Archaeal Plasmids

Warren L Gardner, *University of Georgia, Athens, Georgia, USA*

Archaeal plasmids are extrachromosomal elements isolated from organisms that belong to the domain of Archaea.

Secondary article

Article Contents

- Introduction
- Archaeal Plasmids
- Insertion Sequences and Transposons
- Archaeal Shuttle Vector Development and Use in Biotechnology

Introduction

On the basis of 16S and 18S ribosomal RNA sequences, Woese and Fox (1977) proposed the three domains of life known as Archaea, Bacteria, and Eukarya. The Archaea comprise two kingdoms: one is the Crenarchaeota, which consists of Sulfolobales and Thermoproteales; the second kingdom, the Euryarchaeota is composed of methanogens, extreme halophiles and the orders Thermoplasmales and Thermococcales. Plasmids have been found in members of both kingdoms. Isolation of these plasmids is important for fundamental understanding of archaeal replication mechanisms, identification of origins of replication, and construction of shuttle vectors for basic research and biotechnology.

Archaeal Plasmids

Euryarchaeota

Methanogens

Methanogens are strict anaerobes that use one-carbon compounds or acetate and hydrogen for the production of methane and energy. These anaerobes have been isolated from a variety of locations, including freshwater sediments and marine sediments, the intestinal tracts of cattle, humans and termites, and bioreactors and waste treatment facilities. Methanogenic isolates have been found to grow over a wide range of temperatures, including those that are psychrophilic and hyperthermophilic. This variety of sources has been important to the basic understanding of the diversity in methanogens.

The isolation of plasmids from methanogens began with the description of pMP1 by Thomm *et al*. (1983). pMP1 is a 7.0 kilobase (kb) cryptic plasmid purified from a coccoid methanogen, strain PL-12/M, which was isolated from a fumarole located off of Vulcano Island, Italy. PL-12/M was the only isolate out of 15 isolates to contain a plasmid. The identification of pMP1, which was never sequenced, confirmed the existence of plasmids in methanogens. Later that year, pME2001 was described from the thermophile *Methanobacterium thermoautotrophicum* strain Marburg ('*Methanothermobacter marburgensis* strain Marburg' (note: names presented in quotes are those that have not been validly published)). Isolation of pME2001 refuted a proposal that plasmids could only be purified from newly isolated methanogens and not from strains that had been maintained in the laboratory. Different multimeric forms of this plasmid, ranging from monomers to hexamers, were identified by electron microscopy. In an effort to construct a cloning vector for a *Methanobacterium* species, pME2001 was sequenced (see **Table 1**). *Methanobacterium thermoautotrophicum* ΔH ('*Methanothermobacter thermautotrophicus*' strain ΔH), a close relative to strain Marburg, does not contain plasmids. If the host range of pME2001 is broad enough, vectors derived from it could be used to study basic physiology within both of these well-studied methanogens. To date, vectors based on pME2001 have not been constructed.

Modular structures were found among plasmids isolated from *Methanobacterium* species. *Methanobacterium thermoautotrophicum* strain ZH3 and *Methanobacterium thermoautotrophicum* strain Marburg are closely related, with 99.9% similarity between their 16S rRNA sequences. Each strain also contains a plasmid. These plasmids, pME2200 from strain ZH3 and pME2001, from strain Marburg, were homologous but not identical. PME2200 and pME2001 share a common 3.8 kb region with 99.9% sequence similarity. pME2200 contains three additional regions when compared with pME2001. These 183 bp, 364 bp and 1058 bp regions did not contain open reading frames (ORFs), and were considered nonessential for plasmid replication. Both pME2001 and pME2200 were cryptic plasmids that may be coancestoral rather than derived from an individual plasmid.

Another example of modular structure comes from three plasmids identified in *Methanobacterium thermoformicicum* strains. Two of the plasmids, pFZ1 and pFZ2, from strains Z-245 and FTF, respectively, appeared identical by hybridizations and restriction endonuclease digestions. The third plasmid, from *M. thermoformicicum* strain THF, pFV1 was larger and possessed similarity to pFZ1 and pFZ2. Upon sequencing, pFV1 and pFZ1 were found to contain an 8.2 kb region with greater than 91% sequence identity. In addition, pFV1 possessed a restriction and modification system composed of the DNA methyltransferase *Mth*TIM and the restriction endonuclease *Mth*TIR. This type II deoxyriboendonuclease cleaved at GGCC sequences after the second guanosine residue. *Mth*TIM

