002, *truA,* 0377, 0693, 0816, *cysS*
	Ribosomal proteins (60)	*rplX, rpl31e, rpl40e, rpl37ae, rpl12p, rpl1P, rps17E, rpl34e, rpl24e, rps28e, rpl7ae, rps2P, rps3ae rps6e, rpl10e, rps13p, rps4p, rps11p, rpl18e, rpl13p, rps9p, rps12p, rps7p, rps10p, rpl22p, rps3p, rpmC, rps17p, rpl14p, rpl24p, rps4e, rpl5p, rps14p, rps8p, rpl6p, rpl32e, rpl19e, rpl18p, rps5p, rpl30p, rpl15p, rpl3p, rpl4p, rplW, rpl2p, rps19p, rps15p, rps8e, rpl44e*	*rpl21e, rps19e, rplP0, rpl15, rps24e, rps27ae, rpl37e, rpl30e, rpl11p, rps27e*	*rpl39e*
	Translation factors (14)	*aIF6, infB, aIF2*β, *aIF1A, aIF5A,* 1131, *aIF2*γ, *aEF2,selB, aEF1*α, *aEF1*β	*aIF2α*	0738, *aIF2BI*

[a] The number in parenthesis corresponds to the total number of known genes for that specific sub-process.
[b] Genome annotation code number (MMP#) is used for genes which do not possess a gene symbol.

CHAPTER 4

TESTING GENE FUNCTION

FROM MEASUREMENTS OF MUTANT FITNESS BY TN-SEQ TECHNOLOGY

Introduction

The unusual coenzymes found in the methanoarchaea are integral parts of the methanogenesis pathway and play key roles as C1 carriers or electron donors (Figure 4-1). Surprisingly, work on the biosynthesis of these coenzymes has been developing at a slow pace, and the biosynthetic pathways for these coenzymes are only partially known. The majority of these biosynthetic pathways have been proposed by measuring the incorporation and distribution of isotopes into different molecules using NMR or mass spectrometry. Additionally, the activity of certain enzymes has been identified in cell extracts. In contrast, little information related to the genes which encode the enzymes involved in this pathways or their regulation is available. Genes involved in approximately 23 reactions of the biosynthesis of the different coenzymes remain unidentified (Table 4-1) and have become a critical barrier to the complete understanding of methanogenesis at the physiological and ecological levels. The current state of knowledge of each one of those coenzymes and its biosynthesis is presented as follows.

Methanofuran: The pathway for the biosynthesis of methanofuran has been proposed (12). Isotope incorporation patterns and ^{13}C NMR studies suggest that the furan moiety is biosynthesized from phosphoenolpyruvate (PEP) and dihydroxyacetone-phosphate (DHAP) in the first step of methanofuran biosynthesis (7, 50). The furan moiety is reduced to form a 2-furaldehyde, which reacts with tyramine and two molecules of glutamate. In the final step of

methanofuran biosynthesis, it is proposed that the intermediate forms a peptide bond with 1,3,4,6-hexanetetracarboxylic acid (HTCA)-CoA to form methanofuran. HTCA is derived from 2-oxoglutarate, acetyl-CoA and CO_2 and a pathway for its biosynthesis in *Methanosarcina thermophila* TM-1has been proposed (7, 52). Only one enzyme has been identified in this pathway, a L-tyrosine decarboxylase which produces tyramine from tyrosine. This enzyme is encoded by the gene *mfnA*, which is widely distributed among euryarchaea (26). However, a role for this gene *in vivo* has not been tested in genetic experiments

Tetrahydromethanopterin: The structure of tetrahydromethanopterin (H_4MPT) is similar to that of tetrahydrofolate (H_4F). It has been shown that the pterin ring of both molecules originate from GTP (51). In the first step of H_4MPT biosynthesis, GTP cyclohydrolase catalyzes the conversion of GTP to H_2neopterin 2', 3'-cyclic phosphate (H_2N-Cp) (18). The second step is the hydrolysis of this cyclic phosphate into two different products: H_2neopterin 2'-phosphate and H_2neopterin 3'-phosphate. The enzymes catalyzing these two first steps in the biosynthesis of tetrahydromethanopterin were identified in *Methanocaldococcus jannaschii* (33). The GTP cyclohydrolase is encoded by the gene *mptA*, and the hydrolysis is catalyzed by a cyclic phosphodiesterase encoded by the *mptB* gene. The final two steps of this pathway involve the enzymes 7,8-dihydroneopterin aldolase (DHNA, MptD) and 7,8-dihydro-6-hydroxymethylpterin diphosphokinase (6-HMDPK, MptE), which were recently identified by comparative genomics and biochemically verified after cloning and expression of the *M. jannaschii* genes in *E. coli* (4). These results completed the initial steps in the pterin pathways in Archaea . The biosynthesis of the non-pterin portion of tetrahydromethanopterin involves several other steps. Only the first one is known, and it is catalyzed by a ribofuranosylaminobenzene 5'-phosphate synthase (25).

Coenzyme M: CoM plays a key role in the last step of methanogenesis. The thiol of CoM forms a thioether bond with a methyl carbon transferred from tetrahydromethanopterin. Upon reduction of the methyl carbon with coenzyme B, methane and the heterodisulfide of CoM and CoB are formed. This heterodisulfide bond is reductively cleaved in a key energy-yielding step to regenerate coenzyme M. Five enzymes involved in the biosynthesis of CoM are known and have been characterized (11, 13, 16, 17). In addition, the genes encoding these enzymes have been identified in diverse methanogens, including *M. maripaludis*. However, in the genomes of the *Methanosarcina* species, genes involved in the first three steps of the pathway are absent, suggesting that an alternative pathway exists. The proposed pathway for the biosynthesis of CoM starts with the sulfonation of PEP by a phosphoenolpyruvate sulfotransferase (ComA). Then, a phosphosulfolactate phosphatase (ComB) hydrolyses phosphosulfolactate and a (*R*)-sulfolactate dehydrogenase (ComC) catalyses the oxidation of the (*R*)-sulfolactate intermediate to form sulfopyruvate. In the fourth step, a sulfopyruvate decarboxylase (ComDE) catalyzes the decarboxylation of sulfopyruvate to form sulfoacetaldehyde. For the final postulated step of CoM biosynthesis, the enzyme which catalyses the reductive thiolation of sulfoacetaldehyde to coenzyme M (ComF) has not yet been identified.

Coenzyme B: Originally, coenzyme B was detected as one of the three chromatographically separated components required to reconstitute methyl-coenzyme M reductase (MCR) (21). The biosynthesis of coenzyme B can be split into two parts: (12). the formation of a hydrophobic 7-mercaptoheptanoate group and the subsequent addition and phosphorylation of L-threonine. The first part involves the elongation of 2-oxoacids. The enzyme homocitrate synthase (HCS) catalyzes the addition of an acetyl group to 2-oxoglutarate to produce (*R*)-homocitrate. Then, homoaconitase (HACN) catalyzes, first, the dehydration of (*R*)-

homocitrate, producing *cis*-homoaconitate, and second, the hydration of this product to (*2R*, *3S*)-homoisocitrate. Finally, the homocitrate dehydrogenase enzyme (HICDH) catalyzes an oxidative decarboxylation reaction forming 2-oxodipate. In summary, this series of reaction add one methylene group on the original substrate. Interestingly, the same three enzymes are capable of extending 2-oxoadipate to 2-oxopimelate and 2-oxosuberate. This last molecule is then decarboxylated to produce 7-oxoheptanote which is converted to a thiol to generate 7-mercaptoheptanoate. The three described enzymes have been identified in *M. jannaschii* (6, 22, 23), and homologues are present in other methanogens, including *M. maripaludis*. No enzymes have yet been identified which catalyze the last two steps. In the second part of coenzyme B biosynthesis, 7-mercaptoheptanoate condenses with L-threonine and the product is phosphorylated to generate CoB. No enzymes involved in these steps have been identified.

Coenzyme F_{430}: This unique coenzyme, which plays an essential role as an integral cofactor of the methyl-coenzyme M reductase, has only been found in methanogenic and methanotrophic archaea. The biosynthesis of this nickel tetrapyrrole is probably the most unrevealed of the pathways discussed here. It is predicted that the synthesis of coenzyme F_{430} branches off the cobalamin pathway, and only one of six probable intermediates has been identified (38). No enzymes involved in this pathway have been identified or characterized. A few have been proposed to be involved, such as the cobalt chetalases CobNS and proteins related to the nitrogenase NifH and NifDK (25). However, these last three genes are nonessential for the growth of *M. maripaludis* in rich or minimal medium, which suggest that probably they are not involved in the biosynthesis of coenzyme F_{430} (unpublished data).

Coenzyme F_{420}: This coenzyme is a hydride carrier found not only in archaea but also in some Gram-reaction positive bacteria and cyanobacteria. F_{420} is an essential part of

methanogenesis, functioning as an electron carrier. It is also involved in other processes in the archaea, such as the reduction of sulfite, oxygen detoxification and electron transport (5, 24, 48). The biosynthesis of coenzyme F_{420} can be divided in two main parts. First, 7,8-didemethyl-8-hydroxy-5-deazariboflavin (F_0) is formed by the condensation of 5-amino,6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinedione with 4-hydroxypyruvate in reactions catalyzed by the CofGH enzymes (12). Both of these enzymes have been previously characterized in *M. jannaschii* (14). The second part is the conversion of F_0 to F_{420} by biosynthesizing the phosphodiester bond and lactyl moiety. With the exception of CofB, five different enzymes (CofA-E) involved in this pathway have been identified and characterized in *M. jannaschii*, (15, 19, 20, 30). The final step in coenzyme F_{420} biosynthesis is the ligation of glutamyl residue to the terminal γ-linked glutamate. This reaction is catalyzed by CofF, which has been identified and characterized in *M. jannaschii* (31).

Testing a new strategy for identifying missing biosynthetic genes. The recent advances in sequencing technology have made possible the study of whole genome sequences, generating enormous amounts of data. When high-throughput sequencing is combined with a whole-genome transposon mutant pool, it facilitates the identification of genes and their essentiality assignation for cell survival. Indeed, in our laboratory, a comprehensive whole-genome analysis of essential genes has been successfully implemented for first time in a representative of the Archaea domain (44). After assignation of essential genes, the next major challenge is to use high-throughput techniques, such as Tn-seq, to identify unknown genes involved in specific pathways or processes. The major objective is to assist in the final elucidation of their functions. Several studies for the identification of candidate genes in specific conditions have been performed in bacteria using high-throughput techniques. For instance, 169

candidate genes specifically involved in bile tolerance in *Salmonella typhi* have been identified (27), and 136 genes specifically involved in pathogenesis were identified in *Haemophilus influenza* (10).

In this study, Tn-seq technology was implemented towards the identification of specific unknown genes involved in the biosynthesis of the unusual coenzymes for methanogenesis in the methanoarchaeon *M. maripaludis*. However, only coenzyme M has been chemically synthesized, and the purification of the rest of these coenzymes is difficult and time demanding. So, as a short cut, a coenzyme extract, which was expected to contain intact forms of the unusual coenzymes and other supplements, was used to feed the cells. The expectation was that mutations in the genes involved in coenzyme biosynthesis would have an increased fitness during growth in the present of the coenzymes. In addition, Tn-seq technology was applied towards the elucidation of missing genes involved in the biosynthetic pathways of aromatic amino acids and cysteine. Finally, Tn-seq was applied to compare *M. maripaludis* mutant fitness during steady-state growth with hydrogen or phosphate limitation.

Materials and methods

Strain and culture conditions

Aliquots of a *M. maripaludis* S2 transposon mutant library (44) were grown in minimal (McN) medium reduced with 3 mM coenzyme M and supplemented with 10 mM acetate and 1mM alanine (McNd) (45). Coenzyme extract, 400 µl, 1mM aryl acids (indole acetate, phenylacetate, and hydroxyphenylacetate); or 100 µM cysteine (see below) were added to form McNde, McNda and McNdc respectively. For alanine growth experiments, *M. maripaludis* S2 was grown in minimal (McN) medium reduced with 3 mM cysteine (45). In some experiments McN was supplemented with 10 mM acetate to form McNA. Alanine was added at different

concentrations, and 1 mM proline, lysine or aspartate was supplemented when indicated (refer to appendix 1). For coenzyme extract growth experiments, *M. maripaludis* S2 was grown in minimal (McN) medium reduced with 3 mM cysteine. When indicated 1, 2, 4 or 6 % v/v of coenzyme extract (see below) was added to 5 ml of McN medium. Cultures of 20 ml were grown in 160-ml serum bottles pressurized to 137 kPa with H_2/CO_2 (80:20, v/v). Cultures of 5 mL were grown in 28-ml Balch tubes pressurized to 275 kPa with H_2/CO_2 (80:20, v/v). Before inoculation, 3 mM sodium sulfide was added as the sulfur source. Puromycin (2.5μg/ml) and/or ampicillin (100 μg/ml) were added when indicated.

For steady state conditions, *M. maripaludis* S2 transposon mutant library were shipped to Kyle Costa (University of Washington). Two-ml aliquots of the mutant library were grown in 1L chemostat under conditions of H_2 limitation/phosphate excess or H_2 excess/phosphate limitation at 37 °C as described previously (3) with the following modifications: 0.1% Casamino acids and 100 μg/ml of ampicillin were added to the media. For phosphate limitation, K_2HPO_4 was lowered from the standard 0.8 mM to 0.02 mM. For hydrogen limitation, H_2 was lowered from the standard 110 ml/min to 21 ml/min. Cultures were grown until the OD_{660} remained stable at 0.6-0.8. When the limited nutrients were restored to standard concentrations, the OD_{660} increased to ~1.7.

Coenzyme extract preparation

Approximately 10 g of *M. maripaludis* S2 cells were resuspended in 50 ml of water supplemented with 1 mM DTT and heated in a sand bath for 1 hour at 100 °C under anaerobic conditions. After cooling, cells were re-heated for 30 min at 100 °C. The cell solution was centrifuged for 40 min at ~10000 g under anaerobic conditions and taken into the anaerobic chamber. The supernatant was filter-sterilized using a Stericup vacuum filtration system

(Millipore), and the extract was decanted into a clean and sterile 70-ml serum bottle. Absorbance of the coenzyme extract at 280, 260, 420 and 430 nm was 2.19, 2.27, 0.55, and 0.46. The sterile extract was stored at -20 °C under anaerobic conditions.

Mutant library growth procedures

Aliquots of 400 µl and 200 µl from the mutant libraries 1 and 2 frozen stocks (chapter 3), respectively, were pooled and diluted in 10 ml of McN medium to an absorbance (600 nm) of 0.5-0.6. This suspension was defined as T0. From this dilution, $4x10^7$ viable cells were inoculated into 20 ml of McNd (positive control), McNde, McNda and McNdc media. In every case, ampicillin was added. In the first passage (T1), cells were grown to an absorbance (600 nm) of 0.5-0.6 at 37°C. For the second passage (T2), cells were transferred to fresh medium and grown to an absorbance of 0.5-0.6. Thus, each passage comprised about seven generations and a 100-fold amplification of the cell number. Two biological replicates were performed for each condition. Genomic DNA was extracted from 5 ml of each library at passage T2 using the ZR fungal/bacterial DNA miniprep (Zymo Research) and resuspended in TLE buffer (10mM Tris-HCl buffer [pH 8], 0.1mM EDTA).

For steady state conditions, 5 ml of culture were collected aerobically from the chemostat at zero dilutions (T1), after 3 dilutions (T2), 5 dilutions (T3), 8 dilutions (T4) and 10 dilutions (T5) of the chemostat volume. Cells were decanted, and the pellet was resuspended in 200 µl of dH_2O. The cell suspension was stored at -20 °C. Genomic DNA was extracted using the ZR fungal/bacterial DNA miniprep (Zymo Research) and resuspended in TLE buffer (10mM Tris-HCl buffer [pH 8], 0.1mM EDTA).

Enrichment probe and Illumina sequencing primer

The 61-nucleotide 5' dual biotinylated transposon-chromosome junction enrichment probe (5'-TGTGCAATGTAACATCAGAGATTTTGAGAAGCTTCAGGGTTGAGATGTGTATAAGAGACAG) and 26-nucleotide primer for Illumina sequencing (5'-AGGGTTGAGATGTGTATAAGAGACAG) were designed to bind to the end of the transposon through Primer 3 (43) and purified by HPLC. This design allowed the sequencing to start just after the insertion site (27).

High-throughput insertion tracking by deep sequencing

The procedure was described previously (44). For simplicity, data collected from transposon mutant libraries grown in McNd, McNde, McNda and McNdc will be referenced as Control, CoE, AA and CyS libraries, respectively.

Analysis of sequencing data

Sequence reads from the Illumina FASTQ files were mapped to the genome of *M. maripaludis* S2 using the Burrows-Wheeler Aligner (BWA) (29). Numbers of reads per gene were assigned using software developed in-house by Jan Mrázek. The first 5% of the gene length (at the 5' end) and the last 20% of the gene length (at the 3' end) were not used because it is possible that many start codons were not annotated properly or that insertions close to the end of a gene had little effect on functionality. Duplicates were averaged, and the values were used to calculate the fraction of cells in the population that have an insertion in a particular gene under each specific growth condition:

$$f = {R_g}/{\sum_{i=1}^{N} R_i}$$

where R_g is the number of reads for a specific gene and $\sum_{i=1}^{N} R_i$ is the sum of the total number of reads for every gene in the genome. Malthusian parameters (*m*) (54) for each gene were calculated as follows:

$$m = ln\left(\left(N_{T2}\, x f_{T2}\right) / \left(N_{T0}\, x\, f_{T0}\right)\right)$$

where f_{T2} and f_{T0} are the fractions of cells in the population that have an insertion in a particular gene at time point T2 and T0 respectively for a specific tested condition or the control, and N_{T0} and N_{T2} correspond to the initial and final total number of cells. The difference in fitness (DF score) was calculated by subtracting the Malthusian parameter of a gene in a specific growth condition to the Malthusian parameter of the same gene in the control. A gene with high DF score indicated that mutants with insertions in that gene display an increased fitness under the specific condition compared to mutants growing in the control medium.

For the CoE library, the DF score of genes related to coenzyme M and alanine biosynthesis (MMP0161, 0273, 0411, 1133, 1689 and 1513) were averaged and used as a reference point. These genes were selected because these organic molecules were supplemented in every tested condition and the control. For these genes, the DF scores were $\leq$ 0.36 and $\geq$ -0.36 in CoE libraries. This range was then taken to represent the stochastic variation in DF for genes whose fitness was not affected by the growth conditions. This is only a referential cut-off intended to simplify the analysis of fitness and avoid false negatives. For a more in-deep look, the number of reads per genes in each condition was also carefully analyzed (Datasets S1, S2 and S3). This second analysis proved necessary because the number of reads for many genes under both conditions was very small. In many of these cases, the DF values appeared to be very high due to stochastic variations in the small numbers of reads. In general, the growth conditions

tested in libraries AA and CyS were intended to target specific groups of genes. Some expected results were obtained, and genes expected to change fitness in a specific condition are used to qualitatively calibrate the search for hypothetical proteins that showed an increased DF score.

Results

Growth conditions of the mutants and sequence mapping

The locations of the insertions were mapped in the original library, which was formed after growth with puromycin selection for about 20 generations (T0) and after growth in the absence of puromycin selection for fourteen generations (T2) in the four different media described above. Alanine was supplemented in all media to avoid accumulations of reads in the gene MMP1511 in minimal medium as seen before (44) (appendix 1). For the growth of the mutant library in McNe, different quantities of coenzyme extract were tested in 5-ml cultures of *M. maripaludis* to determine the best conditions for mutant evaluation (Figure 4-2). The results demonstrated an enhanced growth when the coenzyme extract was supplemented to the minimal medium. The best growth was obtained using a 1-2% v/v of coenzyme extract. Growth in the presence of aryl acids has been previously tested (40, 41).

Tn-seq technology was applied only for libraries at T2 for samples growing in batch conditions and at T1 to T5 for samples growing in steady state conditions (Table 4-2 and Table 4-6). Across all libraries, about 89-95% of the sequence reads were mapped to the genome. While insertions were generally evenly distributed around the genome for the mutant library following growth in batch mode, some locations possessed a high number of reads as seen in similar previous experiments (chapter 3). For the mutant library following growth in steady state conditions, some specific genes accumulated large numbers of reads. These genes will be discussed later in the text. The number of reads was determined for every gene, and duplicated

libraries displayed a high Pearson correlation index (over 0.985), with the exception of CyS libraries, which displayed a comparatively lower correlation index of 0.9153 (Figure 4-3). The average number of reads for duplicated libraries per gene was used to calculate the Malthusian parameter (*m*) (paper). Comparing *m* of mutants growing under a specific condition to a control allowed calculation of differential fitness score (DF score) to predict the fitness of mutants with insertions in particular genes (Datasets S1, S2 and S3). However, when the number of reads for a particular gene was low, the DF score was often much higher or lower than expected, probably due to stochastic variations (Figure 4-4). Therefore, to determine an actual change of fitness, not only the DF score of the genes, but the number of reads was also examined.

Relative fitness of mutants growing in the presence of coenzyme extract

In CoE library, 320 genes display a DF score over 0.36 (Dataset S1). Mutants with insertions in those particular genes showed a higher number of reads in medium supplemented with coenzyme extract compared to the absence of supplement, which indicated that these mutants might have an increased fitness. The use of a *M. maripaludis* extract as supplement mimics a rich medium with vitamins, nucleotides, amino acids and cofactors. Methanococci are facultative lithotrophs that only poorly assimilate organic compounds and presumably rely upon their biosynthetic capacity (53). However, uptake of some organic molecules such as alanine, aryl acids and coenzyme M, among others, has been demonstrated (35, 40). If the methanococci were able to assimilate the unusual coenzymes involved in methanogenesis, mutations in their biosynthetic pathways might have an increased fitness in the coenzyme supplemented medium. Genes for specific biosynthetic pathways are analyzed as follow:

Unusual coenzymes. Few genes for the biosynthesis of the unusual coenzymes found in methanogens are known, and a smaller number have been studied in detail. From the four known genes involved in coenzyme B biosynthesis, only one displayed an increased DF score: MMP1480, which encodes one of the subunits of the homoaconitase involved in the 2-oxosuberate pathway. However, the number of reads in both conditions was close to zero. Two of the three known genes for the biosynthesis of tetrahydromethanopterin (H_4MPT) displayed a DF score close to zero (MMP0034 and 0230); the third gene MMP0279 displayed a negative DF score. Two genes for the biosynthesis of deazaflavin coenzyme F_{420} (MMP0915 and 0937) displayed an increased DF score, but the number of reads was close to zero in both conditions. Three other genes (MMP0056, 0057, 0404) displayed a DF score close to zero. The first two encode homologs of the same gene (*cofH*). The final three genes (MMP0170, 0876 and 1487) displayed a negative DF score, but MMP0876 and 1487 had low numbers of reads in both conditions. The only gene known to be involved in the biosynthesis of methanofuran, MMP0131, displayed a slightly increased DF score, but a closer look to the number of reads revealed that in both conditions the number of reads was low Thus, for these coenzymes, there was no convincing evidence for an increased fitness for the biosynthetic genes. Presumably, cells were not able to assimilate these coenzymes from the medium. In summary, the fitness of the methanogen coenzyme genes did not increase during growth with the coenzyme extract.

***Vitamin biosynthesis*.** Riboflavin biosynthesis in *M. maripaludis* is accomplished in three steps from ribulose 5-phosphate. Mutants with insertions in the three genes involved in those steps (MMP0050, 0180 and 0183) displayed an increased fitness in the coenzyme supplemented medium, suggesting that *M. maripaludis* can uptake riboflavin (Figure 4-5A). Genes for the biosynthesis of cobamide can be divided into two groups, genes for the uptake of cobalt and

genes for the actual synthesis of cobamide. Evidence for the role of six genes in a cobalt transport system (cbiAMNOQS) has been previously demonstrated in *M. maripaludis* C5 through an *in silico* reconstruction (2). Homologs for the proteins CbiA, CbiN and CbiO have been found in *M. maripaludis* S2 (MMP1477, 0888 and 1484, respectively). In addition, two homologs for the proteins CbiM (MMP0889 and 1481) and CbiQ (MMP0886 and 1483) are present. All of these genes displayed an increased DF score (Figure 4-5A) with the exception of MMP1477, which had a low number of reads under both conditions. A homolog for the last protein of the transport complex (cbiS) has not been found in *M. maripaludis* S2. Sixteen other genes may be involved in the biosynthesis of cobamide, and only one, MMP1591, displayed an increased DF score. A closer examination of the other genes revealed that the number of reads was low under both the coenzyme supplemented and control conditions with the exceptions of MMP0493 and MMP1227. These observations suggest that *M. maripaludis* synthesized its own cobamide, but the intake of cobalt by the transport system cbiAMNOQS was not necessary when the medium was supplemented with coenzymes. Possibly, the coenzyme extract possess a high concentration of cobalt, which have abrogated the need for a high affinity uptake system. Alternatively, cobalt may be toxic. Indeed, excess of cobalt in *E. coli* has a negative effect on iron-sulfur protein biosynthesis (42). To avoid over accumulation of this metal inside the cells, mutations in the transporter genes may slow the rate of cobalt intake. Of the six known genes for biotin biosynthesis, only one displayed an increased DF score (MMP0126) of 0.7. There are three genes which encode enzymes involved in the molybdenum cofactor biosynthesis which displayed an increased DF score (MMP0571, 1066 and 1235). The molybdenum cofactor is a component of methanofuran dehydrogenase, an abundant enzyme in methanogenesis. These

results suggest that this cofactor might be taken from the medium, and these genes could be involved in its biosynthesis.

***Nucleotide biosynthesis*.** Mutants with insertions in genes related to purine and pyrimidine biosynthesis did not display an increased DF score with the exception of four genes involved in purine biosynthesis (MMP0179, 0540, 1310 and 1146) and three genes involved in pyrimidine biosynthesis (MMP0602, 1589 and 1659) (Figure 4-5B). However, a closer examination revealed that only genes MMP0179 and 0540 displayed an increase of the number of reads the presence of the extract. These results suggest that *M. maripaludis* still biosynthesizes its own nucleotides in the presence of coenzyme extract. In fact, previous studies demonstrated that methanococci readily assimilate nucleobases but they do not uptake the nucleosides and nucleotides likely to be abundant in the cell extract (1).

***Amino acids biosynthesis*.** Previous studies had demonstrated that *M. maripaludis* poorly assimilates amino acids (53). Indeed, the present results demonstrate that, in most of the 69 known genes involved in amino acid biosynthesis, the number of reads was low under both conditions and the DF was not reliable evidence for a change in fitness. The low number of reads could also be explained by low availability of amino acids in the coenzyme extract. Only two genes displayed an actual increased in the number of reads when the coenzyme extract was supplemented. The gene MMP0135 (*thrC*) encodes threonine synthase, which is involved in the last step of the biosynthetic pathway of threonine. This gene had a DF score of 2.6 and an increase in the average number of reads from zero to 12.5 when the coenzyme extract was supplemented. However, the two duplicates of the library in the presence of the coenzyme extract exhibited a number of reads of 0 and 25, indicating a high variability among duplicates. If the observed increase in reads when the coenzyme extract is supplemented is true, *M.*

maripaludis may be absorbing threonine from the supplemented medium, which makes the gene MMP0135 not necessary. Mutants with insertions in the rest of the genes involved in the biosynthetic pathway of threonine (MMP0295, 1017, 1391 and 1702) did not display a difference in fitness among the tested conditions, which suggested that the result with MMP0135 was spurious. Similarly, previous studies demonstrated that methanococci do not take up threonine well (53). The second gene which displayed an increased fitness was MMP0923 (*dapB*), which encodes dihydrodipicolinate reductase and is involved in lysine biosynthesis. Presumably, an alternative enzyme may replace DapB in the pathway. The rest of the known genes involved in the biosynthetic pathway for lysine (MMP0576, 0917, 1017, 1200, 1391 and 1527) accumulated a low number of reads in both conditions.

Hypothetical proteins. Among the 320 genes with a DF score over 0.36, 102 genes were hypothetical proteins of unknown function. From these 102 genes, only 35 of them exhibit an increase in the average number of reads when the coenzyme extract was supplemented that at least duplicates the number of reads in the control (Table 4-3). These genes displayed a higher fitness when mutant libraries were grown in medium supplemented with coenzyme extract and may be playing important roles in different biosynthetic pathways. A complete list of these genes is given in table 4-3, and some interesting examples related to biosynthesis were further analyzed. Over 70% of the protein sequences encoded by the hypothetical gene MMP0567 corresponds to a conserved domain for 4-hydroxybenzoyl-CoA thioesterase,. In addition, a homology search revealed that this protein possesses 71% similarity with a thioesterase from *Methanobacterium* sp. SWAN-1, which seems conserved among methanogens. The hypothetical gene MMP0564 possesses an ATP-grasp domain. Enzymes with this domain are commonly found in several metabolic pathways including *de novo* purine biosynthesis, gluconeogenesis,

fatty acid synthesis and biotin carboxylation (9). The hypothetical gene MMP0543 encodes a probable chorismate lyase. Chorismate is an important intermediary and a precursor for many organic molecules. Indeed, in *E. coli* this enzyme is involved in the biosynthesis of 4HO-benzoate, an intermediate in ubiquinone biosynthesis (36). The hypothetical gene MMP1606 encodes a protein with two domains. The first domain possess the capacity to bind to flavin mononucleotide (FMN), the second domain possesses a bifunctional activity for phosphopantothenoylcysteine decarboxylase and phosphopantothenate-cysteine ligase activity. These observations suggest that this enzyme could be involved in coenzyme A and/or the vitamin pantothenate biosynthesis. Finally, hypothetical genes MMP1243 and 1234 encode proteins with domains that bind and utilize dinucleotides (NAD or FAD) and may be involved in molybdenum cofactor and/or thiamin biosynthesis.

Relative fitness of mutants growing in the presence of aryl acids

Two different biosynthetic pathways for aromatic amino acids (AroAA) have been described in *M. maripaludis* (40, 41). In the *de novo* pathway aromatic amino acids are biosynthesized from 6-deoxy-6-ketofructose-1-phosphate (DKFP). The second pathway utilizes aryl acids (indole acetate, phenylacetate, and hydroxyphenylacetate) as precursors for AroAA. The DF score for the 21 known genes involved in these two biosynthetic pathways when the mutant library was grown in the presence of aryl acids are presented in Table 4-4. For a complete list of DF score and reads under this condition, refer to dataset S2. Mutants with insertions in the genes *aroA'* (MMP0686) and *aroB'* (MMP0006) involved in the common fraction of the *de novo* AroAA pathway did not displayed a change in fitness (Figure 4-6). The gene *aroA'* had only a low number of reads, indicating that the DF was not reliable. In contrast, *aroB'* accumulated a high number of reads in both conditions. Previous studies have demonstrated that this portion of the

de novo AroAA biosynthetic pathway is essential even in the presence of aryl acids because the intermediate 3-dehydroquinic acid (DHQ) is a precursor for the biosynthesis of p-aminobenzoic acid (PABA), which is also a precursor of methanofuran (40). However, mutants for the gene *aroB'* were not affected by the absence or aryl acids or PABA, suggesting that an alternative gene might complement its function (40). Mutants with insertions in the genes *aroD* (MMP1394), MMP0320, and *aroC* (MMP1333), involved in the common fraction of the *de novo* pathway, displayed close to zero DF scores and a low number of reads in both conditions. However, mutants with insertions in the genes *aroE* (MMP0936) and *aroA* (MMP1205) had an increased fitness. These results indicate that some genes of the common pathway for AroAA were essential but others were dispensable in the presence of aryl acids. Genes with low number of reads in both conditions may be playing roles in alternative pathways or their products may be substrates for different processes. The low number of reads for the gene *aroD* in the presence of Aryl acids contradicts previous results which demonstrated that a mutant of the gene *aroD* was successfully constructed and aryl acids fulfilled the requirements for AroAAs (41),. To understand this difference, further experimental analyses areneeded.

After chorismate, the AroAA pathway split into two branches, one of which leads to tryptophan and one to prephenate, the precursor of phenylalanine and tyrosine. Seven genes are involved in the biosynthesis of tryptophan (MMP1002-1008), and mutants with insertion in these genes displayed an increased fitness when aryl acids were supplemented into the medium. However, the gene *trpC* (MMP1008) showed only a slightly increased number of reads in the tested condition compare to the control. Similarly, from the three known genes involved in phenylalanine and tyrosine biosynthesis, two of them, *pheA* (MMP1528) and *tyrA* (MMP1514) had an increased DF score. Mutants with insertions in the gene *aroQ* (MMP0578), which encode

for chorismate mutase and a common step for both of these biosynthetic pathways, did not change fitness. However, the number of reads was low under both conditions, suggesting that the DF score was not reliable. The gene involved in the last step of the biosynthesis of phenylalanine and tyrosine, which encodes an aminotransferase, is not known. Gene *hisC* (MMP1216) and gene MPP1072 have been hypothesized as the probable aromatic aminotransferase, but mutants with insertions in both genes did not display a change in fitness. Interestingly, gene MMP1101, which is described as acetylornithine aminotransferase, displayed a DF score of 2.8, and an increase in the average number of reads from zero to 16 in the presence of aryl acids. These observations suggest that this enzyme may be the aminotransferase for tyrosine and phenylalanine.

Finally, four genes directly involved in the transformation of aryl acids to AroAAs are present in *M. maripaludis* genome. Genes *iorB1A1* (MMP0315 and 0316) and *iorA2B2* (MMP0713 and 0714), which encode two homologs of the indolepyruvate oxidoreductase, did not have an increase in the DF score when aryl acids were present in the medium. This result was expected because the de novo AroAA biosynthetic pathway complements their function.

Hypothetical proteins. Five genes coding for hypothetical proteins displayed an increase in the DF score. However, the number of reads in the two cases was very low, and the DF scores were not reliable (MMP0465 and 784). The gene MMP0604 encodes a kinase protein and possesses a RIO1 like domain. This domain is present in many kinases that are involved in different cellular processes. This gene is of particular interest, because many kinases function as intermediaries in signal transduction and a great deal of regulation is involved in the AroAAs biosynthetic pathway (41). Interestingly, the gene MMP0567, which encodes a thioesterase, displayed an increased DF score in the presence of aryl acids. This same gene had an increased

DF score when coenzyme extract was supplemented into the medium (see above). No increased fitness was observed for aromatic amino acids biosynthetic genes in the presence of the coenzyme extract, indicating that the increased DF score for the gene MMP0567 when the coenzyme extract is present is not related to the biosynthesis of aroAA. These results suggest that this gene could be involved in more than one process. Finally, the gene MMP0412 encodes an unknown protein which presents a TRAM superfamily domain with a nucleic acid binding capacity, which is associated to tRNA modification.

Relative fitness of mutants growing in the presence of cysteine

Cysteine biosynthesis is a major process for the incorporation of inorganic sulfur into organic molecules in bacteria and eukaryotes. At least two cysteine biosynthetic pathways are known in archaea (32). Both pathways use O-phosphoserine (Sep) as substrate, but only one pathway has been identified in methanogenic archaea. In the pathway found in methanococci, Sep is aminoacylated to $tRNA_{cys}$ by *O*-phosphoseryl-tRNA synthetase (SepRS). In the second step Sep-$tRNA_{cys}$ is transformed to Cys-$tRNA_{cys}$ by the enzyme Sep-tRNA:Cys-tRNA synthase (SepCysS) with an unknown sulfur donor (46). The gene MMP0688 encodes the enzyme SepRS, and mutants with insertions in this gene displayed an increase in fitness when cysteine was supplemented into the medium, suggesting that this gene was not required for growth in the presence of cysteine (Table 4-5). Gene MMP1240 encodes the second enzyme involved in this pathway, SepCysS, and mutants with insertions in this gene did not show an increased fitness. However, the number of reads was low, suggesting that the DF score was not reliable. Additionally, it has not been possible to construct mutations for this gene, presumably because they result in the lethal production of Sep- $tRNA_{cys}$ (Y. Liu, unpublished data). A third gene (MMP1217) with unknown function has also been associated to this pathway. Mutants with

insertions in this gene did not show an increased fitness, but the number of reads was low. Construction of mutants for this gene has also not been successful (Y. Liu, unpublished data). Finally, the gene MMP1060, which encodes cysteinyl-tRNA synthetase, did not show an increase in the DF score, and the number of reads was high under both conditions, as expected. For a complete list of DF score and reads under this condition, refer to dataset S3.

An interesting case was observed for the biosynthesis of selenocysteine (Sec). Selenocysteine is biosynthesized in three steps: serylation of tRNASec by seryl-tRNA synthetase (SerRS), phosphorylation of Ser-$tRNA_{Sec}$ by O-phosphoseryl-$tRNA_{Sec}$ kinase (PSTK), and conversion of O-phosphoseryl-$tRNA_{Sec}$ (Sep-$tRNA_{Sec}$) to Sec-$tRNA_{Sec}$ by Sep-tRNA:Sec-tRNA synthase (SepSecS) (49). Homologs for these three genes can be found in *M. maripaludis*, MMP0879 encodes SerRS, MMP1490 possibly encodes PSTK, and MMP0595 encodes SepSecS (56). The number of reads in the gene MMP1490 was low, so the DF was not reliable. Genes MMP0879 and 0595 displayed increased DF scores and a high average number of reads when cysteine was present (Table 4-5). However, the correlation index for the number of reads of the CyS library replicates was comparatively low, and the number of reads for these two genes in CyS replicates was 27 and zero, and 14 and one, respectively. Therefore, the results should be viewed with caution. If this increase in fitness is real, the result suggests a previously unknown relation for the biosynthesis of cysteine and selenocysteine. Alternatively, selenocysteine is not an uncommon contaminant of commercial cysteine.

Mutants growing in steady state conditions

The number of reads per gene were studied after growth in steady state conditions under hydrogen limitation/phosphate excess (H-/P+) or hydrogen excess/phosphate limitation (H+/P-). The results were unexpected. The total number of reads for the mutant library under H-/P+

conditions increased over time (Table 4-6), and mutants of one specific gene (MMP1102) dominated the population in every time point. Reads for this mutant represented 67% of the population at T1and immediately increased to more than 99.5% in T2. This proportion was remained constant over the next three time points. After ten dilutions (T5), only 167 additional mutants were still present in the population, but apart from the MMP1102 mutants they only represented 0.32% of the population. The ten most abundant mutants in T5 are summarized in table 4-7. For a complete list of reads mapped to the genome, refer to dataset S4. In contrast, mutants growing under H+/P- conditions displayed an increase in the total number of reads up to T3. In T4 and T5, the number of reads decreased (table 4-6). In T1 the mutants in MMP0233 dominated the population with over 70% of the reads. After 3 dilutions (T2) the population dynamics changed, and the mutants in gene MMP1102 dominated the population with 98.2% of the reads. By T4 these mutants represented over 99.5% of the reads. After ten dilutions (T5), the abundance of mutants in MMP1102 had declined. At T5, only 211 genes showed at least one read, but apart from the mutant MMP1102 they represented 1.15% of the population. The ten most abundant mutants in T5 are summarized in table 4-8. For a complete list of reads mapped to the genome, refer to dataset S4.

Under hydrogen and phosphate limitation, mutants for the same gene (MMP1102) dominated the population. The presence of similar initial mutants in both steady-state conditions is expected because in both experiments the chemostat was started with two different aliquots of the same mutant library. Moreover, different outcomes were expected in each condition, reflecting the different selective pressure of hydrogen or phosphate limitation. A closer analysis of the reads in the gene MMP1102 in T1 and T5 under both tested conditions revealed that the reads are shared among 15 different sites in this gene, but one of them represents around 93-94

% of the total. At T0 reads for the gene MMP1102 represent only 0.097% of the population, and reads in this specific spot represents around 2% of the total reads for the gene, indicating a strong selection for one specific transposon mutant for the gene MMP1102 under both tested conditions. These observations suggest that the gene MMP1102 may be involved in a process related to the limitation of hydrogen and phosphate, but this seems unlikely because of the possible function of this gene (see below for a complete analysis of gene MMP1102). Alternatively, the selection may be related to the physiological steady-state condition or to the chemostat itself. However, at T1 mutants were still growing in batch mode, suggesting that the selection for mutants with insertions in this gene started earlier and was not a direct consequence of the chemostat.

The gene MMP1102 encodes for a protein with a phospholipase D (PLD) catalytic domain of unknown function in archaea. Enzymes which possess this bifunctional domain are commonly located in the membrane and catalyze the hydrolysis of the distal phosphodiester bond of phospholipids to generate phosphatidic acid (PA) and transphosphatidylation with primary or secondary alcohols to generate different phospholipids (55). The PLD domain superfamily is composed of a large and diverse group of proteins such as plant, mammalian and bacterial phospholipase D (PLD) and bacterial cardiolipin (CL) synthases, among others. Interestingly, the phospholipase D catalytic domain has been proposed to be present in the last common ancestor of archaea, bacteria and eukaryotes (39). In eukaryotes, PLD enzymes are proposed to mediate intracellular signal transduction (37) and are involved in sporulation in yeasts (8). In some bacteria, they are virulence factors (34). CL synthase synthesizes a diphosphatidylglycerol lipid (cardiopilin), which in some prokaryotes is accumulated during stationary growth or environmental stress (47). Structural analogues of this lipid have been

isolated from halophilic archaea (28). There is no clear evidence to support a role for the gene MMP1102 in archaea or to understand why mutants of this gene displayed an increased fitness under these conditions in *M. maripaludis*. Possibly, the gene is playing a role in the structural organization of biological membranes. Alternatively, if the gene MMP1102 encodes a CL synthase, it may be involved in stress response. However, in any case the large number of reads in this gene implied that is not an essential gene under the tested conditions.

Discussion

Tn-seq is a powerful technology that combines transposon mutagenesis and next generation sequencing to provide genome wide assessments of a microorganism's genetic requirements for growth under different conditions. However, experiments using Tn-seq technology should be carefully designed to obtain results that generate laboratory testable hypotheses. The main objective of this work was to use Tn-seq technology as a hypothesis generation system towards the identification of genes that may be involved in specific biosynthetic and metabolic pathways in *M. maripaludis*. Thus, different specific conditions for the growth of *M. maripaludis* were assessed and compared to a common control to study fitness and mutant adaptation. The functions of many of the genes of methanococci are already known, and they served as internal controls for the experiment. However, the functions of over 800 genes in the genome of *M. maripaludis* are unknown. These genes may represent the missing steps of the targeted biosynthetic pathways, and the experiments were designed to generate hypotheses about their nature. In addition, genes with dual functions or genes which are miss-annotated can be found.

In this study three specific conditions were analyzed under batch growth. The study of the mutant library following growth in coenzyme extract supplemented medium was primarily intended to identify genes involved in the biosynthetic pathways for the unusual coenzymes of

methanogens. However, genes already known for these biosynthetic pathways did not display changes in fitness. Possibly, *M. maripaludis* cannot absorb these coenzymes due to their size or steric conformation. In addition, H_4MPT, F_{430} and F_{420} are negatively charge molecules which could interfere with their uptake. Alternatively, the integrity of these coenzymes may be affected by the preparation of the extract and/or *M. maripaludis* may not possess transporters for these coenzymes. The coenzyme extract was rich on many other organic compounds that may be used by the cells, but the results only identified a small number of genes with an increased fitness, principally those involved in biosynthesis of a few vitamins. This observation supports the idea that *M. maripaludis* does not assimilate many organic molecules and relies principally on its own biosynthetic capabilities. These results imply that the hypothetical or unknown proteins which displayed increased fitness are probably not involved in the biosynthesis of the unusual coenzymes but in the biosynthesis of other organic compounds, such as vitamins, which can be further studied by genetic and biochemical methods. For a clearer outcome related to the biosynthesis of the unusual coenzymes, a targeted experiment where the mutant libraries are fed with the isolated coenzymes should be designed. However, first the capabilities of *M. maripaludis* to uptake the coenzymes should be tested.

The study of the mutant library following growth in aryl acids and cysteine were able to identify genes with an increased fitness that were previously known to be involved in their biosynthesis. The design of these experiments was targeted to a small group of genes, and the mutant library was grown under controlled conditions for which a specific outcome was expected. Indeed some previously unknown genes were identified as possible players in aromatic amino biosynthesis following growth of the mutant libraries with aryl acids. The next step is

corroboration of the involvement of the hypothetical proteins in these pathways by genetic and biochemical studies.

Finally, two specific conditions were analyzed under steady-state growth. Surprisingly, in both experiments mutants with insertions in the same gene, MMP1102, dominated the populations, suggesting that the absence of this gene is beneficial for the cell. The outcome of these two experiments was completely unexpected and difficult to explain from a biological point of view. Indeed, possibly the results were not a direct consequence of a biological event which was triggered under the hydrogen or phosphate limitation or the steady-state conditions but due to a problem in the processing of the samples. Actually, the samples were shipped to another laboratory to grow the mutant library in chemostat under anaerobic conditions. For example, any problems in the "cold chain" of proper refrigeration of the samples could affect the outcome.