Table 1 Sequenced archaeal plasmids

Host(s)[a]	Plasmid	Size (bp)	Accession number
Methanosarcina acetivorans strain C2A*	pC2A	5467	U78295
Methanobacterium thermoautotrophicum strain Marburg*	pME2001	4439	X17205
Methanobacterium thermoformicicum strain THF*	pFV1	13 513	X68366
Methanobacterium thermoformicicum strain Z-245*	pFZ1	11 014	X68367
Methanococcus jannaschii strain JAL1*	pURB800/MJECL	58 407	L77118
Methanococcus jannaschii strain JAL1*	pURB801/MJECS	16 550	L77119
Methanococcus maripaludis strain C5*	pURB500	8285	U47023
Halobacterium strain GN101*	pHGN1	1765	X16460
Halobacterium strain GRB*/ *Halobacterium halobium* strain R1	pGRB1	1781	X52610
Halobacterium strain SB3*	pHSB1	1735	X07128
Halobacterium strain SB3*/ *Halobacterium halobium* strain R1	pHSB2	1781	X66324
Haloferax volcanii strain DS2*	pHV2	6354	J03014
Pyrococcus abyssi strain GE5*	pGT5	3444	U49503
Acidianus ambivalens strain LEI10*	pDL10/pSL10	7598	AJ225333
Sulfolobus islandicus strain REN1H1*/ *Sulfolobus solfataricus* strain P1	pRN1	5350	U36383
Sulfolobus islandicus strain REN1H1*	pRN2	6959	U93082
Sulfolobus isolate NOB8-H2*	pNOB8	41 229	AJ010405

[a]Asterisk indicates source.

methylated the third residue to form GG^{Me}CC. Similarly to pFV1, pFZ1 encoded the DNA methyltransferase *Mth*ZIM modification system for the nucleotide sequence CTAG. Both the Z-245 and FTF strains of *M. thermoformicicum* utilized this restriction/modification system. Southern hybridizations further identified a region of pFZ2 homologous to the methyltransferase gene *mthZIM* of pFZ1. An ORF on pFZ1 in the reverse orientation of *mthZIM* may also encode the restriction enzyme for this modification system, *Mth*ZIR. This ORF has yet to be characterized biochemically to confirm this hypothesis. In addition to these shared sequences, both pFZ1 and pFV1 also contained regions that hybridized to the genomic DNA of *Methanobacterium thermoautotrophicum* strain ΔH and five *M. thermoformicicum* strains. The mechanism of replication was not identified for these plasmids. However, both single-stranded plasmid DNA and initiator proteins predicted for rolling circle mechanisms were undetectable in cells containing pFV1 and pFZ1, suggesting that the plasmids replicated by a theta mechanism.

The first shuttle vector developed for a methanogen utilized the cryptic plasmid pURB500 (**Table 1**) from *Methanococcus maripaludis* strain C5. Strain C5 was isolated from Airport Marsh on Sapelo Island, Georgia, in an area that contained the short form of *Spartina alterniflora* as the major flora. The study that isolated strain C5 was conducted to identify plasmid-containing methanococci for the development of a genetic system. These methanogens have a relatively short doubling time and a simple cell morphology, and are easily grown. pURB500 was the only plasmid described in the 21 methanococcal strains screened. This plasmid was found to be maintained at about three copies per genome in strain C5. Plasmid dimers were observed by electron microscopy. Tumbula *et al.* (1997) proposed that pURB500 may replicate using a theta mechanism to generate the dimers. Although pURB500 was sequenced (see **Table 1**) none of the 18 putative ORFs possessed a high degree of similarity to known proteins. Two areas of complex secondary structure that might be origins of replication were identified. However, subcloning of the pURB500 plasmid has yet to yield a minimum replicon containing only these regions. The host range of pURB500 appears to be restricted to strains of *M. maripaludis*.

Of the methane produced in the environment, 70% comes from aceticlastic methanogenesis, which occurs when acetate is disproportionated into methane and carbon dioxide. Interest in this transformation led to a

search of nine acetotrophic and three obligate methylotrophic methanogens for plasmids. A total of three plasmids were identified in strains maintained as laboratory cultures for over 5 years. The three strains, *Methanosarcina acetivorans* strain C2A, isolate C2B, and isolate C2C, were obtained from sediments collected from the Scripps Submarine Canyon off the coast of La Jolla, California. The sizes and restriction endonuclease maps of the plasmids were similar, so only the plasmid (pC2A) from *M. acetivorans* strain C2A was further characterized. pC2A was a low copy number plasmid with about 6 copies per cell. Sequencing identified 4 ORFs larger than 120 amino acids. One ORF had homology to family of recombinases; a second ORF had homology to a replication initiation protein. The other ORFs did not possess known homologues. pC2A was used to construct a shuttle vector for *M. acetivorans* and *Escherichia coli*. The shuttle vector, pWM307 (see below), had a broad host range and was able to replicate in seven *Methanosarcina* species.

Plasmids have also been identified in hyperthermophilic methanococci. Two plasmids have been isolated from *Methanococcus jannaschii* strain JAL-1 ('*Methanocaldococcus jannaschii*'), and one plasmid has been isolated from *Methanococcus fervens* strain AG86 ('*Methanocaldococcus fervens*'). Strain AG86, which is closely related to *M. jannaschii* JAL-1, contains one plasmid called pURB900. Restriction digestions and hybridizations indicated that the 20 kb pURB900 was not related to the plasmids from *M. jannaschii* JAL-1. Both *M. jannaschii* plasmids were sequenced during a total genome sequencing project. pURB800 (also known as MJECL) possessed 44 ORFs, five of which had homology to known proteins. There were regions of nucleotide sequence similarity among the putative genes that suggested recent recombination events between pURB800 and the genome; the origin of replication was not identified in this plasmid. The smaller *M. jannaschii* plasmid pURB801 contained 5 ORFs, none of which had homology to known proteins. The origin of replication was also not identified in pURB801. Neither pURB800 nor pURB801 possessed homology to pURB500 from *M. maripaludis* strain C5. Development of pURB800 or pURB801 into shuttle vectors has not been reported.