Although these experiments were not conclusive, the Tn-seq technology in this study yielded some interesting results which need to be further tested. Because of the speed and resolution of Tn-seq, this technique has potential to efficiently conduct fitness analyses in diverse contexts when the conditions are carefully designed. Further improvement and application of this technique in *M. maripaludis* and other archaea is expected to generate new understanding of the genetic requirements for biosynthesis and metabolism, and could be a powerful tool in the assignment of function for unknown and hypothetical genes.

References

1. **Bowen, T. L., and W. B. Whitman.** 1987. Incorporation of Exogenous Purines and Pyrimidines by *Methanococcus voltae* and Isolation of Analog-Resistant Mutants. Appl Environ Microbiol **53:**1822-6.

2. **Chellapandi, P.** 2011. In silico description of cobalt and nickel assimilation systems in the genomes of methanogens. Syst Synth Biol **5:**105-14.

3. **Costa, K. C., P. M. Wong, T. Wang, T. J. Lie, J. A. Dodsworth, I. Swanson, J. A. Burn, M. Hackett, and J. A. Leigh.** 2010. Protein complexing in a methanogen suggests electron bifurcation and electron delivery from formate to heterodisulfide reductase. Proc Natl Acad Sci U S A **107:**11050-5.

4. **de Crecy-Lagard, V., G. Phillips, L. L. Grochowski, B. El Yacoubi, F. Jenney, M. W. Adams, A. G. Murzin, and R. H. White.** 2012. Comparative genomics guided discovery of two missing archaeal enzyme families involved in the biosynthesis of the pterin moiety of tetrahydromethanopterin and tetrahydrofolate. ACS Chem Biol **7:**1807-16.

5. **Deppenmeier, U.** 2004. The membrane-bound electron transport system of *Methanosarcina* species. J Bioenerg Biomembr **36:**55-64.

6. **Drevland, R. M., Y. Jia, D. R. Palmer, and D. E. Graham.** 2008. Methanogen homoaconitase catalyzes both hydrolyase reactions in coenzyme B biosynthesis. J Biol Chem **283:**28888-96.

7. **Eisenreich, W., and A. Bacher.** 1992. Biosynthesis of methanofuran in *Methanobacterium thermoautotrophicum*. J Biol Chem **267:**17574-80.

8. **Ella, K. M., J. W. Dolan, and K. E. Meier.** 1995. Characterization of a regulated form of phospholipase D in the yeast *Saccharomyces cerevisiae*. Biochem J **307 (Pt 3):**799-805.

9. **Fawaz, M. V., M. E. Topper, and S. M. Firestine.** 2011. The ATP-grasp enzymes. Bioorg Chem **39:**185-91.

10. **Gawronski, J. D., S. M. Wong, G. Giannoukos, D. V. Ward, and B. J. Akerley.** 2009. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. Proc Natl Acad Sci U S A **106:**16422-7.

11. **Graham, D. E., M. Graupner, H. Xu, and R. H. White.** 2001. Identification of coenzyme M biosynthetic 2-phosphosulfolactate phosphatase. A member of a new class of Mg(2+)-dependent acid phosphatases. Eur J Biochem **268:**5176-88.

12. **Graham, D. E., and R. H. White.** 2002. Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics. Nat Prod Rep **19:**133-47.

13. **Graham, D. E., H. Xu, and R. H. White.** 2002. Identification of coenzyme M biosynthetic phosphosulfolactate synthase: a new family of sulfonate-biosynthesizing enzymes. J Biol Chem **277:**13421-9.

14. **Graham, D. E., H. Xu, and R. H. White.** 2003. Identification of the 7,8-didemethyl-8-hydroxy-5-deazariboflavin synthase required for coenzyme F(420) biosynthesis. Arch Microbiol **180:**455-64.

15. **Graupner, M., H. Xu, and R. H. White.** 2002. Characterization of the 2-phospho-L-lactate transferase enzyme involved in coenzyme F(420) biosynthesis in *Methanococcus jannaschii*. Biochemistry **41:**3754-61.

16. **Graupner, M., H. Xu, and R. H. White.** 2000. Identification of an archaeal 2-hydroxy acid dehydrogenase catalyzing reactions involved in coenzyme biosynthesis in methanoarchaea. J Bacteriol **182:**3688-92.

17. **Graupner, M., H. Xu, and R. H. White.** 2000. Identification of the gene encoding sulfopyruvate decarboxylase, an enzyme involved in biosynthesis of coenzyme M. J Bacteriol **182:**4862-7.

18. **Grochowski, L. L., H. Xu, K. Leung, and R. H. White.** 2007. Characterization of an Fe(2+)-dependent archaeal-specific GTP cyclohydrolase, MptA, from *Methanocaldococcus jannaschii*. Biochemistry **46:**6658-67.

19. **Grochowski, L. L., H. Xu, and R. H. White.** 2008. Identification and characterization of the 2-phospho-L-lactate guanylyltransferase involved in coenzyme F420 biosynthesis. Biochemistry **47:**3033-7.

20. **Grochowski, L. L., H. Xu, and R. H. White.** 2006. Identification of lactaldehyde dehydrogenase in *Methanocaldococcus jannaschii* and its involvement in production of lactate for F420 biosynthesis. J Bacteriol **188:**2836-44.

21. **Gunsalus, R. P., and R. S. Wolfe.** 1980. Methyl coenzyme M reductase from *Methanobacterium thermoautotrophicum*. Resolution and properties of the components. J Biol Chem **255:**1891-5.

22. **Howell, D. M., M. Graupner, H. Xu, and R. H. White.** 2000. Identification of enzymes homologous to isocitrate dehydrogenase that are involved in coenzyme B and leucine biosynthesis in methanoarchaea. J Bacteriol **182:**5013-6.

23. **Howell, D. M., K. Harich, H. Xu, and R. H. White.** 1998. Alpha-keto acid chain elongation reactions involved in the biosynthesis of coenzyme B (7-mercaptoheptanoyl threonine phosphate) in methanogenic Archaea. Biochemistry **37:**10108-17.

24. **Johnson, E. F., and B. Mukhopadhyay.** 2005. A new type of sulfite reductase, a novel coenzyme F420-dependent enzyme, from the methanarchaeon *Methanocaldococcus jannaschii*. J Biol Chem **280:**38776-86.

25. **Kaster, A. K., M. Goenrich, H. Seedorf, H. Liesegang, A. Wollherr, G. Gottschalk, and R. K. Thauer.** 2011. More than 200 genes required for methane formation from H(2) and CO(2) and energy conservation are present in *Methanothermobacter marburgensis* and *Methanothermobacter thermautotrophicus*. Archaea **2011:**973848.

26. **Kezmarsky, N. D., H. Xu, D. E. Graham, and R. H. White.** 2005. Identification and characterization of a L-tyrosine decarboxylase in *Methanocaldococcus jannaschii*. Biochim Biophys Acta **1722:**175-82.

27. **Langridge, G. C., M. D. Phan, D. J. Turner, T. T. Perkins, L. Parts, J. Haase, I. Charles, D. J. Maskell, S. E. Peters, G. Dougan, J. Wain, J. Parkhill, and A. K. Turner.** 2009. Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. Genome Res **19:**2308-16.

28. **Lattanzio, V. M., A. Corcelli, G. Mascolo, and A. Oren.** 2002. Presence of two novel cardiolipins in the halophilic archaeal community in the crystallizer brines from the salterns of Margherita di Savoia (Italy) and Eilat (Israel). Extremophiles **6:**437-44.

29. **Li, H., and R. Durbin.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:**1754-60.

30. **Li, H., M. Graupner, H. Xu, and R. H. White.** 2003. CofE catalyzes the addition of two glutamates to F420-0 in F420 coenzyme biosynthesis in *Methanococcus jannaschii*. Biochemistry **42:**9771-8.

31. **Li, H., H. Xu, D. E. Graham, and R. H. White.** 2003. Glutathione synthetase homologs encode alpha-L-glutamate ligases for methanogenic coenzyme F420 and tetrahydrosarcinapterin biosyntheses. Proc Natl Acad Sci U S A **100:**9785-90.

32. **Liu, Y., L. L. Beer, and W. B. Whitman.** 2012. Sulfur metabolism in archaea reveals novel processes. Environ Microbiol **14:**2632-44.

33. **Mashhadi, Z., H. Xu, and R. H. White.** 2009. An Fe2+-dependent cyclic phosphodiesterase catalyzes the hydrolysis of 7,8-dihydro-D-neopterin 2',3'-cyclic phosphate in methanopterin biosynthesis. Biochemistry **48:**9384-92.

34. **McNamara, P. J., W. A. Cuevas, and J. G. Songer.** 1995. Toxic phospholipases D of Corynebacterium pseudotuberculosis, C. ulcerans and A*rcanobacterium haemolyticum*: cloning and sequence homology. Gene **156:**113-8.

35. **Micheletti, P. A., K. A. Sment, and J. Konisky.** 1991. Isolation of a coenzyme M-auxotrophic mutant and transformation by electroporation in *Methanococcus voltae*. J Bacteriol **173:**3414-8.

36. **Nichols, B. P., and J. M. Green.** 1992. Cloning and sequencing of *Escherichia coli* ubiC and purification of chorismate lyase. J Bacteriol **174:**5309-16.

37. **Nishizuka, Y., and S. Nakamura.** 1995. Lipid mediators and protein kinase C for intracellular signalling. Clin Exp Pharmacol Physiol Suppl **22:**S202-3.

38. **Pfaltz, A., A. Kobelt, R. Huster, and R. K. Thauer.** 1987. Biosynthesis of coenzyme F430 in methanogenic bacteria. Identification of 15,17(3)-seco-F430-17(3)-acid as an intermediate. Eur J Biochem **170:**459-67.

39. **Ponting, C. P., L. Aravind, J. Schultz, P. Bork, and E. V. Koonin.** 1999. Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. J Mol Biol **289:**729-45.

40. **Porat, I., M. Sieprawska-Lupa, Q. Teng, F. J. Bohanon, R. H. White, and W. B. Whitman.** 2006. Biochemical and genetic characterization of an early step in a novel pathway for the biosynthesis of aromatic amino acids and p-aminobenzoic acid in the archaeon *Methanococcus maripaludis*. Mol Microbiol **62:**1117-31.

41. **Porat, I., B. W. Waters, Q. Teng, and W. B. Whitman.** 2004. Two biosynthetic pathways for aromatic amino acids in the archaeon *Methanococcus maripaludis.* J Bacteriol **186:**4940-50.

42. **Ranquet, C., S. Ollagnier-de-Choudens, L. Loiseau, F. Barras, and M. Fontecave.** 2007. Cobalt stress in *Escherichia coli.* The effect on the iron-sulfur proteins. J Biol Chem **282:**30442-51.

43. **Rozen, S., and H. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol **132:**365-86.

44. **Sarmiento, F., Mrázek, J and Whitman, WB.** 2013. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis.* Proc Natl Acad Sci U S A. doi: 10.1073/pnas.1220225110.

45. **Sarmiento, F. B., J. A. Leigh, and W. B. Whitman.** 2011. Genetic systems for hydrogenotrophic methanogens. Methods Enzymol **494:**43-73.

46. **Sauerwald, A., W. Zhu, T. A. Major, H. Roy, S. Palioura, D. Jahn, W. B. Whitman, J. R. Yates, 3rd, M. Ibba, and D. Soll.** 2005. RNA-dependent cysteine biosynthesis in archaea. Science **307:**1969-72.

47. **Schlame, M.** 2008. Cardiolipin synthesis for the assembly of bacterial and mitochondrial membranes. J Lipid Res **49:**1607-20.

48. **Seedorf, H., A. Dreisbach, R. Hedderich, S. Shima, and R. K. Thauer.** 2004. F420H2 oxidase (FprA) from *Methanobrevibacter arboriphilus*, a coenzyme F420-dependent enzyme involved in O2 detoxification. Arch Microbiol **182:**126-37.

49. **Sherrer, R. L., J. M. Ho, and D. Soll.** 2008. Divergence of selenocysteine tRNA recognition by archaeal and eukaryotic O-phosphoseryl-tRNASec kinase. Nucleic Acids Res **36:**1871-80.

50. **White, R. H.** 1988. Biosynthesis of the 2-(aminomethyl)-4-(hydroxymethyl)furan subunit of methanofuran. Biochemistry **27:**4415-20.

51. **White, R. H.** 1986. Biosynthesis of the 7-methylated pterin of methanopterin. J Bacteriol **165:**215-8.

52. **White, R. H.** 2001. Biosynthesis of the methanogenic cofactors. Vitam Horm **61:**299-337.

53. **Whitman, W. B., S. Sohn, S. Kuk, and R. Xing.** 1987. Role of Amino Acids and Vitamins in Nutrition of Mesophilic Methanococcus spp. Appl Environ Microbiol **53:**2373-8.

54. **Woods, R. J., J. E. Barrick, T. F. Cooper, U. Shrestha, M. R. Kauth, and R. E. Lenski.** 2011. Second-order selection for evolvability in a large *Escherichia coli* population. Science **331:**1433-6.

55. **Yang, H., and M. F. Roberts.** 2002. Cloning, overexpression, and characterization of a bacterial Ca2+-dependent phospholipase D. Protein Sci **11:**2958-68.

56. **Yuan, J., S. Palioura, J. C. Salazar, D. Su, P. O'Donoghue, M. J. Hohn, A. M. Cardoso, W. B. Whitman, and D. Soll.** 2006. RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. Proc Natl Acad Sci U S A **103:**18923-7.

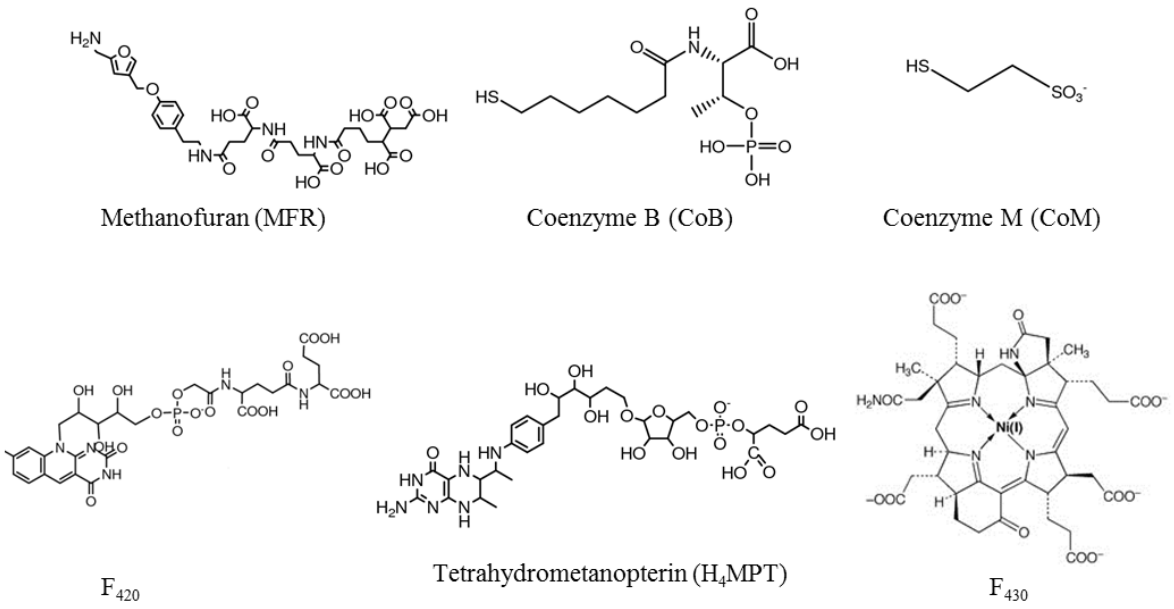

Figure 4-1. Structure of the coenzymes involved in methanogenesis pathway.

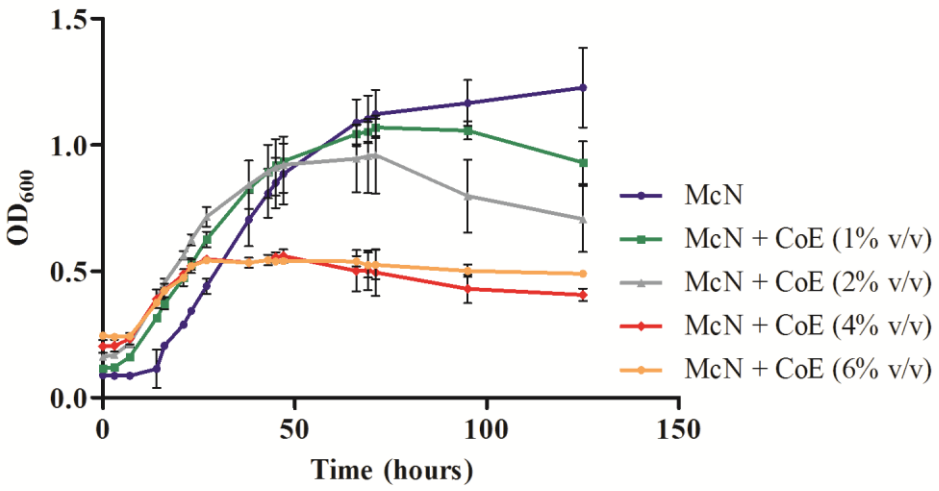

Figure 4-2. Growth of *M. maripaludis* S2 in the presence of different concentrations of coenzyme extract. The error bars indicate the SD of four independent replicates.

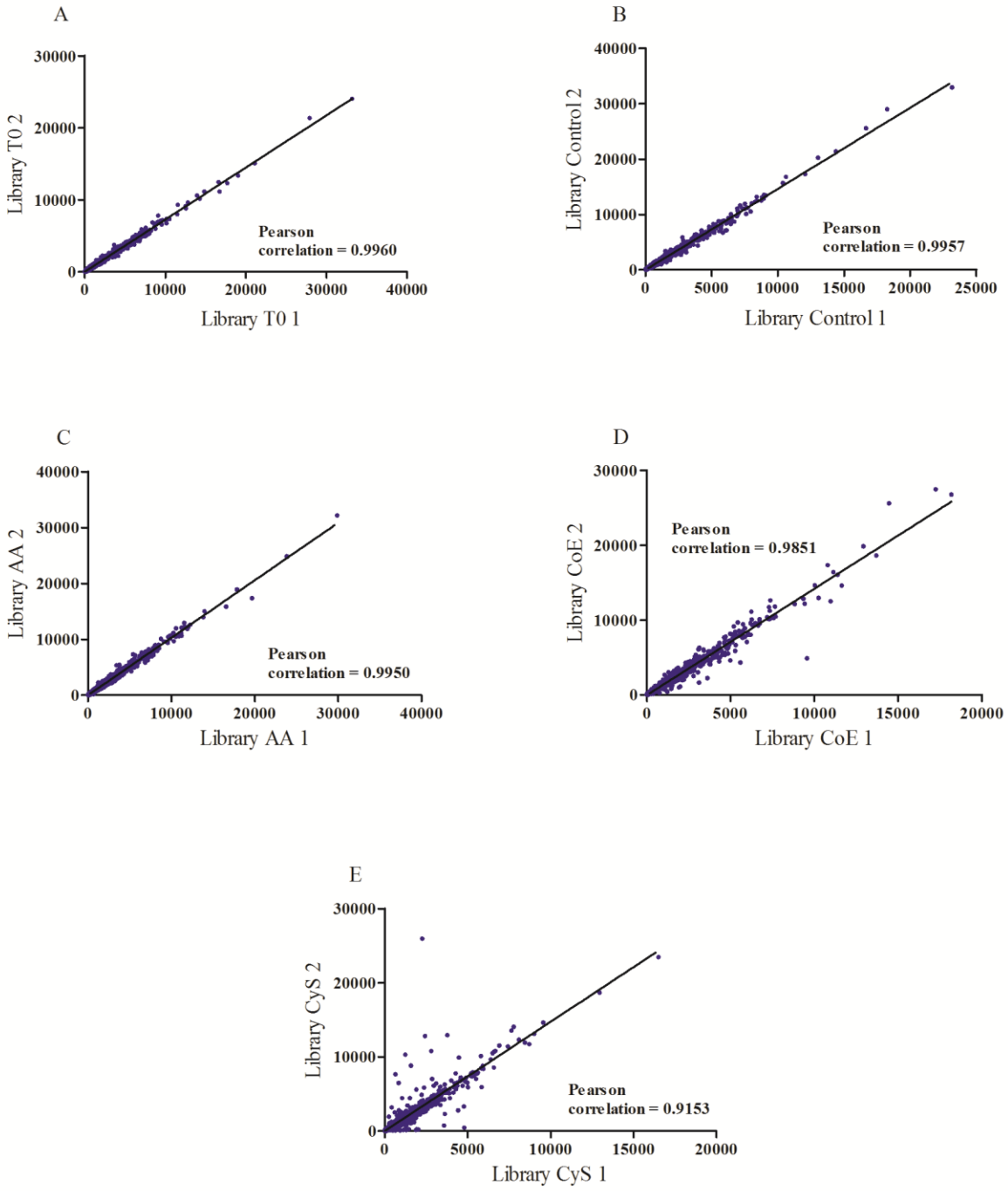

Figure 4-3. Reproducibility and stability of the number of reads. (A) Correlation of the number of reads in duplicates of the T0 library, (B) the control library, (C) the AA library, (D) the CoE library, and (E) CyS library.

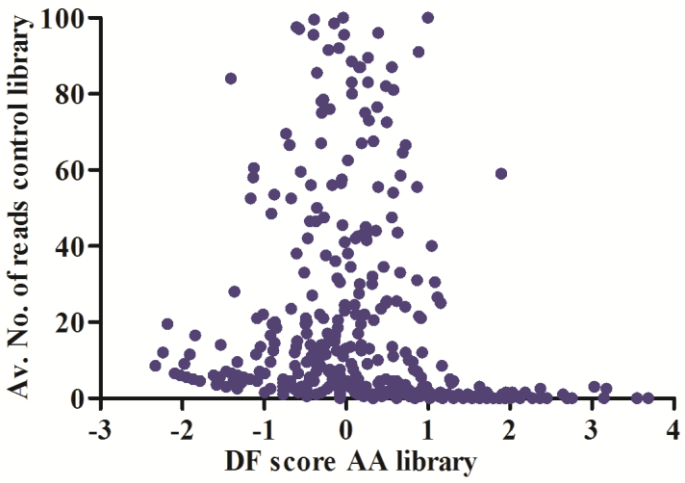

Figure 4-4. Variation of the DF scores for the AA library by the average number of reads of the control library. The graph shows only average number of reads below 100.