Methanogenic isolates from mesophilic and thermophilic molasses stillage fermenters as well as rabbit faeces were screened for plasmids. Of the 79 strains examined, five contained plasmids. A 26.5 kb plasmid was found in *Methanosarcina barkeri* strain DSM 800, which had previously been described as lacking plasmids and was used as a negative control for this experiment; this example illustrates the difficulty of identifying plasmids. Biavati *et al*. (1992) suggested that the methodology for plasmid isolation contributed to the discovery of the plasmid in DSM 800 and that low copy number plasmids could also be underrepresented by their methodology.

Including the study of Biavati *et al*. (1992), 137 methanogenic strains have been screened for plasmids. Plasmids have been reported in 16 (12%) of these strains. Of the sequenced plasmids described, none contained ORFs with homology to known antibiotic resistance markers. This lack of antibiotic resistance markers was not surprising because archaea are insensitive to many of the antibiotics for bacteria and eukarya.

Halophiles

In the Euryarchaeota the organisms collectively referred to as the halophiles include extreme halophiles (organisms that grow optimally at 2.5 to 5.2 mol L^{-1} salt) as well as moderate halophiles (organisms that grow at 0.5–2.5 mol L^{-1} salt). The extremely halophilic methanogens are not usually included in this largely aerobic group. In 1963, the initial identification of plasmid DNA was confirmed in two extremely halophilic archaea. Since that discovery, plasmids have been found in many halophilic archaea, where multiple plasmids are common and can comprise up to 31% of the deoxyribonucleic acid in some species.

The moderate halophile *Haloferax volcanii* strain DS2 (formerly known as *Halobacterium volcanii*), contained four different plasmids that contribute up to 25% of the cellular DNA. These plasmids were the 86 kb pHV1, the 6.4 kb pHV2, the 442 kb pHV3 and the 690 kb pHV4. Only pHV2 was studied in great detail; about 6 copies per chromosome were present in the cell. Sequence analysis identified four ORFs with sizes greater than 189 amino acids and numerous smaller ORFs; none of the ORFs possessed homology to known proteins. pHV51 was a natural variant of pHV2 that possessed an insertion element; this cryptic plasmid was developed into the shuttle vector pWL102 (see below), replicated in three different halophilic genera that spanned the 16S rRNA phylogenetic tree for halophilic archaea and included *Haloferax*, *Halobacterium* and *Haloarcula*.

Haloferax strain Aa 2.2 possessed two plasmids. The smaller 10.5 kb cryptic plasmid was named pHK2 and was maintained at 7–8 copies per cell. With the addition of a selectable marker, pHK2 was transformed into *H. volcanii*. Both pHK2 + marker and pHV2 were recovered from the transformed cells (Holmes and Dyall-Smith, 1990). While never completely sequenced, a 3359 bp fragment of pHK2 was found to contain the minimum region required for replication. This region contained three ORFs and four inverted repeats. Although ORF1 had homology to rolling-circle initiator replication proteins, single-stranded DNA replication intermediates could not be detected.

Plasmids initially were found in *Halobacterium* based upon the observation that many species contained satellite DNA with a different mol% G + C from that in genomic DNA.

In *Halobacterium salinarium* strain 5, a 66.7 kb plasmid was found to be associated with gas vacuole production. This strain also contained two additional plasmids of 39.4 kb and 130.3 kb that were never linked to a phenotype. Gas vacuole production was also linked to the 150 kb plasmid pHH1 from *Halobacterium halobium* strain NRC817. This plasmid was maintained at four copies per chromosome. Other *H. halobium* strains also contained plasmids with homology to pHH1. *H. halobium* strain DSM670, *H. halobium* strain DSM671 and *Halobacterium cutirubrum* maintained plasmids that varied in size from 75.8 kb to 150 kb with homology to pHH1. These strains have not been characterized further.

Halobacterium halobium contained a dynamic population of plasmids formed by a high rate of deletions. pHH9, a 5.7 kb derivative of pHH1, was formed through successive spontaneous deletions of pHH1. This plasmid had a copy number identical to pHH1 and presumably contained all the information necessary for stable replication. Sequencing and subcloning parts of pHH9 identified a minimum replicon of 2.9 kb. This DNA fragment contained a 1893 bp ORF and a 350 bp AT-rich region. The ORF did not possess homologues in the databases. The AT-rich region was similar to origins of replication for some bacterial plasmids. A shuttle vector was developed from pHH9 (see below). This vector, pUBP2, had a broad host range and replicated in *H. halobium*, *H. volcanii*, *Haloarcula hispanica*, and *Haloarcula vallismortis*.