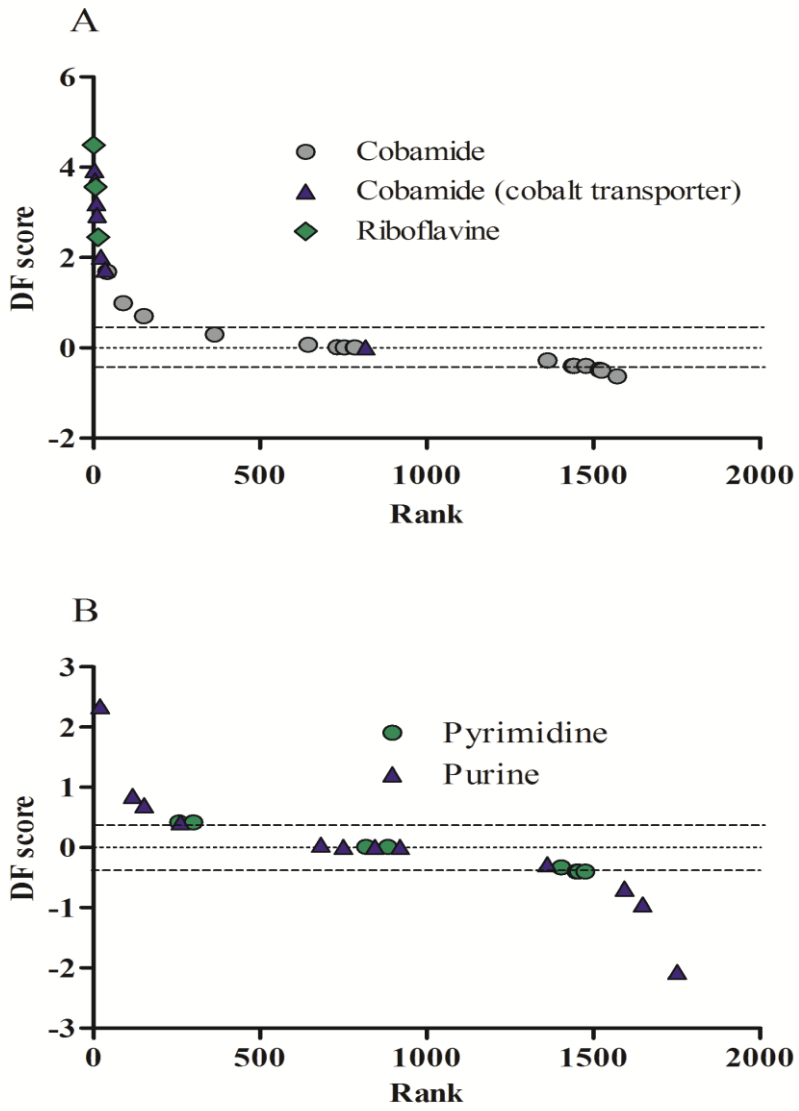

Figure 4-5. DF score distribution by rank for genes involved in different biosynthetic pathways following growth in coenzyme extract supplemented medium (CoE). (A) Genes involved in riboflavin and cobamide biosynthesis. (B) Genes involved in nucleotide biosynthesis. Rank represents a hierarchical organization of the genes in the genome of *M. maripaludis* modeled by the DF score. Dashed lines represent the limits for null DF score ($\leq$ 0.36 and $\geq$ -0.36) and dotted line represents zero in the Y axis.

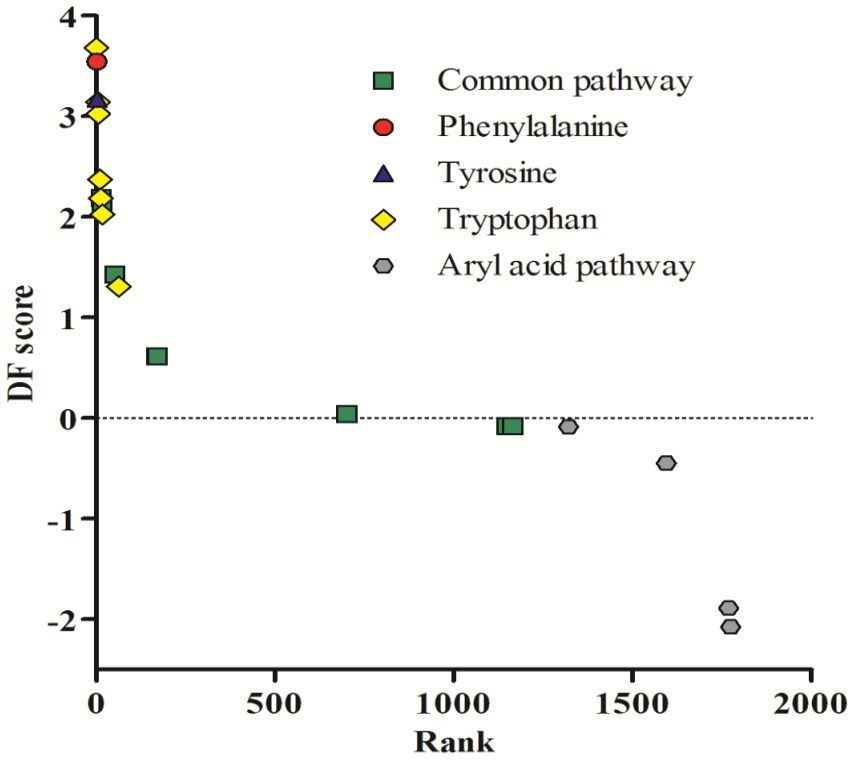

Figure 4-6. DF score distribution by rank for genes involved in aromatic amino acids biosynthesis following growth in aryl acids supplemented medium (AA). Rank represents a hierarchical organization of the genes in the genome of *M. maripaludis* modeled by the DF score. The dotted line represents zero in the Y axis.

Table 4-1: Genes involved in the biosynthesis of the unusual coenzymes of methanogenesis[a].

Biosynthetic pathway	Known genes	Predicted but unidentified genes
Methanofuran	1	7
Tetrahydromethanopterin	5	5
Coenzyme M	5	1
Coenzyme B	4	4
Coenzyme F_{430}	0	7
Coenzyme F_{420}	8	1

[a]Predicted genes were selected by literature search. It was assumed that one gene is involved in each missing reaction.

Table 4-2. Number of reads for mutant libraries T0 and following growth at T2.

Library name[a]	Total No of reads	Reads mapped to the genome	Percentage	Reads per gene
T0 (1)	4238439	3871524	91.34	2176
T0 (2)	3058661	2800415	91.56	1574
AA T2 (1)	3530968	3273882	92.72	1840
AA T2 (1)	3675362	3344440	91.00	1879
C+ T2 (1)	2721262	2501270	91.92	1406
C+ T2 (2)	3917146	3604829	92.03	2026
CoE T2 (1)	2758869	2517800	91.26	1415
CoE T2 (2)	3842260	3523115	91.69	1980
Cys T2 (1)	2006114	1837187	91.58	1032
Cys T2 (2)	2994100	2769177	92.49	1556

[a]Each library was sequenced after 14 generations (T2) of growth in control medium (McNd), coenzyme extract supplemented medium (McNde), aryl acids supplemented medium (McNda) or cysteine supplemented medium (McNdc) without puromycin selection.

Table 4-3. Hypothetical genes with increased DF score in CoE library and their predicted function.

Locus tag	Length (bp)	Predicted function[b]	Av. No. of reads (T2)[a]		DF score	Rank
			Control lib.	CoE lib.		
MMP0567	411	Thioesterase	59	786.5	2.58	13
MMP1343	807	β-lactamase domain	371.5	4028.5	2.39	17
MMP0269	1260	Unknown function	5	46	2.07	20
MMP1593	924	Predicted DNA binding domain	0.5	10	2	23
MMP0564	921	ATP-grasp domain	1	13.5	1.99	24
MMP1204	1035	Metallo-aminopeptidase	91	654	1.97	25
MMP1218	1269	amino acid biosynthesis (lysine)?- SDH domain	12.5	69.5	1.66	41
MMP1392	1467	Unknown function	2	14.5	1.65	42
MMP1499	522	predicted transcriptional regulator	67	343.5	1.63	43
MMP0543	531	Probable chorismate lyase	7	34.5	1.5	46
MMP1606	1191	phosphopantothenoylcysteine decarboxylase/phosphopantothenate-cysteine ligase	0.5	5.5	1.48	48
MMP1109	1119	Not known function (possible transmembrane protein)	14	56.5	1.35	57
MMP0656	627	Unknown function	4	18	1.35	58
MMP1275	576	Predicted transcriptional regulator	364.5	1386	1.34	59
MMP1243	657	ThiF/moaA/HesA family protein	125	471.5	1.33	60
MMP0422	483	Unknown function	82	295	1.28	63
MMP0406	798	Unknown function	33	111	1.2	71
MMP1196	561	Unknown function	127	419	1.2	72
MMP1594	570	Unknown function	25	83	1.18	73
MMP1234	756	ThiF/MoeB sulfur transfer protein	40	128.5	1.16	75
MMP0022	339	Unknown function	27.5	78.5	1.04	84
MMP1065	513	Probable Transcriptional regulator	448.5	1211.5	1	87
MMP0618	342	Unknown function	24	64	0.97	94
MMP1554	1320	SAM superfamily	308.5	792	0.95	99
MMP0236	417	Unknown function	292.5	738	0.93	101

MMP0944	1074	Putative transporter	1573	3957	0.93	102
MMP0053	771	Unknown function	432.5	1084	0.93	104
MMP0828	354	Unknown function	87	215	0.91	111
MMP0076	216	Putative deoxyribonuclease	7.5	19.5	0.89	114
MMP1118	354	Probable DNA polymerase II large subunit	69.5	165	0.87	116
MMP0631	444	Putative iron dependent repressor	435.5	1008.5	0.85	120
MMP1685	243	Unknown function	708.5	1620	0.84	122
MMP0352	939	Putative oxidoreductase (NADH dependent)	1325	2921	0.8	125
MMP1033	2946	Unknown function	3072.5	6510	0.76	130
MMP0121	321	Unknown function	140.5	291	0.73	135

[a] Genes with low number of reads (0-5) in the control library were considered on this table only if they possessed at least an increase of at least 5 reads in the CoE library.

[b] Prediction of the function was based primarily in the NCBI gene database but was curated by using additional bioinformatics tools like BLAST, IMG database and NCBI conserved domains.

Table 4-4. Known genes involved in the aromatic amino acid biosynthetic pathways in *M. maripaludis*.

Locus tag	Length (bp)	Gene symbol	Description	Av. No. of reads (T2) Control Lib.	Av. No. of reads (T2) AA Lib.	DF score	Rank
***de novo* Common pathway**							
MMP0936	873	aroE	shikimate 5-dehydrogenase	1.5	23	2.18	14
MMP1205	1320	aroA	3-phosphoshikimate-1-carboxyvinyltransferase	0.5	12.5	2.12	16
MMP1333	1161	aroC	Chorismate synthase	0	3.5	1.43	52
MMP0686	825	aroA'		0	1	0.62	169
MMP1394	666	aroD	3-dehydroquinate dehydratase	0	1	0.62	171
MMP0006	1167	aroB'?	Conserved hypothetical protein	141	158.5	0.04	700
MMP0320	921	-	shikimate kinase	0	0	-0.08	1149
MMP0578	348	aroQ	Chorismate mutase	0.5	0.5	-0.08	1164
***de novo* phenylalanine, tyrosine and tryptophan specific pathways**							
MMP1528	840	pheA	Prephenate dehydratase	0	36.5	3.55	2
MMP1514	1356	tyrA	Prephenate dehydrogenase	2.5	89.5	3.17	3
MMP1003	1194	trpB	Tryptophan synthase, beta chain	0	42	3.68	1
MMP1002	777	trpA	Tryptophan synthase, alpha chain	0	24	3.14	4
MMP1005	615	trpG	Anthranilate synthase component II	3	88	3.02	5
MMP1004	645	trpF	N-(5'phosphoribosyl)anthranilate isomerase	2.5	39.5	2.37	10
MMP1006	1323	trpE	Anthranilate synthase component I	1.5	23	2.18	13
MMP1007	975	trpD	Anthranilate phosphoribosyltransferase	1.5	19.5	2.03	17
MMP1008	699	trpC	Indole-3-glycerol phosphate synthase	0	3	1.31	63
pathway for conversion of aryl acids to aromatic amino acids							
MMP0315	612	iorB1	indolepyruvate oxidoreductase subunit beta 1	429.5	426	-0.09	1321
MMP0316	1860	iorA1	indolepyruvate oxidoreductase subunit alpha 1	3522.5	2429.5	-0.45	1594
MMP0713	1800	iorA2	indolepyruvate oxidoreductase subunit alpha 2	6579	1074	-1.89	1768
MMP0714	579	iorB2	indolepyruvate oxidoreductase subunit beta 2	400	53.5	-2.07	1774

Table 4-5. Known genes involved in the biosynthesis of cysteine and selenocysteine in *M. maripaludis*.

Locus tag	Length (bp)	Gene symbol	description	Av. No. of reads (T2) Control lib.	Av. No. of reads (T2) CyS lib.	DF score	Rank
Cysteine biosynthesis							
MMP0688	1656	sepRS	Phosphoseryl-tRNA synthetase	3	103	3.54	1
MMP1217	690	-	Hypothetical archaeal protein	0	3	1.67	32
MMP1060	1446	cysS	Cysteinyl-tRNA synthetase	2247.5	1732	0.02	849
MMP1240	1170	sepCysS	Sep-tRNA:Cys-tRNA synthase	1173	1768	-0.12	1187
Selenocysteine biosynthesis							
MMP0595	1329	SepSeCS	Sep-tRNA:Sec-tRNA synthase	0.5	13.5	2.55	3
MMP0879	1557	serS	Seryl-tRNA synthetase	0	7.5	2.42	5
MMP1490	822	-	O-phosphoseryl-$tRNA_{Sec}$ kinase	0	0	0.28	484

Table 4-6. Number of reads for the mutant library under steady state growth conditions.

Library name[a]	Total No of reads	Reads mapped to the genome	Percentage	Reads per gene
H-/P+ (T1)	2847925	2702999	94.91	1519
H-/P+ (T2)	3192572	3042138	95.29	1710
H-/P+ (T3)	2448603	2348162	95.90	1319
H-/P+ (T4)	3594424	3441625	95.75	1934
H-/P+ (T5)	4846589	4439051	91.59	2495
H+/P- (T1)	2671253	2453369	91.84	1379
H+/P- (T2)	3145108	2837581	90.22	1595
H+/P- (T3)	7497516	6671593	88.98	3750
H+/P- (T4)	2557492	2412491	94.33	1356
H+/P- (T5)	1993597	1872739	93.94	1052

[a] DNA was sequence for samples collected after 0 (T1), 3 (T2), 5 (T3), 8 (T4) and 10 (T5) dilutions during steady state growth without puromycin selection under hydrogen limitation/phosphate excess (H-/P+) or hydrogen excess/phosphate limitation (H+/P-). At T1, the fermentor had reach a culture absorbance of ~0.6 and the dilution was begun.

Table 4-7. Top ten genes with the highest number of reads growing under hydrogen limitation/phosphate excess condition in a steady state system.

			Number of reads at each time point				
Locus tag	**Length (bp)**	**Description**	**T1**	**T2**	**T3**	**T4**	**T5**
MMP1102	693	Phospholipase D/ Transphosphatidylase	1565966	3026360	2337696	3424263	4414539
MMP0233	408	Hypothetical archaeal protein	398878	2566	2276	3847	6002
MMP0234	1233	conserved hypothetical protein	14239	1187	1050	1340	1418
MMP0235	897	conserved hypothetical protein	18835	1334	859	1172	1254
MMP0680	699	Uracil Phosphoribosyltransferase	19993	1979	725	806	1140
MMP0601	897	hypothetical protein	10023	475	435	730	948
MMP0232	1065	Conserved Hypthetical Protein	15421	267	219	397	654
MMP0039	2729	conserved hypothetical protein	16179	477	336	523	526
MMP0681	1272	Xanthine/uracil permease family	28414	750	196	268	491
MMP0038	912	conserved hypothetical protein	18524	278	233	463	412

Table 4-8. Top ten genes with the highest number of reads growing under hydrogen excess/phosphate limitation condition in a steady state system.

			Number of reads at each time point				
Locus tag	**Length (bp)**	**description**	**T1**	**T2**	**T3**	**T4**	**T5**
MMP1102	693	Phospholipase D/ Transphosphatidylase	151708	2767482	6602015	2396257	1843969
MMP1303	1929	Sensory transduction histidine kinase	3431	1068	2515	4280	15800
MMP0233	408	Hypothetical Archael protein	863170	11192	11377	1664	1585
MMP0038	912	conserved hypothetical protein	1708	2266	3334	919	969
MMP0039	2729	conserved hypothetical protein	2033	3905	2689	556	565
MMP0601	897	hypothetical protein	1504	3383	3697	626	372
MMP0241	1767	Hypothetical archaeal protein	2745	6154	5951	482	248
MMP0680	699	Uracil Phosphoribosyltransferase	74617	3255	3003	352	233
MMP0681	1272	Xanthine/uracil permease family	83443	2604	1828	278	210
MMP0235	897	conserved hypothetical protein	2591	6213	4802	309	177

CHAPTER 5

CONSTRUCTION OF COENZYME M AUXOTROPHS IN *METHANOCOCCUS MARIPALUDIS*[3]

[3]**Sarmiento, F., Ellison, C.K. and W.B. Whitman.** To be submitted to *Archaea*.

Introduction

Hydrogenotrophic methanogens, such as *Methanococcus maripaludis*, possess a specialized metabolism. To generate methane, they only reduce CO_2 using H_2 or formate as the electron donor, in a process known as methanogenesis. Other methanoarchaea can use different substrates to perform this process including acetate, methylamine and other methyl-group containing compounds (8). The beginning of the methanogenesis pathways varies depending on the initial substrate, but the final step remains the same, where coenzyme M, the smallest known organic cofactor, plays a key role as the last methyl carrier (4). In short, methanogenesis ends when methyl-CoM reductase (MCR) catalyzes the reduction of methyl-CoM using coenzyme B (CoB) as an electron donor and releases methane. The oxidation of CoB yields a heterodisulfide with CoM (CoM-S-S-CoB), which is reduced to regenerate the thiols by heterodisulfide reductase (Hdr) (10). Without coenzyme M present to complete the biosynthesis of methane, the organism is unable to produce the necessary energy for function and growth.

The biosynthetic pathway of coenzyme M (CoM) in *Methanocaldococcus jannaschii*, an organism closely related to *M. maripaludis*, proceeds in five steps. Four enzymes involved in the biosynthesis of CoM have been biochemically characterized in *M. jannaschii* (3, 5-7). In addition, the genes encoding these enzymes have been identified in diverse methanogens, including *M. maripaludis*. The proposed pathway for the biosynthesis of CoM starts with the sulfonation of PEP by a phosphoenolpyruvate sulfotransferase (ComA). Then, a phosphosulfolactate phosphatase (ComB) hydrolyses phosphosulfolactate and a R-sulfolactate dehydrogenase (ComC) catalyses the oxidation of the (*R*)-sulfolactate intermediate to form sulfopyruvate. In the fourth step, a sulfopyruvate decarboxylase (ComDE) catalyzes the decarboxylation of sulfopyruvate to form sulfoacetaldehyde (5). For the final postulated step of

CoM biosynthesis, the enzyme which catalyzes the reductive thiolation of sulfoacetaldehyde to coenzyme M (ComF) has not yet been identified. In addition, the genomes of *Methanosarcina* species lack the genes for the first three steps of the pathway, suggesting that an alternative pathway for CoM biosynthesis exists (4).

M. maripaludis may have the ability to uptake coenzyme M from the medium. Previous studies have identified an energy-dependent transport system for coenzyme M within *Methanococcus voltae* (2). Coenzyme M auxotrophs have been randomly constructed and tested in this same methanoarchaea, but the specific mutated genetic loci was not identified (11). *M. maripaludis* is closely related to *M. voltae*, and they share many metabolic and physiologic features. Thus, is probable that the same uptake system is present in *M. maripaludis*.

In the present study, coenzyme M auxotrophs for the methanogenic archaeon *M. maripaludis* were made by *in vitro* transposon mutagenesis followed by transformation into the genome. This system was previously used to construct tryptophan auxotrophs in *M. maripaludis* (12). The selected target was the gene *comE* (MMP1689), which encodes one of the subunits of the enzyme that catalyzes the decarboxylation of sulfopyruvate to form sulfoacetaldehyde in the fourth step of coenzyme M biosynthesis. In minimal medium in the absence of coenzyme M, the mutant grew poorly and normal growth was restored by the addition of coenzyme M. Thus, coenzyme M stimulated but was not absolutely required for growth. These results confirmed the role of ComE in coenzyme M biosynthesis and the ability of *M. maripaludis* to take up CoM.

Materials and methods

Strains, primers, and plasmids.

Strains, primers and plasmids are summarized in Table 5-1. All primers made in this work were designed using the Primer3Plus software (13).

Culture conditions

E.coli was grown in Luria-Bertani medium at 37°C. For solid medium preparation, 1% agar was added. Ampicillin (100μg/ml) and kanamycin (50μg/ml) were supplemented when indicated.

Methanococcus maripaludis strain S2 was grown in 28-ml Balch tubes pressurized to 275 kPa with H_2/CO_2 (80:20, v/v) at 37°C in 5 ml of minimal (McN) and complex (McCV) media reduced with 3 mM cysteine in anaerobic conditions as described previously (14). When indicated, McCV was supplemented with 3 mM coenzyme M to form McCm medium, and McN was supplemented with 10 mM acetate to form McNA. Before inoculation, 3 mM sodium sulfide was added. Puromycin (2.5 μg/ml) was added when indicated. For solid medium preparation, 1% agar was added. To avoid residual coenzyme M contamination, the glassware used in this study was heated for 4-5 hours at 180 °C.

Construction of *comE::*Tn5 mutant

PCR amplification of the *comE* genetic region was performed using primers RegcomEF and RegcomER and *Taq* DNA Polymerase (Fermentas). The cycling program proceeded as follows. After 3 min of denaturation at 95°C, the following steps were performed for 30 cycles: denaturation at 95°C for 30 s, annealing at 58°C for 3 min and extension at 72°C for 3 min. A final extension was performed at 72 °C for 10 min. The 3089 bp-amplification product and the plasmid vector pUC18 were digested using restriction enzymes BamHI and XbaI at 37°C for 90 min and ligated together using T4 ligase (New England Biolab) at 25°C for 30 min. The resulting plasmid, pcomE (5775 bp), was transformed into Transformax^{TM} EC100^{TM} electrocompetent cells (Epicentre) by electroporation in a 2-mm electrode gap cuvette (2.36 kV). Samples were resuspended in 1ml of LB medium, incubated for 1 hour at 37°C, and spread on LB agar plates in the presence and absence of ampicillin (100μg/ml) to determine transformation

efficiency. Clones were picked, and plasmid was extracted using a Zyppy Plasmid miniprep Kit (Zymo Research). For screening, 1 μg of plasmid was digested using the following sets of restriction enzymes (Sph1/Nde1/Alwn1, and Kpn1/Nco1 (New England Biolab)) to produce fragments of predicted and detectable sizes.