Another minimal region for replication was identified in a plasmid from another strain of *H. halobium*. Strain NRC-1 contained one plasmid of 350 kb (pNRC200) and one of 200 kb (pNRC100); it also contained a population of smaller plasmids derived from pNRC100. A minimum replication region of only 3874 bp was identified for pNRC100. This region contained a 3027 bp ORF with homology to replication proteins and an AT-rich region. The ORF also had homology to an ORF from the *H. volcanii* plasmid pHV2. However, these ORFs were not homologous to the ORFs from pHH1. This 3.9 kb region of pNRC100 allowed plasmid replication in strain NRC-1 and *H. volcanii* strain WFD11. A shuttle vector was made by ligating the 3.9 kb fragment to an *E. coli* plasmid containing a selectable marker for *Halobacterium* and *Haloferax*. The vector was not stable in the absence of antibiotic selection in *H. volcanii*, so, even though the plasmid replicated, portions required for plasmid stability were lacking.

A large number of other plasmids have been described in *Halobacterium* strains (see **Table 1**). *Halobacterium* strain GN101 possessed the 1.7 kb pHGN1 and four large plasmids of 39, 43, 49 and 65 kb. Strain GRA contained a 38 kb plasmid and a 65 kb plasmid. Strain SB3 had a 52 kb plasmid and a 34 kb plasmid with homology to an insertion sequence on pHH1. Strain SB3 also possessed the 1.7 kb plasmids pHSB1 and pHSB2. Strain GRB maintained the 1.7 kb pGRB1 and two large plasmids at 35 kb and 65 kb. A single-stranded intermediate of pGRB1 was found in the host's cells, but the mechanism of replication was not determined. pHGN1 and pHSB1, the 1.7 kb plasmids from strain SB3, were also thought to replicate through single-stranded intermediates. These small plasmids, which were sequenced, all possessed a homologous ORF that was proposed to be a replication protein even though no direct evidence was reported.

Another common feature among the plasmids was the sequence GATTT(C/G). pGRB1, pHSB1 and pHGN1 contained multiple copies of this element. The location of the hexanucleotide element corresponded to similar regions among the plasmids. This sequence was thought to be common in bacterial origins of replication and was speculated to be the origin for these plasmids (Hackett *et al*., 1990).

During sequencing of pHSB1, pHGN1 and pGRB1, some positions appeared to contain more than one nucleotide. Thus, what were thought to be single plasmids were populations of closely related plasmids. Two forms of heterogeneity were described for these plasmids. Macroheterogeneity, or a large region with multiple nucleotide changes, was found in a 250 bp region of pHSB1 and pHSB2, which were 80% identical at the nucleotide level. Microheterogeneity, an isolated nucleotide change between plasmids, was detected in the sequence analysis of pHSB1 that led to the discovery of a pHSB1 population. Microheterogeneity was not described for pHSB2. No hybrids between pHSB1 and pHSB2 were detected. This observation was interesting since many *Halobacterium* species had high recombination rates, pHGN1 from strain GN101 also contained microheterogeneity. pHSB1, pHSB2 and pHGN1 were also thought to be maintained at a high copy number. pGRB1 was maintained in *Halobacterium* strain GRB at 180 copies/cell. pGRB1, after being developed as a shuttle vector for use with *E. coli*, also replicated in *H. halobium* strain R1 following transformation. Additionally, *H. halobium* strain R1 maintained pHSB2 after transformation.

Thermoplasmales and Thermococcales

Within the Euryarchaeota are the orders Thermococcales and Thermoplasmales. Thermococcales is composed of the genera *Thermococcus* and *Pyrococcus*. With the exception of a few thermophiles, these organisms are hyperthermophiles with optimum growth occurring at 80°C or higher. Thermoplasmales currently consists of the genera *Thermoplasma* and *Picrophilus*. Both genera are thermophiles with growth optimums near 60°C. These organisms are also hyperacidophilic and capable of growth below a pH of 1.

In 1995 two articles reported the isolation of plasmids from the Thermoplasmales. pTA1 from *Thermoplasma acidophilum* strain HO-1012 was a 15.2 kb cryptic plasmid maintained at 7–13 copies/cell. *T. acidophilum* strains HO-63 and HO-121 also contained plasmids that hybridized to

pTA1. An additional three strains contained plasmids that were not further characterized. Twelve *T. acidophilum* strains were screened for metal and novobiocin resistance. No significant differences were found between the plasmid-free and the plasmid-containing strains. Plasmids were also found in eight strains of *Picrophilus oshimae*. Six strains contained an 8.3 kb plasmid. Another strain contained an 8.8 kb plasmid, while the last strain carried both plasmids. Restriction analysis was not reported for these plasmids, but Southern hybridizations suggested that the 8.3 kb and 8.8 kb plasmids shared homologous regions. The closely related species *Picrophilus torridus* did not contain plasmids.

Many plasmids have been isolated from the order Thermococcales. Screening of the Thermococcales was conducted to increase the diversity and number of plasmids for basic research and biotechnological applications. During the screening of 52 strains from a Pacific hydrothermal vent, plasmids were found in 11 strains. Of these 11 strains, six were chosen for further characterization since they had a higher concentration of plasmid DNA. By restriction fragment length polymorphisms (RFLPs), strains I559, I560 and I690 were shown to be related to *Thermococcus stetteri*. Each strain contained a 3.5 kb and a 24 kb plasmid. Through the use of Southern hybridizations, the small plasmids were shown to hybridize to each other but not to the larger plasmids. Likewise, the large plasmids also hybridized to each other. A 5.3 kb plasmid (pGN31) was found in *Thermococcus* sp. strain GE31. pGN31 did not hybridize to the other plasmids within this study (Benbouzid-Rollet *et al*., 1997). In addition, plasmids were found in two *Pyrococcus abyssi* strains. *P. abyssi* strain GE23 contained a unique 16.8 kb plasmid called pGN23. This strain had 79% DNA hybridization to the type strain *Pyrococcus abyssi* strain GE5 (Benbouzid-Rollet *et al*., 1997).