The Tn5<KAN-2-pac> transposon (12) was PCR amplified using 5' phosphorylated oligonucleotides ME-Plus9-3' and ME-Plus9-5' (Epicenter) and Phusion high-fidelity DNA polymerase (New England Biolabs). The cycling program proceeded as follows. After 3 min of denaturation at 98°C, the following steps were performed for 30 cycles: denaturation at 98°C for 30 s, annealing at 55°C for 60 s and extension at 72°C for 140 s. A final extension was performed at 72 °C for 10 min. The PCR product was purified using the DNA clean and concentrator -5 kit (Zymo Research) and resuspended in TE buffer (10mM Tris-HCl buffer, pH 7.5, 1mM EDTA). *In vitro* transposition was achieved by incubation of 80 ng of the transposon, 200 ng of pComE and 1 μl of EZ-Tn5 transposase (1U/μl, Epicentre) following the manufacturer instructions (Epicentre). The mixture was transformed into Transformax^{TM} EC100^{TM} electrocompetent cells (Epicentre) following the manufacturer instructions. Dilutions of the mixture were spread onto plates containing ampicillin (100μg/ml) or kanamycin (50μg/ml). Isolated Kanamycin resistance colonies were picked and grown in broth in the presence of ampicillin and kanamycin. Plasmids with the transposon insertions were extracted using a Zyppy Plasmid miniprep Kit (Zymo Research). For screening, insertions were sequenced from the end of the transposon using the primers KAN-2RP-1out2 (12) at the University of Michigan DNA Sequencing Core. Two plasmids with different insertions in the *comE* gene were found and named pComET1 and pComET2.

Transformation into *M. maripaludis* with plasmid pComET1and pComET2 was performed as described previously (14). Briefly, wild-type *M. maripaludis* S2 cells were grown to an absorbance of 0.9 in 5ml of McCV and washed and resuspended in transformation buffer (50 mM Tris Base, 0.35 M sucrose, 0.38 M NaCl, 0.00001% resazurin, and 1 mM $MgCl_2$ with pH adjusted to 7.5). One μg of pComET1 or pComET2 was added followed by 0.225ml of 40% PEG transformation buffer. Cells were incubated for 1 hour at 37^{o}C without shaking, washed and resuspended in 5 mL McCm medium containing 3 mM sodium sulfide. Cells were grown overnight at 37°C.

The transformation mixture was serially diluted, and 500 μl were spread onto McCm agar plates supplemented with puromycin. After incubation for 6 days in the presence of ~150 kPa of H_2/CO_2 (80:20, v/v) at 37°C, isolated puromycin-resistant colonies were picked and re-streaked onto McCm agar plates supplemented with puromycin. After growth under the same conditions, isolated puromycin-resistant colonies were transferred to stoppered culture tubes containing 5 ml of McCm medium plus puromycin, pressurized to 275 kPa and incubated at 37°C. Frozen stocks were prepared in McCm medium +30% glycerol (v/v), and the suspensions were stored at -80°C. The new strains with the integrated plasmids pComET1 and pComET2 were named M. *maripaludis* S201 and S202, respectively (Table 1). To reisolate the mutant S201, frozen stocks were inoculated in 5 ml of McCm plus puromycin, pressurized to 275 kPa of H_2/CO_2 (80:20, v/v) and incubated at 37°C. The cell suspension was serially diluted and plated in serum-bottle agar slants with McCm medium plus puromycin (14). After 5 days of incubation in the presence of 137 kPa of H_2/CO_2 (80:20, v/v) at 37°C, isolated puromycin-resistant colonies were picked and transferred to stoppered culture tubes containing 5 ml of McCm medium plus puromycin. Frozen stocks were prepared in McCm medium + 30% glycerol (v/v), and the suspensions were stored at

-80°C. Genomic DNA was extracted using ZR Fungal/Bacterial DNA MiniPrep (Zymo Research).

Verification of the genotype of the mutant was determined through PCR using two sets of primers (Table 5-1). The first set consisted of a forward primer (KAN-2RP-1out2) from the end of the transposon and a reverse primer from the surrounding gene MMP1688 (MMP1688R), and was used to demonstrate the transposon insertion. The second set consisted of forward primer comEF and reverse primer comER, which amplified the wild type *comE* gene. The products were PCR amplified using Phusion high-fidelity DNA polymerase (New England Biolabs). The cycling program for both set of primers proceeded as follows. After 1 min of denaturation at 98°C, the following steps were performed for 30 cycles: denaturation at 98°C for 15 s, annealing at 59°C for 30 s and extension at 72°C for 30 s. A final extension was performed at 72 °C for 10 min. The PCR products were loaded and run in a 1% agarose gel and stained with ethidium bromide.

To determine the phenotype of the mutant, ~$1x10^5$ mutant or wild type S2 cells were inoculated in 5 mL of McNA in the presence or the absence of 146 µM of coenzyme M. The cultures were grown in 28-ml Balch tubes pressurized to 275 kPa with H_2/CO_2 for 90 hours in the absence of antibiotic. Second passages were done by inoculating ~$1x10^4$ mutant or wild type S2 cells from the previous cell suspensions into 5 mL of McNA in the presence or the absence of 146 µM of coenzyme M, and incubated under the same conditions for 100 hours. Third passages were performed by inoculating ~$1x10^4$ mutant or wild type S2 cells from the second passage into 5 mL of McNA in the presence or the absence of 146 µM of coenzyme M, and incubated under the same conditions for 480 hours.

Results

Disruption of *comE* gene impaired growth in the absence of coenzyme M

The gene encoding comE was disrupted by *in vitro* transposon mutagenesis with the Tn5<KAN-2-pac> transposon (12), which encodes puromycin resistance in methanococci. In the first step, the transposon was randomly inserted in a plasmid containing a 3089 bp chromosomal region of *M. maripaludis* harboring the *comE* gene (Figure 5-1A). Two independent transposon insertions in the *comE* gene were found and mapped by sequencing (Figure 5-1B). The *comE* gene was replaced in the genome of *M. maripaludis* by the transposon-disrupted versions through homologous recombination, generating strains S201 and S202 with insertions locations at 336 bp and 395 bp from the 5'start of the gene, respectively (Figure 5-1B). Both insertions were localized close to each other in the middle of the gene, so only one of the strains (S201) was selected for the following experiments. The replacement was genotypically confirmed by two independent PCR amplifications (Figure 5-2). When using primers to amplify the *comE* gene, only the wild type strain displayed a band at approximately 600 bp. Under these conditions, the mutant did not yield an amplification product, presumably because of the length and high GC content of the *pac* cassette inside the transposon prevented amplification under these conditions. In contrast, using a primer from the end of the transposon into the *comE* gene, a product of the expected size was found with the mutant strain S201 but not the wild type strain.

When grown in minimal medium plus acetate in the absence of coenzyme M, the mutant strain S201 displayed a longer lag phase, slower growth rate during exponential phase, and lower growth yield than the wild type (Figure 5-2C). This "sick" phenotype was restored by the addition of 146 µM coenzyme M (Figure 5-2C). These results were observed after a mutant culture grown with 146 µM coenzyme M was diluted 5,000-fold into McNA, which gives a final

concentration of around 29 nM from CoM carryover. Previous studies in *M. voltae* demonstrated that concentrations of coenzyme M less than 70 nM cannot support growth of this methanoarchaeon (11). This "sick" phenotype was accentuated after two sequential passages of the mutant cells in minimal medium in the absence of coenzyme M (Figure 5-3). For each sequential passage, the cells were further diluted 25,000-fold into McNA, to yield carryovers of coenzyme M of 1.2 pM or less. In the third passage the doubling time in exponential phase (means $\pm$ standard deviations of three cultures) of S2 in McNA medium in the absence of coenzyme M and S201 in McNA medium in the absence and presence of coenzyme M were 4.9 $\pm$ 0.4, 162.2 $\pm$ 21.7, and 5.9 $\pm$ 0.4 hours, respectively. The growth yields (means $\pm$ standard deviations of three cultures) of S2 in McNA medium in the absence of coenzyme M and S201 in McNA medium in the absence and presence of coenzyme M were 0.53 $\pm$ 0.01, 0.22 $\pm$ 0.03, and 0.53 $\pm$ 0.02 mg dry weight/ml of culture, respectively. Assuming a coenzyme M content of 0.43 nmol/mg dry weight (9), the minimum rates of CoM biosynthesis necessary to support the observed growth rates were 0.06 $\pm$ 0.004 and 0.002 $\pm$ 0.0002 nmol/mg dry weight hour for S2 and S201 respectively, indicating that the mutant retained no more than 3% of the wild type level of coenzyme M production.

The glassware used to prepare the medium was heat-treated to remove traces of contaminating CoM (1). Different times and temperatures were evaluated (data not shown). After heat-treating clean glassware and glassware with the prior addition of CoM sufficient for a final medium concentration of 250 nM, the growth rates for the mutant strain S201 (means $\pm$ standard deviations of three cultures) were 160 $\pm$ 17 and 165 $\pm$ 13, which was similar to the growth rate of the mutant strain S201 on the third passage in glassware heat-treated for 4 hours at 180°C. Thus,

the low amount of growth of the mutant was not the result of small amounts of contaminating CoM in the medium.

The growth of the mutant strain S201 was tested in the presence of different concentrations of coenzyme M (Figure 5-4). The doubling times of the mutant strain S201 in McNA medium supplemented with 146 µM, 1.4 µM, 250 nM, 10 nM, 0.5 nM and the absence of coenzyme M were: 5.3 $\pm$ 0.3, 5.3 $\pm$ 0.5, 7.2 $\pm$ 1, 77 + 3, 90 $\pm$ 11, and 153 $\pm$ 16 hours, respectively. Concentrations of coenzyme M over 1.4 µM fully restored the growth rates. In contrast, when the concentration of coenzyme M was 10 nM or less, the mutant grew very poorly. Thus, growth of the mutant was severely impaired in the absence of CoM, but CoM was not essential for growth. This conclusion is supported by the increase in yields when the medium was supplemented with small amounts of CoM. Based upon the levels measured in whole cells of 0.43 nmol/mg dry weight (13), the expected growth yield in the presence of 10 nM CoM is approx. 0.023 mg dry weight/ml, however, the observed increase in growth yield was 0.32 $\pm$ 0.01 mg dry weight/ml, or 14 times higher (Figure 5-4).

Discussion

A partial auxotroph for coenzyme M was constructed in *M. maripaludis* through the disruption of the gene *comE*, which encodes the sulfopyruvate decarboxylase β-subunit involved in the fourth step of the proposed pathway of CoM biosynthesis (4). The mutant exhibits impaired growth in minimal medium + acetate, which is fully restored upon the addition of coenzyme M, indicating that ComE is important for the biosynthesis of coenzyme M. However, the transposon insertion of *comE* did not yield a complete CoM auxotroph, suggesting that *M. maripaludis* still possess a source of coenzyme M. The carryover amounts of coenzyme M from the original inoculum were not enough to support a continuous growth, and the heat-treatment of the

glassware suggest that the mutant strain S201 did not grow because of exogenous coenzyme M, but it is still able to produce small amounts of this coenzyme. The manner that coenzyme M is still produced is not known. A possible explanation is that the transposon disrupted *comE* gene still retains some low activity, which allows *M. maripaludis* to grow slowly. Alternatively, a different promiscuous enzyme could replace the sulfopyruvate decarboxylase in the CoM biosynthetic pathway. A protein homology search against the genome of *M. maripaludis* S2 revealed the presence of the protein MMP0142, which possesses 28% sequence similarity in a 120 aa region of overlap with ComE. Like ComE, this enzyme belongs to the thiamine pyrophosphate (TPP) superfamily. MMP0142 exhibits a pyrimidine (PYR) binding domain which is found in many key metabolic enzymes that use TPP as cofactor. One of the subgroup of the TPP superfamily involves decarboxylases which can use pyruvate derivatives as substrate such as phosphonopyruvate decarboxylase from *Streptomyces hygroscopicus* and sulfopyruvate decarboxylase (ComDE) from *Methanocaldococcus jannashii*. Presumably, MMP0142 is involved in a different biosynthetic pathway in *M. maripaludis* but possesses some fortuitous affinity with sulfopyruvate. A final possibility is that *M. maripaludis* possesses an unknown alternative pathway for coenzyme M. Nevertheless, to solve this enigma more experiments have to be performed.

References

1. **Balch, W. E., and R. S. Wolfe.** 1976. New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressureized atmosphere. Appl Environ Microbiol **32:**781-91.

2. **Dybas, M., and J. Konisky.** 1989. Transport of coenzyme M (2-mercaptoethanesulfonic acid) and methylcoenzyme M [(2-methylthio)ethanesulfonic acid] in *Methanococcus voltae*: identification of specific and general uptake systems. J Bacteriol **171:**5866-71.

3. **Graham, D. E., M. Graupner, H. Xu, and R. H. White.** 2001. Identification of coenzyme M biosynthetic 2-phosphosulfolactate phosphatase. A member of a new class of Mg(2+)-dependent acid phosphatases. Eur J Biochem **268:**5176-88.

4. **Graham, D. E., and R. H. White.** 2002. Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics. Nat Prod Rep **19:**133-47.

5. **Graham, D. E., H. Xu, and R. H. White.** 2002. Identification of coenzyme M biosynthetic phosphosulfolactate synthase: a new family of sulfonate-biosynthesizing enzymes. J Biol Chem **277:**13421-9.

6. **Graupner, M., H. Xu, and R. H. White.** 2000. Identification of an archaeal 2-hydroxy acid dehydrogenase catalyzing reactions involved in coenzyme biosynthesis in methanoarchaea. J Bacteriol **182:**3688-92.

7. **Graupner, M., H. Xu, and R. H. White.** 2000. Identification of the gene encoding sulfopyruvate decarboxylase, an enzyme involved in biosynthesis of coenzyme M. J Bacteriol **182:**4862-7.

8. **Hedderich, R., and W. B. Whitman.** 2006. Physiology and biochemistry of the methane-producing Archaea, p. 1050-1079. *In* M. Dworkin, S. Falkow, E. Rosenberg, K. Schleifer, and E. Stackebrandt (ed.), The Prokaryotes, Third ed, vol. 2. Springer-Verlag, New York.

9. **Liu, Y., M. Sieprawska-Lupa, W. B. Whitman, and R. H. White.** 2010. Cysteine is not the sulfur source for iron-sulfur cluster and methionine biosynthesis in the methanogenic archaeon *Methanococcus maripaludis*. J Biol Chem **285:**31923-9.

10. **Liu, Y., and W. B. Whitman.** 2008. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. Ann N Y Acad Sci **1125:**171-89.

11. **Micheletti, P. A., K. A. Sment, and J. Konisky.** 1991. Isolation of a coenzyme M-auxotrophic mutant and transformation by electroporation in *Methanococcus voltae*. J Bacteriol **173:**3414-8.

12. **Porat, I., and W. B. Whitman.** 2009. Tryptophan auxotrophs were obtained by random transposon insertions in the *Methanococcus maripaludis* tryptophan operon. FEMS Microbiol Lett **297:**250-4.

13. **Rozen, S., and H. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol **132:**365-86.

14. **Sarmiento, F. B., J. A. Leigh, and W. B. Whitman.** 2011. Genetic systems for hydrogenotrophic methanogens. Methods Enzymol **494:**43-73.

15. **Whitman, W. B., J. Shieh, S. Sohn, D. S. Caras, and U. Premachandran.** 1986. Isolation and characterization of 22 mesophilic methanococci. Systematic and Applied Microbiology **7:**235-240.

Table 5-1. List of strains, plasmids and primers

Strains	**Description**	**Source**
M. maripaludis S2	Wild type	(15)
M. maripaludis S201	*comE::*Tn*5*<KAN-2-pac >	This work
M. maripaludis S202	*comE::*Tn*5*<KAN-2-pac>	This work
E. coli EC100	F *mcrA* Δ(*mrr-hsdRMS mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139* Δ(*ara, leu*)*7697 galU galK* λ *rpsI* (str) *nupG*	Epicentre
Plasmids		
pUC18	*E. coli* plasmid cloning vector	GenScript
pComE	Plasmid pUC18 containing the genetic region of gene *comE*	This work
PComET1 and 2	pComE containing Tn*5* transposon insertions in *comE* gene	This work
pMEVI	Shuttle plasmid for *M. maripaludis*	(14)
Primers		
RegcomEF	5’AAAAAAGGATCCCGGATCTGACCCATACAATAGAG	This work
RegcomER	5’-AAAAAATCTAGAATGGATGGATTG GCAGTTTC	This work
ME-Plus9-3’	5’-CTGTCTCTTATACACATCTCAACCATCA	Epicentre
ME-Plus9-5’	5’-CTGTCTCTTATACACATCTCAACCCTGA	Epicentre
KAN-2RP-1out2	5’-GCAATGTAACATCAGAGATTTTGAC	(12)
MMP1688R	5’-CGAATGGATTCTTTTGAACTTTT	This work
comEF	5’-TTGCGTTCATAAATCTGTGTTT	This work
comER	5’-AATGGAATACGTGACCGATG	This work
M13R	5’-CAGGAAACAGCTATGACC	

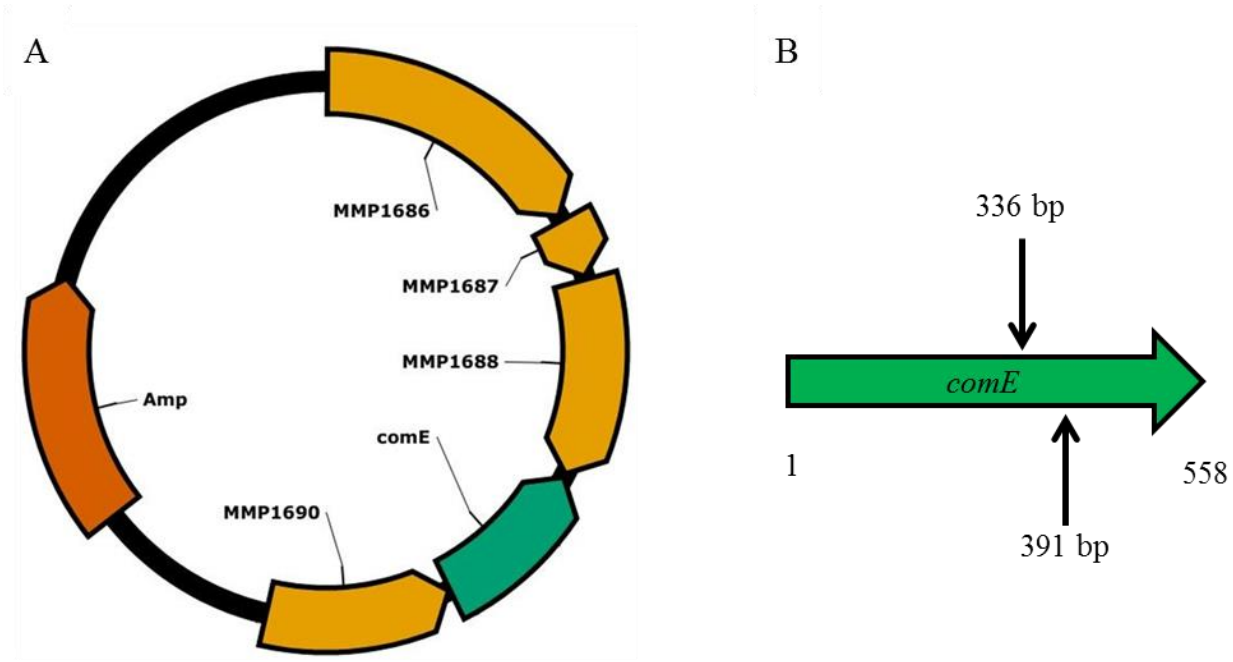

Figure 5-1. Map of pComE and the sites of the transposon insertions. (A) Plasmid pComE made by cloning the *comE* genetic region in a pUC18 plasmid. The region of the *M. maripaludis* genome cloned (3089 bp) includes the genes MMP1686, 1687, 1688 and1690 orange) and the gene *comE* (green). (B) Insertion positions of the Tn5<KAN-2-pac> transposon into the gene *comE* (558 bp). Insertion at 336 bp into the gene corresponds to the mutation in strain *M. maripaludis* S201, insertion at 395 bp corresponds to the mutation in strain *M. maripaludis* S202.

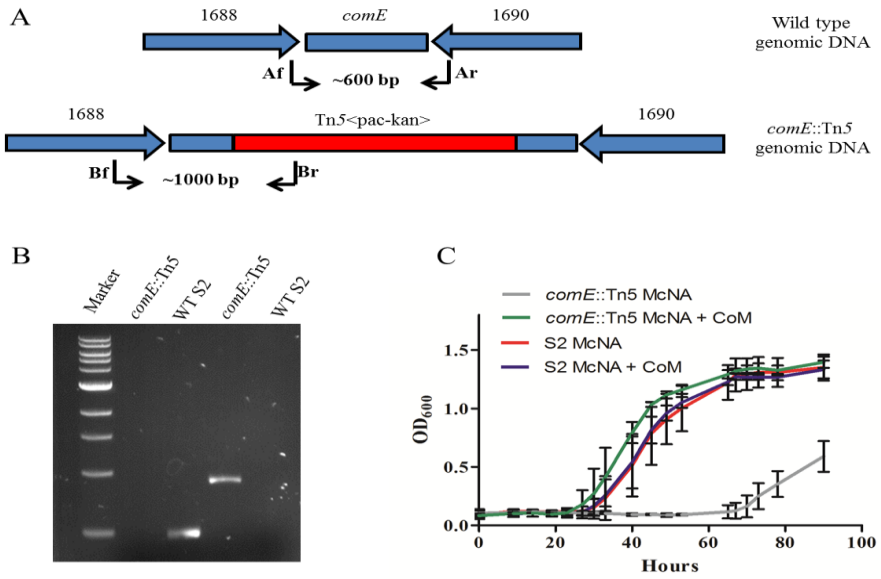

Figure 5-2. Characterization of the *comE*::Tn5 mutant strain S201. (A) Genetic maps of the gene *comE* in the wild-type strain and the disrupted version by the insertion of the Tn5<KAN-2-pac> transposon in the *comE*::Tn5 mutant strain S201. Numbers indicate the MMP identification, and the black arrows indicate primers used for PCR amplification. (B) Genotypic characterization of the *comE*::Tn5 mutant strain 201 by PCR amplification. Lane 1, Standard 1kb ladder (New England Biolab); Lane 2 and 3, PCR amplifications of *comE* using primers comEF (Af) and comER (Ar) for genomic DNA of the *comE*::Tn5 mutant strain S201 and wild type strain S2, respectively; Lane 4 and 5, PCR amplifications using a primer from the end of the transposon (KAN-2RP-1out2; Br) and a primer that binds upstream of the gene *comE* (Bf)for genomic DNA of *comE*::Tn5 mutant strain S201 and wild type strain S2, respectively. (C) Growth curves of the wild-type and *comE*::Tn5 mutant strain 201 in minimal medium + acetate (McNA) and McNA supplemented with coenzyme M. A culture of the mutant grown with 146 µM of coenzyme M was diluted 5,000-fold into McNA containing the indicated concentration of added CoM. The error bars indicate the standard deviation of three independent cultures.