A plasmid, pGT5, was also found in *P. abyssi* strain GE5, an organism from a Pacific hydrothermal vent. pGT5 was maintained at 25–30 copies per chromosome. It was 3444 bp, and was the only Thermococcales plasmid sequenced. From the nucleotide sequence, two open reading frames were identified. When translated, ORF1 possessed motifs similar to proteins involved in rolling-circle replication mechanisms. ORF2 did not have homology to proteins within the databases. However, based on location and putative identification of ORF1 to the pC194 family of plasmids that replicate by a rolling-circle mechanism, it was proposed that ORF2 may be involved in recombination. Plasmids that utilize a rolling-circle mechanism contain two origins of replication. The plus origin is the site recognized by a replication protein that binds and then nicks the plasmid. This nick allows the synthesis of single-stranded (ss)DNA from the plasmid. After the ssDNA synthesis is complete, the replication protein terminates the reaction and forms a single-stranded product and a double-stranded plasmid. The minus origin is then used for the conversion of the single-stranded intermediate into a double-stranded plasmid. A region of 11 nucleotides identical to the double-stranded origin of replication used by other rolling-circle type plasmids was identified in pGT5. A region of low identity corresponding to the second origin of replication was also suggested. To confirm a rolling-circle mechanism of replication, a single-stranded DNA intermediate was identified in cell extracts for the positive strand of the plasmid. In addition, *P. abyssi* strain GE27 possessed a 3.5 kb plasmid (pGN27) that was thought to be closely related to if not identical to pGT5 (Benbouzid-Rollet *et al*., 1997). pGT5 was the basis for a shuttle vector for *Pyrococcus furiosus*, *Sulfolobus acidocaldarius*, and *E. coli* (see below).

Crenarchaeota

The order Sulfolobales within the Crenarchaeota is composed of the genera *Sulfolobus*, *Acidianus*, *Metallosphaera*, *Sulfurisphaera*, and *Stygiolobus*. Species of the genus *Sulfolobus* oxidize hydrogen, elemental sulfur and sulfide under autotrophic growth conditions in media of low ionic strength. The genus *Acidianus* contains salt-tolerant autotrophs that also oxidize sulfides, elemental sulfur and hydrogen. Under anaerobic conditions, members of this genus reduce elemental sulfur to sulfide while oxidizing hydrogen.

Plasmids were discovered in the chemolithoautotroph *Acidianus ambivalens* strain LEI10 in 1985. Strain LEI10, formerly known as *Desulfurolobus ambivalens*, was isolated at pH 2.3 and 94°C in a solfataric waterhole in Iceland and contained three plasmids. The most abundant plasmid, pDL10, previously known as pSL10, was 7.7 kb. Although pDL10 was initially thought to be involved in the reduction of elemental sulfur to hydrogen sulfide, strains cured of plasmids were capable of sulfur reduction, and this hypothesis was disproved. pDL10 had around 50 copies per chromosome under anaerobic conditions. Under aerobic conditions, cells contained 20-fold less plasmid. Plasmids similar to pDL10 were also found in 20 out of 22 isolates of *A. ambivalens*. Although the sequence of pDL10 appears in the databases (see **Table 1**), the description of the nucleotide sequence has not been published.

Three plasmids were identified during a screening of Icelandic solfataric field isolates belonging to the order Sulfolobales. These plasmids came from *Sulfolobus islandicus* strains HEN7H2 and REN1H1. pHE7 from *S. islandicus* strain HEN7H2 was maintained at 15 copies per chromosome and had a size of 7 kb. This cryptic plasmid has never been sequenced. *S. islandicus* strain REN1H1 contained the plasmids pRN1 and pRN2 at 20 and 30 copies per chromosome, respectively. pRN1 had a host range of *S. islandicus* and *Sulfolobus solfataricus* but not *Sulfolobus acidocaldarius*. *S. solfataricus* was proposed as a

genetic model for crenarchaeotes since electroporation and plating techniques were developed, and pRN1 was the first plasmid completely sequenced from the Crenarchaeota. pRN1 contained six ORFs within its 5350 bp. One of these ORFs had homology to the repressor protein CopG, which controls the expression of a protein that initiates plasmid replication in bacteria. The remaining ORFs did not possess homology to known proteins. The replication origin for this plasmid was not identified. pRN2, which was recently sequenced, had the same host range as pRN1 and three ORFs in pRN2 were in the same orientation in pRN1. The first ORF was the CopG homologue; the two remaining ORFs were conserved between the two plasmids but did not have any other homologues. Noncoding regions of the plasmids were also conserved in five distinct blocks. The fifth block contained multiple copies of the interon CTAACTCT close to a polycytosine track. While this unusual structure was conserved between pRN1 and pRN2, no functioning role was identified.