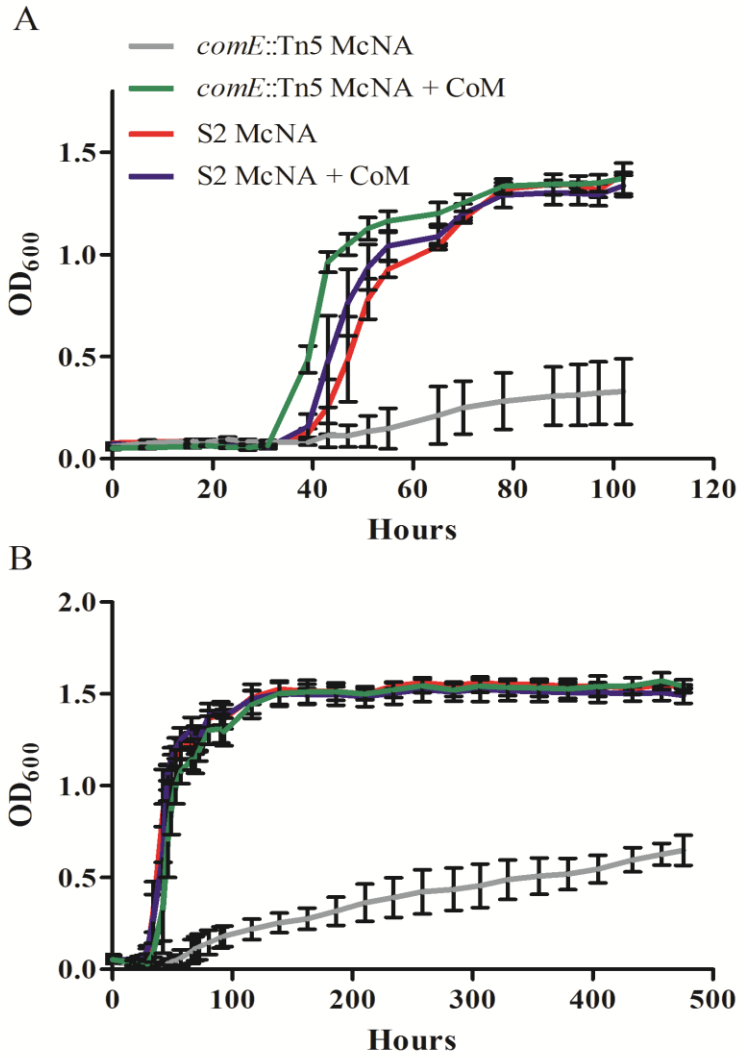

Figure 5-3. Growth of the wild-type and *comE*::Tn5 mutant strains in minimal medium plus acetate (McNA) in the presence and the absence of coenzyme M after two passages (A) and three passages (B). The error bars indicate the standard deviation of three independent replicates.

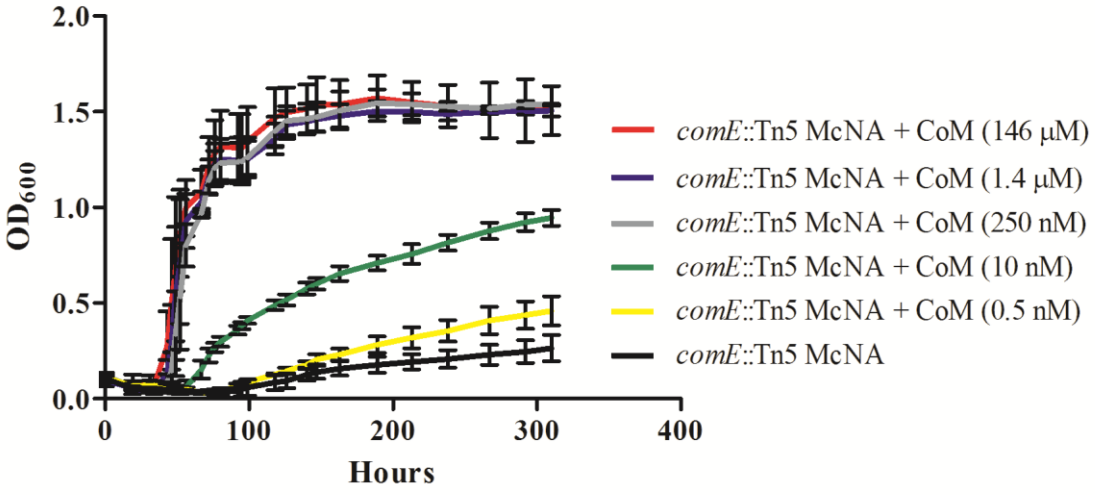

Figure 5-4. Growth of the *comE*::Tn5 mutant strain 201 in response to varying concentrations of coenzyme M. Cultures were grown in McNA after three transfers in medium without CoM. The error bars indicate the standard deviation of three independent replicate cultures.

CHAPTER 6

CONCLUSIONS

Since the pioneering work of Carl Woese and his collaborators reclassified the prokaryotes into the Archaea and the Bacteria, scientists have tried to understand the core features that define the domain Archaea. Archaea are ancient forms of life that have likely inhabited our planet for more than three billion years in different marine environments. Over time they have adapted to all sort of conditions and have developed unique characteristics and novel metabolisms that expand our biological and biochemical understanding of life, providing key pieces to the complex puzzle of evolution.

At first glimpse, Archaea resemble Bacteria. They are both unicellular microorganisms which share structural similarities in size and shape. However, they present important differences at genetic, molecular and metabolic levels. Indeed, Archaea exhibit a mosaic nature in the functionality and machinery of the information processing systems (replication, transcription, and translation), where they share a higher similarity with Eukarya, but retain some characteristics of Bacteria. In addition, they harbor unique features which belong only to the domain Archaea, increasing the complexity of these microorganisms.

In an effort to increase our understanding of the physiology of Archaea, comprehensive whole-genome analysis of gene function by transposon mutagenesis and deep sequencing (Tn-seq) has been successfully implemented in *Methanococcus maripaludis*, a hydrogenotrophic methanogenic representative of the domain Archaea (chapter 3). Many genes involved in the information processing system proved possibly essential for growth and provided direct evidence for their roles. Previously, these genes have been identified based upon their similarity to

eukaryotic homologs, indicating a close relation between archaea and eukaryotes. Surprisingly, the archaeal-specific DNA polymerase D (PolD) was essential for growth, suggesting that it performs a fundamental role in DNA replication. In contrast, DNA polymerase B (PolB), found in all three domains of life, was non-essential and was not required for viability. Interestingly, PolD is present in all archaeal phyla with the exception of crenarchaeotes, suggesting a clear evolutionary division within the archaeal domain. These observations demonstrate a fundamental change in the replication mechanism among domains and suggest an unanticipated variability in archaeal DNA replication, one of the most conserved processes in cellular life. The observed differences in replication among the Archaea could be a direct consequence of a division in an early stage of evolution before the replication system was fully formed. Alternatively, the differences observed in the replication mechanism may have evolved relatively late. If this is true, DNA replication may be more variable that thought in other groups of organisms.

Other interesting results were drawn from the functional analysis of *M. maripaludis*. For example, from four known homologs of the minichromosome maintenance (MCM) genes, only one was essential for growth, suggesting an interesting assembly of the replicative helicase. Genes for the two types of RNases H and for the flap endonuclease 1 (Fen-1) were non-essential, suggesting a certain level of functional redundancy. Only two of the three homologs of the replication protein A (RPA) proved essential, suggesting a different configuration of these proteins around ssDNA. These functional observations in *M. maripaludis* added to the intriguing initiation process of DNA replication suggest a unique information processing system in Methanococcales, which may have been shaped by different evolutionary pressures over time.

In addition, many genes encoding hypothetical proteins proved to be essential and may enable important but unidentified functions unique to Archaea, methanogens, the order

Methanococcales, or specifically to *M. maripaludis* S2. For instance many of these genes may be involved in poorly described methanogenesis related processes, including coenzyme biosynthesis. Further detailed genetic and biochemical studies will help in their characterization.

In chapter four, the newly developed Tn-seq technology was also used to identify unknown genes involved in specific metabolic processes of *M. maripaludis*. When batch conditions were tested, previously unknown genes possibly involved in the biosynthesis of diverse vitamins, coenzymes and cofactors were detected. For instance, when cells were grown in medium supplemented with aryl acids, the gene MMP1101, which encode an aminotransferase, displayed an increased fitness, suggesting that this enzyme may be the missing aminotransferase for tyrosine and phenylalanine. Thus, this study yielded some interesting but inconclusive results which need to be further tested by using other genetic and biochemical tools. However, under steady-state growth conditions, the results were surprising and difficult to explain in a biological context. Because of the speed and resolution of Tn-seq, this technique has potential to efficiently conduct fitness analyses in diverse contexts when the conditions are carefully designed. Future improvements and subsequent application of this technique not only on *M. maripaludis* but in other Archaea are expected to generate many new hypotheses of gene function that will complement the current knowledge and understanding of metabolic pathways in Archaea.

Finally, in chapter five, a leaky coenzyme M auxotroph was constructed in *M. maripaludis* by transposon disruption of the *comE* gene, which encodes a step in the coenzyme M biosynthetic pathway. These results validate *in vivo* the proposed pathway for coenzyme M biosynthesis. However, the disruption of the *comE* gene was not lethal, suggesting that the *M.*

maripaludis possess an alternative source for biosynthesis of coenzyme M or the transposon insertion allowed expression of a gene product with partial activity.

APPENDIX A

CHAPTER 3 SUPPLEMENTARY INFORMATION

Material and Methods

Strain and culture conditions

Methanococcus maripaludis strain S2 was grown in minimal (McN) and rich (McCV) media reduced with 3 mM cysteine (25). In some experiments McCV was supplemented with 3 mM coenzyme M to form McCm. Cultures of 5 mL were grown in 28-ml Balch tubes pressurized to 275 kPa with H_2/CO_2 (80:20, v/v). Cultures of 20 ml were grown in 160-ml serum bottles pressurized to 137 kPa with H_2/CO_2 (80:20, v/v). Before inoculation, 3 mM sodium sulfide was added as the sulfur source. Puromycin (2.5 μg/ml) and/or ampicillin (100 μg/ml) were added when indicated.

Enrichment probe and Illumina sequencing primer

The 61-nucleotide 5' dual biotinylated transposon-chromosome junction enrichment probe (5'-TGTGCAATGTAACATCAGAGATTTTGAGAAGCTTCAGGGTTGAGATGTGTATAAGAGACAG) and 26-nucleotide primer for Illumina sequencing (5'-AGGGTTGAGATGTGTATAAGAGACAG) were designed to bind to the end of the transposon through Primer 3 (24) and purified by HPLC. This design allowed the sequencing read to start just after the insertion site (16).

In vivo transposon mutagenesis

The Tn5<KAN-2-pac> transposon (23) was PCR amplified using 5' phosphorylated oligonucleotides ME-Plus9-3' and ME-Plus9-5' (Epicenter) and Phusion high-fidelity DNA polymerase (New England Biolabs). The cycling program proceeded as follows. After 3 min of denaturation at 98°C, the following steps were performed for 30 cycles: denaturation at 98°C for 30 s, annealing at 55°C for 60 s and extension at 72°C for 140 s. A final extension was performed at 72 °C for 10 min. The PCR product was purified using the DNA clean and

concentrator -5 kit (Zymo Research) and resuspended in TE buffer (10mM Tris-HCl buffer, pH 7.5, 1mM EDTA). Production of stable transposomes was achieved by incubation of 100 ng of the transposon with 2 µl of EZ-Tn5 transposase (1U/µl, Epicentre) in the presence of 2.5 µl of 100% glycerol at room temperature for 30 min. The transposome was stored at -20°C for up to 12 months.

The transposomes were transformed into *M. maripaludis* S2 by the polyethylene glycol method (25) with the following modifications. Cultures were grown in 5 ml of McCV broth to an absorbance (600 nm) of 0.8-0.9 and washed twice with McN buffer prepared without Mg^{+2}. Cells were resuspended in modified transformation buffer (TB, 50 mM Tris-Base, 0.35 M sucrose, 0.38 M NaCl, 0.00001% resazurin), and 4 µl of the anaerobic transposome suspension was added. Polyethylene glycol (8000) diluted in modified TB was then added, and the cells were incubated for 1 hour at 37°C.

After transformation, 4.5 ml of cells were spread onto eight McCm agar plates supplemented with puromycin. The remaining 0.5 ml were serially diluted and plated to calculate the transformation efficiency. After incubation for 6 days in the presence of 100 kPa of H_2/CO_2 (80:20, v/v) at 37 °C, puromycin-resistant colonies were resuspended in McCm medium + 30% vol/vol glycerol using a sterilized glass spreader. The suspension was stored at -80°C.

High-throughput insertion tracking by deep sequencing

Five µg of genomic DNA were sheared to an average fragment size of 500 bp using Adaptive Focused Acoustic Technology (AFA, Covaris). Illumina DNA libraries were prepared by ligating specific indexed linkers to the DNA fragments. The ten resulting libraries were pooled in equal concentrations. DNA fragment size and concentration were measured using a 2100 Bioanalyzer (Agilent) on an Agilent DNA7500 chip according to the manufacturer's instructions.

Transposon-chromosome junctions enrichment was based on a methodology to isolate microsatellites DNA loci (5) with the following modifications. In a 0.2-ml PCR tube, 300 ng of the pooled DNA fragments were mixed with 25 µl of 2 x Hyb solution and incubated in a thermal cycler at 95°C for 5 min followed by a quick ramp down to 85°C. Then, 10 pmol of enrichment probe were added into the mixture. Hybridization of the probe to the DNA was achieved in a gradient thermal cycler through the following steps: a gradient step from 80°C to 60°C with a decrease of 0.2°C every 5s followed by a second annealing step at 60°C for 3 hours. After pull-down using Dynabeads M-280 streptavidin (Invitrogen), enriched DNA fragments were eluted with 30 µl of TLE (10 mM Tris pH 8; 0.1 mM EDTA) and snap cooled in ice for 2 min. Recovered DNA fragments were PCR amplified using primers TruSeq PCR for (5'-AATGATACGGCGACCACCGA) and TruSeq PCR rev (5'-CAAGCAGAAGACGGCATACGA) (Oligonucleotide sequences © 2007-2011 Illumina, Inc. All rights reserved. Derivative works created by Illumina customers are authorized for use with Illumina instruments and products only. All other uses are strictly prohibited.) and Phusion high-fidelity DNA polymerase (New England Biolabs). The cycling program proceeded as follows. After 30s of denaturation at 98°C, the following steps were performed for 12 cycles: denaturation at 98°C for 10 s, annealing at 60°C for 30 s and extension at 72°C for 30 s. A final extension was performed at 72°C for 5 min. Then, 15 µl of the PCR product were used for a second enrichment with the biotinylated probe. After elution in 50 µl of TLE, 21 µl of the DNA were used for a second PCR following the same procedure as above, but running for only 8 cycles. The PCR product was purified using the DNA clean and concentrator -5 kit (Zymo Research) and stored at -20°C.

Sequencing of the enriched DNA fragment library was performed at the Genome Services Laboratory at the Hudson Alpha Institute for Biotechnology, Huntsville, Alabama. Custom primer single-end sequencing (50 bp) was carried out on a HiSeq Flowcell v1.5 using a HiSeq2000 sequencer.

Analysis of sequencing data

Sequence reads from the Illumina FASTQ files were mapped to the genome of *M. maripaludis* S2 using the Burrows-Wheeler Aligner (BWA) (18). Of these, 92% mapped to the genome. Of the remaining 8%, about 40% corresponded to fragments of the *M. maripaludis* S2 genome, but the quality of the sequences was poor with internal mismatches or a low coverage. The remainder were either short, possessed numerous runs of single nucleotide bases, or included numbers of unidentified bases and were assumed to be sequencing errors.

Subsequent analysis was performed using in-house software developed specifically for this work. First, the BWA output was converted into a list of unique insertions identified by their position in the genome and the orientation of the inserted transposon. Insertions at the same position but in opposite orientations were both considered unique because they represented different insertion events. Next, the essentiality of individual genes was assessed by a probabilistic approach and a sliding window method.

In the probabilistic approach, the number of unique insertions in each gene was compared to the probability that the same or lower number of insertions would be found in a gene of the same length in the absence of selection and under the assumption that the insertions were equally likely to occur at any site in the genome. Based on these assumptions and considering the insertions as "rare events", the number of unique insertions is approximately Poisson-distributed, with the expected count $\lambda = \Phi \frac{l}{L}$, where l is the length of the gene at hand, L is the length of the genome, and Φ is the number of unique insertions in the whole genome in the absence of selection. Difficulties arise in estimating the parameter Φ because selection against growth of

some mutants eliminates a number of insertions from the dataset. While the number of unique insertions found can be used as a lower bound estimate for Φ, this leads to an overestimation the *p*-values and conservative assessments of the statistical significance of gene essentiality. Nevertheless, gene rankings by the *p*-values can be used to obtain an approximate relative assessment of essentiality of individual genes. An additional drawback of this approach is that short genes are classified as nonessential even if they contain no insertions because the expected counts λ are too low to provide sufficient confidence to reclassify the gene.

The sliding-window assessment of gene essentiality was designed to partially overcome the drawbacks of the probabilistic approach. To avoid comparing counts of insertions in segments of variable length (e.g., genes), the number of unique insertions N_w in each window of a fixed size *l* was determined starting at position *w* in the chromosome. Using the insertion counts N_w, an essentiality index was assigned to any segment composed of one or more overlapping windows as the maximum N_w among all windows fully embedded in that segment. Formally, the essentiality index for a segment between chromosomal positions *x* and *y* (measured in base pairs) is

$$S_{x,y} = \max_{x \le w \le y-l+1} N_w$$

When assigning essentiality index S_g to a gene g, which can be smaller than the window size, the formula is

$$S_g = \min_{x \le \alpha, y \ge \beta} S_{x,y}$$

where α and β are coordinates of the left and right ends of the gene, respectively. In other words, for genes larger than the sliding window the essentiality index S_g is the largest N_w among all windows embedded in that gene. For genes smaller than the window size, S_g is the smallest N_w among all windows that fully encompass the gene at hand. This software also allows adjusting

the position of the start and end of each gene (values α and β) by a fixed number of nucleotides or by a fixed percentage of the gene length. In this work, the first 5% of the gene length (at the 5' end) and the last 20% of the gene length (at the 3' end) were not used because it is possible that many start codons were not annotated properly or that insertions close to the end of a gene had little effect on functionality. Thus, mutation in these portions of the gene might not affect viability even for essential genes (14). In fact, while the results were similar using the complete genes, for some possibly essential genes, such as RNA polymerase genes (MMP1364), tRNA synthetase genes (MMP0326, MMP1496 and MMP1614) and replication related genes (MMP0026) among others, could not be assigned as possibly essential without excluding the extremes of the gene sequences. A window size of 800 bp was used, which is close to the *M. maripaludis* average gene size (873 bp), and the sliding window was moved in steps of 100 bp. The software is available from the authors upon request.

Both methods generated similar results, but the sliding-window method was preferred for the following reasons. First, for many small genes, the probabilistic method was unable to provide enough confidence to call the genes essential. While absence of insertions in a single short gene still did not provide evidence for essentiality in the sliding-window method, clusters of two or more small possibly essential genes were detected. For instance, the sliding-window method detected essentiality of a potential operon of 14 genes (MMP1408 – 1421) which encoded ribosomal proteins. However, small genes which are known to be essential but surrounded by genes with insertions probably will not be classified as essential because the size of the windows is larger than the size of the target gene and will cover the insertions of the surrounding genes. Thus, small genes cannot be confidently assigned by this methodology. Second, the probabilistic method depends on the assumption that the insertions of the Tn5

transposon are random in the genome of *M. maripaludis*, but some hotspots were observed in the different libraries. Finally, the sliding windows can be used to explore the complete genome of *M. maripaludis* S2, searching for unknown possibly essential genes or possibly essential chromosomal regions.

Real-time PCR for genome copy number quantification

The real-time PCR approach was based on previous methodology used to determine genome copy numbers in Archaea (11) with the following modifications. DNA fragments of ~1Kb were PCR amplified from the gene *cdhA* using primers 1kb cdhA F (5'-TTGATGACGATTTTTGGAAGA) and 1kb cdhA R (5'-CGAGTATAAGTGCACCATCG) and from the Tn5 transposon using primers 1kb Tn5 F (5'-GGTTTTGTATTTCCGGTAGTAATC) and 1kb Tn5 R (5'-GCGATTAAATTCCAACATGG) and Phusion high-fidelity DNA polymerase (New England Biolabs). The cycling program proceeded as follows. After 30s of denaturation at 98°C, the following steps were performed for 30 cycles: denaturation at 98°C for 10 s, annealing at 52°C for 30 s and extension at 72°C for 40 s. A final extension was performed at 72°C for 5 min. The fragments were cleaned by using DNA Clean and Concentrator -5 Kit (Zymo Research). The DNA concentrations were determined by spectrophotometry using a Nanodrop 1000 (Thermo Scientific). Each of the standard fragments was serially diluted to a range between 10^4 to 10^8 molecules, and 1 µl of the dilutions were used for real-time PCR analysis. The real-time PCR amplified fragments were internal to the ~1kb standards and were ~130 bp in size. The primers used to generate these internal fragments were Internal cdhA F (5'-CAGGTCAGGAAGGAAACTCTCTTG), Internal cdhA R (5'-CTACAAGTGGTTCGTCACCATCTC), Internal Tn5 F (5'-CAGCCAGTTTAGTCTGACCATCTC), and Internal Tn5 R (5'-

GGCAATCAGGTGCGACAATCT ATC). Genomic DNAs from library 1 were diluted to 5 ng/µl and were used as templates with the same internal primers used for the standards. Real-time PCR assays were done in 10 µl final volumes, using 1 µl of template, 1 µM of each primer and the SYBR GreenER qPCR SuperMix Universal (Invitrogen). The PCR reaction conditions proceeded as follows. After 5 min of denaturation at 95°C, the following steps were performed for 40 cycles: denaturation at 95°C for 30 s, annealing at 60°C for 1 min, and extension at 72°C for 1 min. The real-time PCRs were performed in Mx3005P Real-Time PCR System (Stratagene). For each sample, the number of cycles before reaching the fluorescence threshold was determined (Ct value). Using the standard Ct values, a standard curve was constructed to calculate the gene copy numbers in the samples. By dividing the copy number of the Tn5 internal fragment over the copy number of *cdhA* internal fragment, the fraction of genomes that contain the transposon insertions can be calculated.