The order Sulfolobales contained the first example of an archaeal plasmid that underwent horizontal transfer via a contact-mediated mechanism. The plasmid was the 41 kb pNOB8 from *Sulfolobus* isolate NOB8H2. The cryptic pNOB8 had around 20 copies per genome when the cells were grown in a liquid culture. Upon plating on a solid medium, cells contained only 2–5 copies of pNOB8/genome. Schleper *et al*. (1995) proposed that this low copy number may be representative of environmental growth conditions. Mating allowed pNOB8 to be transferred to different *Sulfolobus* species. In three strains of *S. solfataricus* and three strains of *S. islandicus*, the copy number was 20–40 copies of pNOB8 per genome in liquid culture. The reduction in copy number was repeated upon plating these strains. In *S. islandicus* strain HEN7H2, which possessed both pNOB8 and the nonconjugative pHE7, only pNOB8 was transferred. Re-isolation of plasmid DNA from the host and recipients showed that the cells contained wild-type pNOB8 and smaller pNOB8-derived plasmids. The major variant of the smaller plasmids was named pNOB8-33. pNOB8-33 was found only in recipient strains and not in isolate NOB8H2 and was missing about 8 kb of the 41 kb plasmid. pNOB8-33 was still able to replicate and to be transferred among *Sulfolobus* species.

pNOB8 was recently sequenced because of the interest in archaeal conjugation and the development of genetic tools for *Sulfolobus*. pNOB8 was a 41 229 bp plasmid containing multiple repeats. The largest perfect repeat was 85 bp separated by 7942 bp. This repeat was identified to be the point of recombination that yielded pNOB8-33. About 50 ORFs were identified in pNOB8. Of these ORFs, fewer than 20% possessed homology to known proteins. However, homologies were identified to transposases (see section below on Insertion Sequences) and proteins involved in partitioning plasmids to daughter cells. Some of these latter ORFs were lost upon the formation of pNOB8-33, which may have contributed to its instability.

pNOB8 has also been electroporated into *S. solfataricus* (Schleper *et al*., 1995). An expression system for *S. solfataricus* strain PH1 was developed by cloning the strong promoter for ribosomal protein S12 from *S. acidocaldarius* and the β-galactosidase gene from *S. solfataricus* strain PI into pNOB8. This expression vector was called pNOB8:*lacS*, but this expression system was not stable. After transformation of pNOB8:*lacS* into *S. solfataricus* PH1, both large and small colonies that contained β-galactosidase (*lacS*) were discovered. Using Southern hybridizations, these larger colonies were found to have lost the pNOB8 but contained *lacS* integrated into the genome. The smaller colonies initially maintained the expression vector, but restreaking frequently generated plasmid-minus colonies.

After the description of pNOB8, eleven additional conjugative plasmids were identified during the screening of nearly 300 strains of *Sulfolobus islandicus*. These plasmids were proposed to belong to a single family that includes pNOB8. While not all of the plasmids were able to initiate conjugation, all could be mobilized during conjugation. For instance, although the 6.0 kb plasmid pING3 was nonconjugative, it could be mobilized by pING6. Many of the plasmids were closely related and some plasmids, such as the pINGs and pSOG2/4 plus clones, formed subfamilies. Nine of the 11 new plasmids were between 25.0 kb and 26.6 kb. The largest new plasmid was 36.5 kb. Although some of the new plasmids were incompatible, others were stably maintained in the same strain, albeit at different copy number. To aid the understanding of compatibility, members of the pING family are being sequenced.

Currently, all of the plasmids from the Crenarchaeota are from Sulfolobales. The plasmid size ranges from 5.3 kb to 41 kb. Most of the plasmids were maintained at a high copy number in the original host. Some plasmids had an initially high copy number after transferral into a new *Sulfolobus* isolate. However, the copy number decreased during subsequent growth of the isolate. None of the sequenced plasmids contained identifiable antibiotic resistance genes, and all of them were cryptic. For a review, see Zillig *et al*. (1998).

Insertion Sequences and Transposons

Insertion sequences

Insertion sequences (IS) are self-contained genetic elements that are capable of insertion into new regions of DNA; they can also contain gene(s). Bacterial IS frequently encode transposases essential for transposition of the element and IS can also influence the expression of neighbouring gene(s) outside of the element. Many IS have

been discovered in the Archaea (Derkacheva *et al*., 1993; Pfeifer, 1987).

Using libraries from *H. halobium* strains NRC-1 and R-1, over 50 IS families were identified on genomic DNA and plasmids. IS were proposed to contribute to the rapid rearrangements that occurred in the plasmids from this genus. Three plasmids from *Halobacterium* species were found to contain insertion sequences, and the 150 kb pHH1 from *H. halobium* strain NRC817 had seven different types of insertion sequences. Many of these elements were present in multiple copies on the plasmid. The 200 kb pNRC100 from *H. halobium* strain NRC-1 contained three of the same IS found in pHH1 plus one additional IS. Upon sequencing, no IS were identified in pGRB1 from *Halobacterium* strain GRB. Upon passage through *H. halobium* strain R-1, pGRB1 acquired an additional region of DNA that corresponded to a new IS, called ISH11. The 1068 bp ISH11 was typical of the IS found in *H. halobium* species. It contained an inverted repeat of 15 bp at the terminal ends of the IS and an ORF that corresponded to 334 amino acids. Although a function was not proposed for the putative protein, other IS contained an ORF that appeared necessary for transposition. ISH27 had observed transpositions into pHH4 and pHH6. When Northern hybridizations were performed on *H. halobium* ORF1167 from ISH27, messenger RNA for this ORF was detected in *H. halobium*. Evidence for ORF1167's involvement in transposition was discovered during the analysis of ISH51 transposition events (see below). For reviews of individual elements, see Derkacheva *et al*. (1993) and Pfeifer (1987).