Construction of a Δ*polB*::*pac* mutant

The gene *polB* in *M. maripaludis* S2 genome was deleted and replaced by the *pac* cassette as shown in Figure S3-3 by using pIJ03FS. To construct the plasmid pIJ03FS, an 822 bp DNA region upstream *polB* was amplified by PCR using the primers polB up F (5'-ATATATCTCGAGCGTCAGCAATCCCTCTTAAA) and polB up R (5'-ATATATGGATCCTGAAGTCTGGAAATCGGGTA), which introduced the *Xho*I and *Bam*HI restriction sites respectively (underlined) into the PCR amplicon. The product was PCR amplified using Phusion high-fidelity DNA polymerase (New England Biolabs). The cycling program proceeded as follows. After 1 min of denaturation at 98°C, the following steps were performed for 30 cycles: denaturation at 98°C for 30 s, annealing at 61°C for 30 s and extension at 72°C for 30 s. A final extension was performed at 72 °C for 7 min. The PCR product was

purified using the DNA clean and concentrator -5 kit (Zymo Research), digested with *Xho*I and *Bam*HI, and gel purified using the QIAquick Gel Extraction Kit (Qiagen). The recovered product was ligated into the multi cloning site 1 (MCS1) upstream of the *pac* cassette, which confers puromycin resistance, in the plasmid pIJ03 (1). Similarly, a 961 bp region immediately downstream of *polB* was amplified using the primers polB down F (5'-ATATATGGTACCGAGCCTTTCAAATTCCATCAA) and polB down R (5'-ATATATGCTAGCCCGCCTTCATCCATTATATCT), which introduced the underlined *Kpn*I and *Nhe*I sites, respectively, into the amplicon. The PCR product was amplified by PCR following the same conditions as the upstream region, cleaned, digested, gel purified and ligated into MCS2 downstream of the *pac* cassette in the plasmid pIJ03. The plasmid pIJ03FS was linearized by digestion with *Sal*I, purified and transformed into *M. maripaludis* S2 cells using polyethylene glycol mediated transformation (1). Transformed cells were serially diluted, and 500 µl of the dilutions were spread in serum-bottle agar slants. Random puromycin-resistant colonies were picked and transferred to stoppered culture tubes containing 5 ml of McCv medium plus puromycin. The culture tubes containing the isolates were pressurized to 275 kPa and incubated at 37°C. After growth, cultures were serially diluted and replated in serum-bottle agar slants with McCv medium plus puromycin. Colonies were then picked into broth with the same composition, and 1 ml of the grown culture was used to prepare frozen stocks. The rest was used to extract DNA using the ZR fungal/bacterial DNA miniprep (Zymo Research) for further characterization.

Verification of the genotype of the mutant was determined through PCR using two sets of primers (Figure S3-4). Internal primers for the *polB* gene were: IndF (5'-CCAATCCCGCCTTACTACTT) and IndR (5'- ATGAGCAGTAGCCAAATGGT). The second

set of primers amplified the junction between *M. maripaludis* genomic DNA and the *pac* cassette: JudF (5'- CCCAGGTATTGCACATTTTT) and JudR (5'- CGGTCATGAGAATCACTCCT). The products were PCR amplified using Phusion high-fidelity DNA polymerase (New England Biolabs). The cycling program for both set of primers proceeded as follows. After 1 min of denaturation at 98°C, the following steps were performed for 30 cycles: denaturation at 98°C for 10 s, annealing at 63°C for 30 s and extension at 72°C for 40 s. A final extension was performed at 72 °C for 7 min. The PCR products were loaded and run in a 1% agarose gel and stained with ethidium bromide. Control experiments indicated that the PCR reactions could detect the wild-type allele in a thousand-fold excess of the mutant allele (Figure S3-4C).

Second verification of the genotype of the mutant was achieved by Southern blot (Figure S3-4). An 822 bp probe, which corresponded to the DNA region upstream of the *polB* gene that had been used in the construction of the plasmid pIJ03FS (see above), was PCR amplified as described above and labeled with digoxigenin (DIG)-dUTP using the High Prime DNA Labeling and Detection Starter Kit according to the manufacturer instructions (Roche). Genomic DNA of *M. maripaludis* S2 and the mutant Δ*polB*::*pac* were digested overnight using the restriction enzyme EcoRI-HF (New England Biolabs) and run in a 0.7% agarose gel. Gel preparation and transfer to a Nytran nylon membrane (Whatman) was achieved using the TurboBlotter system according to the manufacturer's instructions (Whatman). DNA was UV-crosslinked to the nylon membrane by exposing the membrane to UV light at an intensity of 120 mJ/cm^2 for 1 minute; this step was repeated 3 times. Hybridization and immunological detection was achieved using the DIG High Prime DNA Labeling and Detection Starter Kit according to the manufacturer instructions (Roche) with a few modifications. Membranes were prehybridized for two hours in

20 ml of DIG Easy Hyb buffer in roller bottles. Hybridization was done at 41.1°C in 20 ml of probe/hybridization mix (probe concentration 25 ng/ml) and incubated overnight. For every step of detection, the volume of the respective solution or buffer was increased to 100 ml, with the exception of the color substrate solution which was increased to 60 ml. Incubation with Blocking solution was increased to 1 hour, and 3 washes of 15 minutes were done with washing buffer after incubation with the antibody. After 8 hours of color development, the reaction was stopped using 100 ml of TE buffer. To determine the phenotype of the mutant, $2x10^7$ mutant or wild type cells were inoculated in 5 mL of McCV and McA (McN + acetate) media reduced with 3 mM cysteine and with the addition of 3 mM sodium sulfide. The cultures were grown in 28-ml Balch tubes pressurized to 275 kPa with H_2/CO_2 for over 70 hours in the absence of antibiotic.

Additional Results and Discussion

Energy metabolism. In this category, 53 genes were possibly essential. These included, the 9 genes encoding the ATP synthase (MMP1038-1046) and 22 genes directly involved in methanogenesis. Six genes encoding the molybdenum-containing formylmethanofuran dehydrogenase were nonessential, but every gene encoding the tungsten-containing enzyme was possibly essential. Although both W and Mo were abundant in the medium, the tungsten form appeared to be preferred (13). Methanococci possess both the Mtd and Hmd pathways for the reduction of methenyltetrahydromethanopterin (9). Because of this redundancy, the genes encoding both pathways were nonessential. *M. maripaludis* possesses six nickel hydrogenases. Only two, Eha and Vhu, were possibly essential under these growth conditions, which agreed closely with expectations. Eha is encoded by 20 genes clustered together (MMP1448-1467) in which 15 genes were possibly essential. It is required to replenish intermediates in

methanogenesis (19). Four of the five genes (MMP1693-1696) encoding Vhu were possibly essential. Vhu provides electrons for the last step of methanogenesis during growth on H_2 (19).

The *mcr* genes encoding the methylcoenzyme M reductase, which catalyzes the last step in methanogenesis, are arranged in one operon (MMP1555-1559), all genes of which were possibly essential. In most methanogens, the *mcr* operon is flanked by a conserved ORF with homology to radical S-adenosyl-L-methionine proteins. In spite of its wide conservation, this gene (MMP1554) was nonessential. *M. maripaludis* encodes two homologs, MMP0620 and MMP1630, to component A2 (*atwA*), which is required for activation of the methylcoenzyme M reductase (15). Both were possibly essential. However, whether or not both play important roles in methylreductase activation is not known.

Transcription. In *M. maripaludis*, 12 genes encode RNA polymerase, of which 11 were possibly essential (MMP0248, 0261, 0440, 1322, 1326, 1327, 1360-1364). One of these, MMP0248, was previously annotated as a RNAP-related protein, but reexamination of the annotation indicated that it likely encoded subunit P of RNAP. The remaining gene, *rpoF* (MMP0092) was unassigned. These results are in agreement with an *in vitro* reconstruction of the RNAP of *M. jannaschii,* where a minimal RNAP complex possessed only subunits A1, A2, B1, B2, D, L, N, and P, and the incorporation of subunits E and F had no detectable effect on activity (30). Similarly, *rpoE* and *rpoF* are nonessential in *T. kodakarensis* (12)

In archaea, two transcription factors, TFB (MMP0041) and TBP (MMP0257), are required for cell-free transcription (10, 22), and both were possibly essential. The gene MMP0036, which encodes transcription factor TFE, was also classified as possibly essential and has been previously shown to facilitate transcription initiation in archaea (1, 8). Four transcriptional elongation factors are known in archaea (7). NusA (MMP1366) and Spt5, the

methanococcal homolog of the bacterial elongation factor NusG (MMP1434), were possibly essential. Spt4 (MMP0441) was classified as unassigned, and TFS (MMP1429) was nonessential. Lastly, MMP1015, which is homologous to the eukaryotic CBF/NF-Y, was classified as unassigned.

Most genes with homology to transcriptional regulators were nonessential, but there were exceptions. NrpR (MMP0607) is a transcriptional repressor which regulates nitrogen metabolism and was possibly essential (20). In addition, the methanococcal homolog to the Lon protease (MMP1186) was possibly essential. In bacteria and eukaryotes, this protease regulates gene expression by selectively degrading transcriptional regulators (29). Possibly, it plays a similar role in methanococci.

Translation. Many of the 124 genes with a function in translation were possibly essential (Fig. S2). With the exception of cysteinyl-tRNA synthetase, the genes encoding the aminoacyl-tRNA synthetases were possibly essential. Like other methanogenic archaea, *M. maripaludis* possesses two different pathways to charge $tRNA^{Cys}$. The first one employs the canonical cysteinyl-tRNA synthetase (MMP1060), which is nonessential in methanococci (27). In the second pathway, cysteine is biosynthesized *de novo* by phosphoseryl-tRNA synthetase (MMP0688) and Sep-tRNA:Cys:tRNA synthase (MMP1240), which was possibly essential (26). Three other aminoacyl-tRNA synthetase-related genes, MMP0377, 0693, and 0816, were nonessential, suggesting that they play less fundamental roles.

A total of 27 genes encoding the large subunit and 21 genes encoding the small subunit ribosomal proteins were possibly essential. Many of these possibly essential genes were clustered in putative operons. Although free of insertions, the gene MMP0151 could not be reliably classified due to its small size. Even after careful examination of the location of

insertions, ten genes encoding ribosomal proteins remained unassigned and one gene was nonessential. These results are not uncommon. In *Salmonella typhi*, eight genes that encode ribosomal proteins were nonessential (16). Presumably, these proteins might be conditionally expressed or redundant with other proteins.

In *M. maripaludis*, eight genes are annotated as translation initiators factors (2). Of these, MMP0603 (aIF-1A), 0297 and 1208 (β and γ subunits of aIF2), 0284 (aIF2/5B), 0952 (aIF5A), and 0061 (aIF6) were possibly essential. MMP1707, which encodes the α-subunit of aIF2, was unassigned, and MMP1618 (aIF2B, subunit 1) was nonessential. Four translation elongation factors are also annotated. MMP1370, 1401 and 1369 encode the α and β subunits of elongation factor 1 and elongation factor 2, respectively, and were possibly essential. MMP0738, which encodes domain 2 of the elongation factor TU was nonessential. Finally, MMP1336, which encodes the special archaeal translation factor for selenocysteine insertion (SelB), was possibly essential**.**

The genome of *M. maripaludis* contains multiple copies of the rRNA genes, and none of them were essential. MMP0606 encodes a "23S rRNA methyltransferase related protein" and was possibly essential. Forty-one other genes encode structural RNAs. The three genes encoding the RNA component of RNAse P and the RNA components of both the bacterial and eukaryotic type signal recognition particles were initially classified as nonessential. However, upon careful examination, these small genes contained 0-1 insertions and, hence, were possibly essential. Although the 38 tRNA genes are too small for determination of their essentiality by this method as no more than two insertions would be expected by chance, most tRNAs possessed fewer insertions. However, some tRNA genes possessed more insertions than expected. Methanococci contain two copies of asparagine tRNA with the same anticodon (GTC), and they contained 2 or

4 insertions, respectively (averages of both libraries). Presumably, either one can support growth. Similarly, two of the four methionine tRNA genes possessed 7-13 insertions and were likely nonessential.

Anabolism. Many of the biosynthetic genes were possibly essential even in rich medium (Fig. S2). Methanococci are facultative lithotrophs that only poorly assimilate organic compounds (31). For instance, methanococci do not take up sugars, and nearly all the genes required for gluconeogenesis from pyruvate to fructose-6-phosphate were possibly essential. The exception was MMP0293, which encodes fructose bisphosphate aldolase. Presumably, there is another aldolase with similar activity. Likewise, stable isotope labeling previously indicated that the nonoxidative pentose phosphate pathway was the major route of pentose biosynthesis in *M. maripaludis* (26). Consistent with this hypothesis, the genes encoding this pathway, including MMP1113-1115, 1189, and 1308, were possibly essential. In contrast, the methanococcal homolog (MMP1270) to the genes encoding the ribulose monophosphate pathway, which is used for pentose biosynthesis by *T. kodakarensis* (21, 28), was nonessential, and this pathway does not appear to play a major role in methanococcal pentose biosynthesis. Similarly, eight of the 10 genes identified for lipid biosynthesis were possibly essential. Because the rich medium did not contain the isoprenoid precursors for the archaeal lipids, these results were not unexpected. Similarly, while methanococci readily assimilate nucleobases, they do not take up the nucleosides and nucleotides likely to be abundant in rich medium, and many of the genes for nucleotide biosynthesis were possibly essential (3). *M. maripaludis* poorly assimilates amino acids (31), and many of the genes encoding amino acid biosynthetic pathways were possibly essential (Dataset S4). Similarly, even though thiamine and nicotinamide were present in the medium, nearly every gene for their biosynthesis was possibly essential. Presumably,

methanococci rely upon their biosynthetic capacity for these vitamins (17). In contrast, the genes identified in biotin and riboflavin biosynthesis were nonessential. Presumably, these coenzymes are taken up from the medium.

Methanogenesis requires unusual coenzymes for which the biosynthetic pathways are only partially known (6). The identity of only a few of these genes has been confirmed by genetic methods. Five of the seven genes for the biosynthesis of the deazaflavin coenzyme F_{420} were possibly essential: MMP0056, 0404, 0876, 0915, and 0937. The four genes known for coenzyme B biosynthesis (MMP0153, 0381, 0880 and 1480) were all possibly essential (4). Two genes, MMP0034 and 0279, of the three known for H_4MPT biosynthesis were possibly essential. The only known gene involved in methanofuran biosynthesis, MMP0131, was possibly essential. The essentiality of these genes supports their roles in coenzyme biosynthesis. Because coenzyme M was added to the medium, the genes for its biosynthesis were unassigned or nonessential.

References

1. **Bell, S. D., A. B. Brinkman, J. van der Oost, and S. P. Jackson.** 2001. The archaeal TFIIEalpha homologue facilitates transcription initiation by enhancing TATA-box recognition. EMBO Rep **2:**133-8.
2. **Benelli, D., and P. Londei.** 2011. Translation initiation in Archaea: conserved and domain-specific features. Biochem Soc Trans **39:**89-93.
3. **Bowen, T. L., and W. B. Whitman.** 1987. Incorporation of Exogenous Purines and Pyrimidines by *Methanococcus voltae* and Isolation of Analog-Resistant Mutants. Appl Environ Microbiol **53:**1822-6.

4. **Drevland, R. M., Y. Jia, D. R. Palmer, and D. E. Graham.** 2008. Methanogen homoaconitase catalyzes both hydrolyase reactions in coenzyme B biosynthesis. J Biol Chem **283:**28888-96.

5. **Glenn, T. C., and N. A. Schable.** 2005. Isolating microsatellite DNA loci. Methods Enzymol **395:**202-22.

6. **Graham, D. E., and R. H. White.** 2002. Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics. Nat Prod Rep **19:**133-47.

7. **Grohmann, D., and F. Werner.** 2011. Recent advances in the understanding of archaeal transcription. Curr Opin Microbiol **14:**328-34.

8. **Hanzelka, B. L., T. J. Darcy, and J. N. Reeve.** 2001. TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFIIEalpha. J Bacteriol **183:**1813-8.

9. **Hendrickson, E. L., and J. A. Leigh.** 2008. Roles of coenzyme F420-reducing hydrogenases and hydrogen- and F420-dependent methylenetetrahydromethanopterin dehydrogenases in reduction of F420 and production of hydrogen during methanogenesis. J Bacteriol **190:**4818-21.

10. **Hethke, C., A. C. Geerling, W. Hausner, W. M. de Vos, and M. Thomm.** 1996. A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus*. Nucleic Acids Res **24:**2369-76.

11. **Hildenbrand, C., T. Stock, C. Lange, M. Rother, and J. Soppa.** 2011. Genome copy numbers and gene conversion in methanogenic archaea. J Bacteriol **193:**734-43.

12. **Hirata, A., T. Kanai, T. J. Santangelo, M. Tajiri, K. Manabe, J. N. Reeve, T. Imanaka, and K. S. Murakami.** 2008. Archaeal RNA polymerase subunits E and F are

not required for transcription in vitro, but a *Thermococcus kodakarensis* mutant lacking subunit F is temperature-sensitive. Mol Microbiol **70:**623-33.

13. **Hochheimer, A., R. Hedderich, and R. K. Thauer.** 1998. The formylmethanofuran dehydrogenase isoenzymes in Methanobacterium wolfei and *Methanobacterium thermoautotrophicum*: induction of the molybdenum isoenzyme by molybdate and constitutive synthesis of the tungsten isoenzyme. Arch Microbiol **170:**389-93.
14. **Hutchison, C. A., S. N. Peterson, S. R. Gill, R. T. Cline, O. White, C. M. Fraser, H. O. Smith, and J. C. Venter.** 1999. Global transposon mutagenesis and a minimal Mycoplasma genome. Science **286:**2165-9.
15. **Kuhner, C. H., B. D. Lindenbach, and R. S. Wolfe.** 1993. Component A2 of methylcoenzyme M reductase system from *Methanobacterium thermoautotrophicum* delta H: nucleotide sequence and functional expression by *Escherichia coli*. J Bacteriol **175:**3195-203.
16. **Langridge, G. C., M. D. Phan, D. J. Turner, T. T. Perkins, L. Parts, J. Haase, I. Charles, D. J. Maskell, S. E. Peters, G. Dougan, J. Wain, J. Parkhill, and A. K. Turner.** 2009. Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. Genome Res **19:**2308-16.
17. **Leigh, J. A.** 1983. Levels of water-soluble vitamins in methanogenic and non-methanogenic bacteria. Appl Environ Microbiol **45:**800-3.
18. **Li, H., and R. Durbin.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:**1754-60.

19. **Lie, T. J., K. C. Costa, B. Lupa, S. Korpole, W. B. Whitman, and J. A. Leigh.** 2012. Essential anaplerotic role for the energy-converting hydrogenase Eha in hydrogenotrophic methanogenesis. Proc Natl Acad Sci U S A.

20. **Lie, T. J., and J. A. Leigh.** 2003. A novel repressor of nif and glnA expression in the methanogenic archaeon *Methanococcus maripaludis*. Mol Microbiol **47:**235-46.

21. **Orita, I., T. Sato, H. Yurimoto, N. Kato, H. Atomi, T. Imanaka, and Y. Sakai.** 2006. The ribulose monophosphate pathway substitutes for the missing pentose phosphate pathway in the archaeon *Thermococcus kodakaraensis.* J Bacteriol **188:**4698-704.

22. **Ouhammouch, M.** 2004. Transcriptional regulation in Archaea. Curr Opin Genet Dev **14:**133-8.

23. **Porat, I., and W. B. Whitman.** 2009. Tryptophan auxotrophs were obtained by random transposon insertions in the *Methanococcus maripaludis* tryptophan operon. FEMS Microbiol Lett **297:**250-4.

24. **Rozen, S., and H. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol **132:**365-86.

25. **Sarmiento, F. B., J. A. Leigh, and W. B. Whitman.** 2011. Genetic systems for hydrogenotrophic methanogens. Methods Enzymol **494:**43-73.

26. **Sauerwald, A., W. Zhu, T. A. Major, H. Roy, S. Palioura, D. Jahn, W. B. Whitman, J. R. Yates, 3rd, M. Ibba, and D. Soll.** 2005. RNA-dependent cysteine biosynthesis in archaea. Science **307:**1969-72.

27. **Stathopoulos, C., W. Kim, T. Li, I. Anderson, B. Deutsch, S. Palioura, W. Whitman, and D. Soll.** 2001. Cysteinyl-tRNA synthetase is not essential for viability of the archaeon *Methanococcus maripaludis.* Proc Natl Acad Sci U S A **98:**14292-7.

28. **Tumbula, D. L., Q. Teng, M. G. Bartlett, and W. B. Whitman.** 1997. Ribose biosynthesis and evidence for an alternative first step in the common aromatic amino acid pathway in *Methanococcus maripaludis*. J Bacteriol **179:**6010-3.

29. **Van Melderen, L., and A. Aertsen.** 2009. Regulation and quality control by Lon-dependent proteolysis. Res Microbiol **160:**645-51.

30. **Werner, F., and R. O. Weinzierl.** 2002. A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. Mol Cell **10:**635-46.

31. **Whitman, W. B., S. Sohn, S. Kuk, and R. Xing.** 1987. Role of Amino Acids and Vitamins in Nutrition of Mesophilic *Methanococcus* spp. Appl Environ Microbiol **53:**2373-8.