Haloferax volcanii also contained IS. Of the 20–30 copies of the ISH51 found, five were located on the 86 kb plasmid pHV1. The ISH51 family was found to have at least three members. ISH51-3, found by a recent transposition into pHV2, was the third ISH51 element sequenced. ISH51-3 had 91% identity in 1200 nucleotides to ISH27-1 from the *H. halobium* plasmid pHH4. ISH51-3 contained a homologue of ORF1167, which was thought to be essential for transposition. Two members of the ISH51 family, ISH51-1 and ISH51-2, contained mutations that disrupted the ORF1167 homologue. ISH51-1 and ISH51-2 were thought to be nontransposable elements. ISH51-3, which contained a 21-nucleotide deletion at the 3′-terminal end of ORF1167 homologue when compared to ISH27, was observed to transpose into pHV2.

Plasmids from methanogens were also found to contain insertion sequences. The first insertion sequences were found in pFV1 and pFZ1 from *Methanobacterium thermoformicicum* strains THF and Z-245, respectively. FR-I was an IS element that resided in pFV1 and not pFZ1. This element was also found in the genomes of *M. thermoautotrophicum* strain ΔH and four *M. thermoformicicum* strains, but not strain Z-245. FR-I was 1501 bp with direct repeats at both ends of the element. The direct repeats were a palindromic sequence of AAATTT. Although pFZ1 lacked FR-I, this sequence was in a similar position, and may represent the integration site for FR-I. Upon comparison with the genomic-derived elements from *M. thermoformicicum* strains THF and ΔH, two conserved regions, 1118 bp box 1 and 383 bp box 2, were identified. Those regions encoded ORFs that did not have homologues in the databases. The second IS sequence was located in both plasmids. FR-II was 2510 bp in pFZ1 and 3034 bp in pFV1. The difference in size was caused by the addition of a 524 bp direct repeat in pFV1. The ORFs found in FR-II did not possess any homologues in the databases.

The genomic sequencing of *M. jannaschii* revealed a putative IS in pURB800. This IS belonged to an 11-member *M. jannaschii* family known as ISAMJ1. The pURB800 IS was named ISAMJ1-C, and was the smallest member of the family at 265 bp. ISAMJ1-C did not appear to encode a protein and had a 16 bp inverted repeat at the terminal regions of the element.

To date, insertion elements have been discovered on only one plasmid from the Crenarchaeota pNOB8 from *Sulfolobus* contains two putative insertion sequences. ORFs 406 and 413 appear to encode transposases. ORF 406 had 89% identity at the amino acid level to the Ro2 elements from *Sulfolobus solfataricus* strain P2 but did not contain terminal inverted repeats. The lack of repeats was also similar to the Ro2 elements from *S. solfataricus* strain P2. Repetitive sequences are frequently involved in the transposition of insertion sequences. The second potential IS was ORF 413, which had 28% identity at the amino acid level to the IS256 family. The IS256 family is a group of transposases from bacteria. The putative IS256 homologue had an imperfect inverted repeat of 32 bp flanking the ORF. Neither ORF 406 nor ORF 413 was proved experimentally to be a mobile insertion sequence.

Transposons

Transposons are genetic elements that usually comprise insertion sequences, genes required for transposition and accessory gene(s). These mobile elements can be used to identify genes by insertional inactivation, add exogenous genes, or map the location of genes within a genome. Unfortunately, no natural archaeal transposons are known. Artificial transposons were developed for *Haloarcula hispanica*. These transposons used the insertion sequences ISH2, ISH26 and ISH28 from *Halobacterium halobium*. These ISH elements were chosen because they had been sequenced, were very mobile, and did not require a specific integration sequence. The *Haloferax volcanii* gene for mevinolin resistance was cloned into each IS, and the artificial transposon was ligated into an *E. coli* plasmid. After transformation into *H. hispanica* and selection on medium containing mevinolin, no transformats were recovered with the element constructed from ISH2. However, the artificial elements constructed from ISH26

and ISH28 had average transformation frequencies of 60 and 840 colony-forming units per microgram of DNA. Subsequent experiments confirmed that *H. hispanica* did not contain ISH26-like or ISH28-like elements prior to transformation and that the transformants contained the artificial elements. Upon analysis of individual transpositions, some isolates appeared to have duplicated the transposon region, and in one case the *E. coli* plasmid was co-integrated. This co-integration was thought to occur after the primary integration event or possibly by transposition of a plasmid dimer.