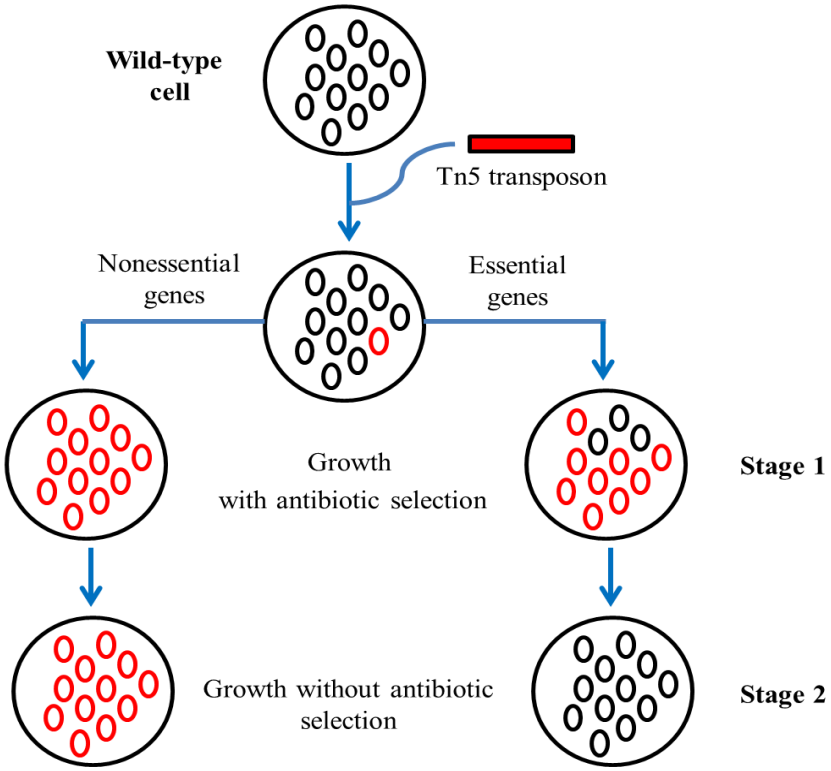

Figure S3-1. Schematic representation of gene conversion and selection in polyploid methanococcal cells. The transposition event introduces the transposon into one genome in each cell. In stage one, the cells grew as colonies on solid medium to limit competition for about 20 generations. The presence of the antibiotic puromycin selected for the maintenance of the genomes containing the transposon. After 20 generations, the wild type allele of nonessential genes is expected to be eliminated by gene conversion, and only the mutated allele expected to remain. For essential genes, gene conversion is unable to eliminate the wild type allele, and both alleles are retained in different proportions. In stage two, the cells are grown in broth in the absence of antibiotic selection for 7 (T1) and 14 (T2) generations. While only the mutated allele remains for nonessential genes, the abundance of the mutants can fluctuate because of competition with other mutants and wild-type cells. However, for essential genes the mutated allele can now be lost by gene conversion to restore the wild-type cells.

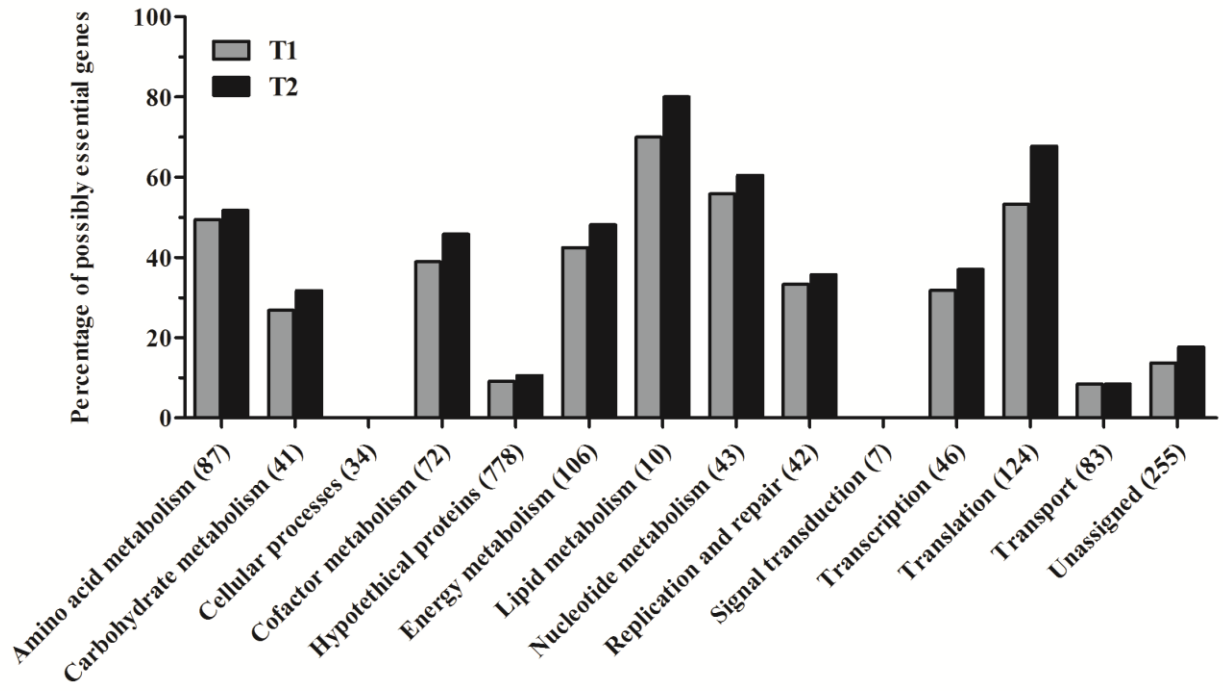

Figure S3-2. Functional categories of possibly essential protein-coding genes. Genes were assigned to the functional classes updated from the *M. maripaludis* S2 genome annotation. The percentage of possibly essential genes in each category was calculated as the (number of possibly essential genes)/(total number of genes) x100. The total number of genes in each category is shown in parentheses.

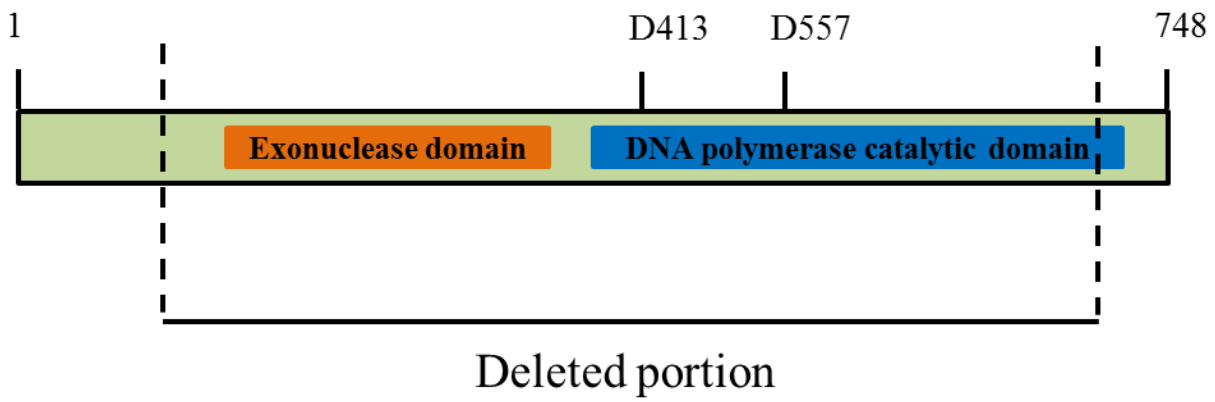

Figure S3-3. Schematic representation of the methanococcal *polB* gene in the Δ*polB*::*pac* mutant strain S123. Approximately 1795 bp of the *polB* gene were deleted (598 amino acids). The exonuclease domain and the polymerase domain are represented in orange and blue, respectively. The essential aspartyl residues for the polymerase function are indicated.

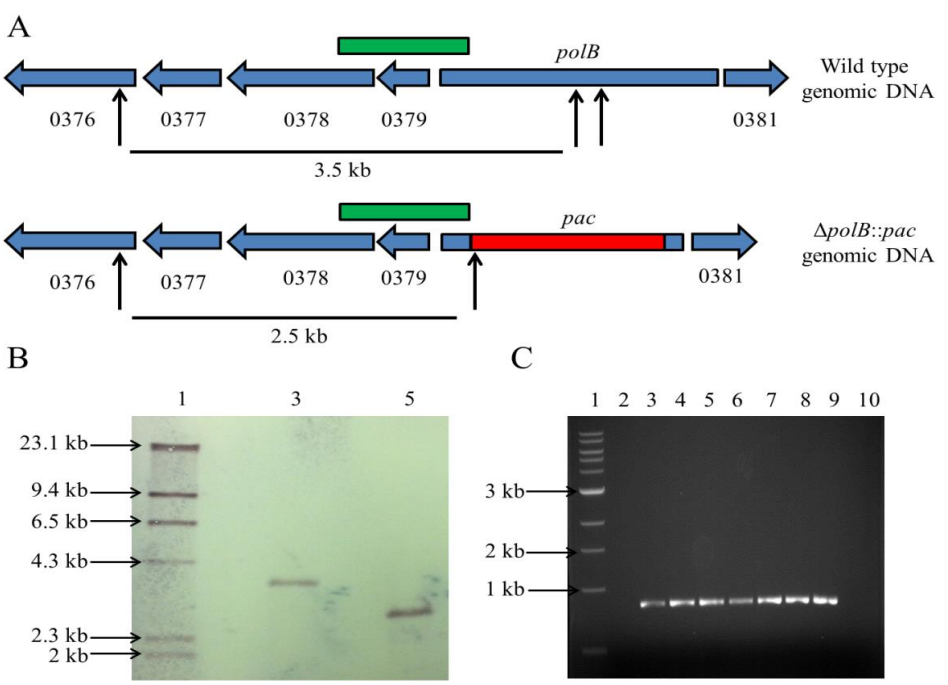

Figure S3-4**.** Verification of the genotype of the mutant S123 Δ*polB*::*pac*. (A) Graphic representation of the Southern blot procedure for verification of the deletion of the gene *polB*. Numbers indicate the MMP identification, black arrows indicate restriction sites for the enzyme EcoRI-HF, black line indicate the digested fragment, and green rectangles represent the binding site of the DNA probe. (B) Genotypic of the Δ*polB*::*pac* mutant by Southern blot. Lane 1, DNA Molecular Weight Marker II, DIG-labeled (Roche); Lane 3, Southern blot for genomic DNA of *M. maripaludis* wild type strain S2; Lane 5, Southern blot for genomic DNA of the Δ*polB*::*pac* mutant strain 123 (C) PCR amplification of an internal portion of *polB* using primers IndF and IndR for a mixture of different proportions of genomic DNA of the Δ*polB*::*pac* mutant strain S123 and wild type strain S2. Lane 1, Standard 1kb ladder (New England Biolab); Lane 2, 100% genomic DNA of the Δ*polB*::*pac* mutant strain S123; Lane 3, 99.9% genomic DNA of strain S123 and 0.1% of wild type strain S2; Lanes 4-9, 99.5, 99, 98, 95, 90, and 0% of genomic DNA of the Δ*polB*::*pac* mutant strain S123 and the balance genomic DNA of wild type strain S2. Lane 10, negative control without template.

Table S3-1. Number of sequences and insertions for each library

Library	Condition[a]	No. of reads mapped to genome	No. of unique insertions	Avg. no. of insertions per gene
Library 1	T0	3,952,536	59,079	33
	McCm (T1)	3,632,478	49,270	28
	McCm (T2)	2,592,650	44,619	25
	McN (T1)	1,542,104	23,447	13
	McN (T2)	2,911,505	13,849	8
Library 2	T0	3,318,618	28,724	16
	McCm (T1)	2,593,856	23,962	14
	McCm (T2)	4,222,958	27,458	16
	McN (T1)	1,908,608	17,314	10
	McN (T2)	4,312,878	15,734	9

[a]Each library was sequenced at T0 and after 7 (T1) and 14 generations (T2) of growth in rich (McCm) or minimal (McN) media without antibiotic selection.

APPENDIX B

CHAPTER 4 SUPPLEMENTARY INFORMATION: ALANINE PSEUDOAUXOTROPHY IN *METHANOCOCCUS MARIPALUDIS*

Methanococcus maripaludis S2 mutant libraries grown in minimal medium (McN) show a enrichment for mutants with insertions in the gene MMP1511 (2). Indeed, the abundance of reads in these gene, which encodes an alanine/sodium symporter, increased from 0.6% of the total reads in the T0 libraries to 85% in the T2 libraries following growth in minimal medium (Figure S4-1A). This result indicates that the inactivation of the symporter stimulated the growth rate in minimal medium. Indeed, the relative fitness of the mutant, calculated by the Malthusian parameter (5), was 1.5 times that of mutants of nonessential genes (Figure S4-1B). Because this effect was not observed in rich medium containing Casamino acids, it appears to depend upon the absence of amino acids.

One hypothesis for the basis of the growth stimulation of the mutation is as follows. M. maripaludis is a facultative autotroph and can make all of its amino acids for growth. The biosynthesis of alanine is achieved in one step by the transamination of pyruvate, which is derived from autotrophic fixation of CO2 into acetyl-CoA. However, in the absence of external alanine, the symporter encoded by the gene MMP1511 could allow alanine to escape from the cell, and deplete the internal pool of this amino acid. If this was the case, addition of alanine to the medium should enhance growth by preventing the loss of alanine. To test this hypothesis, *M. maripaludis* S2 wild type was grown in minimal medium in the presence and the absence of different concentrations of alanine (Figure S4-2A). The addition of alanine did enhance the growth, but *M. maripaludis* also could be using this amino acid as an alternative source of nitrogen or carbon. However, this is not consistent with the growth stimulation by 1 μM of alanine, which contributes with only 0.02% and 0.03% of the cellular carbon and nitrogen needed for a 5-ml culture (OD_{600}= 1), respectively. In addition, ammonia, the prefer nitrogen source has been already added to the medium, suggesting that nitrogen is not source of the

enhanced growth. However, the experiment was repeated in minimal medium plus acetate as an alternative carbon source (Figure S4-2B). Interestingly, even low concentrations of alanine still enhance the growth of *M. maripaludis* (Figure S4-2C). To check that is not a common amino acids effect, *M. maripaludis* was grown in the presence of proline, lysine and aspartate, which did not enhance *M. maripaludis* growth (Figure S4-2D).

Previous experiments in *Methanococcus voltae* demonstrated that this microorganism is capable of synthesizing leucine, isoleucine and acetate, but still these nutrients are required for growth, a phenomenon call pseudoauxotrophy (3). The observations here presented may not confirm that *M. maripaludis* is an alanine pseudoauxotroph, but certainly support this hypothesis. In addition, a different strain of *M. maripaludis* (JJ) has been observed to excrete alanine (Whitman, personal communication). The reasons why *M. maripaludis* excrete alanine are unknown, but it has been demonstrated that some bacteria excrete alanine in symbiotic processes with plants. For example, In the Rhizobia-leguminous plant symbiosis, *Rhizobium* sp., in its bacteroid form excrete alanine as a carrier of nitrogenase-generated ammonium to the plant (4). *M. maripaludis* is also a nitrogen-fixingr prokaryote (1). Possibly, it has a symbiotic relationship with native plants of its natural habitat in tidal marshes.

References

1. **Kessler, P. S., C. Blank, and J. A. Leigh.** 1998. The nif gene operon of the methanogenic archaeon *Methanococcus maripaludis*. J Bacteriol **180:**1504-11.
2. **Sarmiento, F., Mrázek, J and Whitman, WB.** 2013. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. Proc Natl Acad Sci U S A. doi: 10.1073/pnas.1220225110.

3. **Shieh, J., M. Mesbah, and W. B. Whitman.** 1988. Pseudoauxotrophy of *Methanococcus voltae* for acetate, leucine, and isoleucine. J Bacteriol **170:**4091-6.

4. **Waters, J. K., B. L. Hughes, 2nd, L. C. Purcell, K. O. Gerhardt, T. P. Mawhinney, and D. W. Emerich.** 1998. Alanine, not ammonia, is excreted from N2-fixing soybean nodule bacteroids. Proc Natl Acad Sci U S A **95:**12038-42.

5. **Woods, R. J., J. E. Barrick, T. F. Cooper, U. Shrestha, M. R. Kauth, and R. E. Lenski.** 2011. Second-order selection for evolvability in a large *Escherichia coli* population. Science **331:**1433-6.

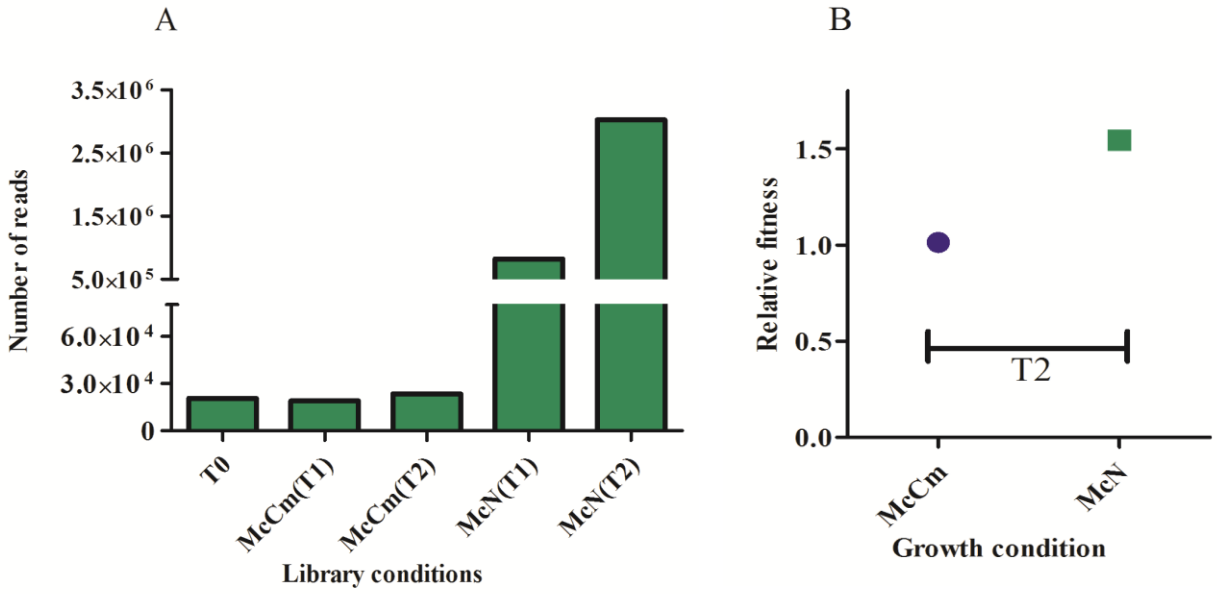

Figure S4-1. Enhanced growth of *M. maripaludis* MMP1511 mutant. (A) Number of reads of mutants with insertions in the gene MMP1511 when the library was sequenced after T0, or following growth in minimal (McN) or rich (McCm) media after 7 (T1) or 14 (T2) generations (1). (B) Relative fitness of the mutants in MMP1511 following growth in minimal (McN) or rich (McCm) media after 14 (T2) generations.

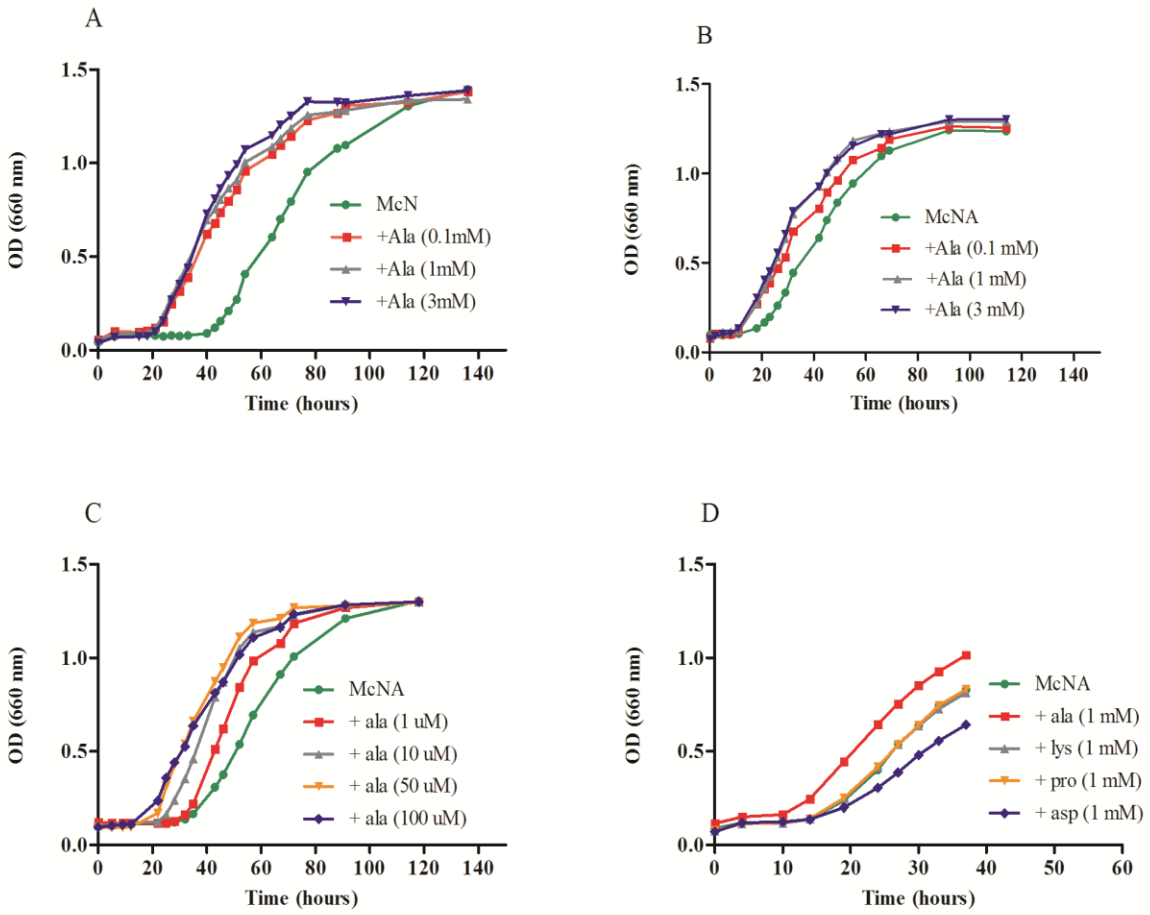

Figure S4-2. Effect of alanine and other amino acids on growth of *M. maripaludis* S2. (A) Stimulation of growth curves by alanine. Growth in McN supplemented with 0.1, 1 and 3 mM of alanine. (B) Stimulation of growth by alanine in the presence of acetate. Growht in minimal medium plus acetate (McNA) supplemented with 0.1, 1 and 3 mM of alanine. (C) The minimum concentrations of alanine required for growth stimulation. Growth in minimal medium plus acetate (McNA) supplemented with 1, 10, 50 and 100 µM of alanine. (D) Growth stimulation by other amino acids. Growth in minimal medium plus acetate (McNA) supplemented with 1mM of alanine, lysine, proline and asparagine.