Archaeal Shuttle Vector Development and Use in Biotechnology

Archaeal shuttle vector development is important because assembly of an active enzyme may require an archaeon's unusual cofactors and enzymatic pathways. The vectors must contain replicons for *E. coli* and at least one archaeon, genes that encode antibiotic resistance, and species-specific promoters for controlled expression of genes. There is significant interest in these vectors for both basic research and biotechnology.

Shuttle vectors are a recent occurrence for thermophiles and methanogens. A challenge for the thermophiles has been the identification of selectable markers and thermostable reporters. pAG1, the first thermophilic shuttle vector (**Table 2**), utilized the alcohol dehydrogenase (ADH) from the thermophile *Sulfolobus solfataricus* as a selectable marker. The vector was also found to replicate in *Pyrococcus furiosus* and *Sulfolobus acidocaldarius*. A thermostable β-galactosidase from *S. solfataricus* was utilized as a reporter. For methanogens, the obstacle was the absence of transformation systems. Uptake of genomic DNA by *Methanococcus voltae*, for example, was inefficient with only 2–100 transformants/μg of genomic DNA. Recently, *M. maripaludis* and *M. acetiovorans*, using a poly(ethylene glycol) or liposome-mediated methodology for the respective organism, were transformed with plasmid DNA. These transformations had efficiencies of 10^7 to 10^8 transformants/μg of plasmid using pDLT44 for *M. maripaludis* or pWM307 for *M. acetiovorans*. However, neither shuttle vector has been tested for *in vivo* heterologous gene expression. The plasmid pWLG30, which was similar to pDLT44, expressed the *E. coli* β-galactosidase gene in *M. maripaludis*.

Ventosa and Nieto (1995) reviewed the potential applications of halophiles and their products. An example was the membrane protein bacteriorhodopsin. Bacteriorhodopsin forms a hexagonal array in the membrane surface and composes up to 75% of the membrane dry weight of *H. halobium*. This protein is a light-driven proton pump, and is used for ATP synthesis in *H. halobium*. Commercial products include the development of sensors and (possibly) data storage devices. pMPK35, a shuttle vector used in *E. coli* and *H. halobium*, was able to express bacteriorhodopsin at 25–40% of wild-type levels in *H. halobium* strains lacking the gene. This vector will allow intensive site-directed mutagenesis studies with bacteriorhodopsin.

Table 2 Engineered vectors for the Archaea

Plasmid	Size (kb)	Host(s)
pAG1	4.730	*Pyrococcus furiosus* *Sulfolobus acidocaldarius* *Escherichia coli* strain JM109
pDLT44	12.691	*Methanococcus maripaludis* strain JJ *Escherichia coli* strain XL1-Blue MRF
pMPK35	4.5	*Halobacterium halobium* strain R1 *Halobacterium* strain GRB *Escherichia coli*
pUBP2	12.3	*Halobacterium halobium* *Haloferax volcanii* strain WFD11 *Haloarcula hispanica* *Haloarcula vallismortis* *Escherichia coli* strain DH5α
pWL102	10.5	*Haloferax volcanii* strain DS2 *Haloarcula hispanica* *Haloarcula vallismortis* *Halobacterium halobium* strain PO3 *Escherichia coli* strain DH5α
pWM307	8.715	*Methanosarcina acetivorans* strain C2A *Methanosarcina barkeri* strains Fusaro, MS, W *Methanosarcina mazei* strain S-6 *Methanosarcina thermophilia* strain TM-1 *Methanosarcina siciline* strain C2J *Methanosarcina* spp. WH1 and WH2 *Escherichia coli* strain DH5α/λpir

References

Biavati B, Sgorbati B and Palenzona D (1992) Plasmid DNA from methanogenic bacteria. *Current Microbiology* **24**: 285–287.

Benbouzid-Rollet N, López-García P, Watrin L, Erauso G, Prieur D and Forterre P (1997) Isolation of new plasmids from hyperthermophilic Archaea of the order Thermococcales. *Research in Microbiology* **148**: 767–775.

Derkacheva NI, Kagramanova VK and Man'kin AS (1993) Genetic variability in halophilic archaebacteria (a review). *Molecular Biology* **27**: 287–295.

Hackett NR, Krebs MP, DasSarma S *et al*. (1990) Nucleotide sequence of a high copy number plasmid from *Halobacterium* strain GRB. *Nucleic Acid Research* **18**: 3408.

Holmes ML and Dyall-Smith ML (1990) A plasmid vector with a selectable marker for halophilic archaebacteria. *Journal of Bacteriology* **172**: 756–761.

Pfeifer F (1987) Genetics of halobacteria. In: Rodriguez-Valera F (ed.) *Halophilic Bacteria*, pp. 105–133. Boca Raton, FL: CRC Press.

Schleper C, Holz I, Janekovic D, Murphy J and Zillig W (1995) A multicopy plasmid of the extremely thermophilic archaeon *Sulfolobus* effects its transfer to recipients by mating. *Journal of Bacteriology* **177**: 4417–4426.

Thomm M, Altenbuchner J and Stetter KO (1983) Evidence for a plasmid in a methanogenic bacterium. *Journal of Bacteriology* **153**: 1060–1062.

Tumbula DL, Bowen TL and Whitman WB (1997) Characterization of pURB500 from the archaeon *Methanococcus maripaludis* and construction of a shuttle vector. *Journal of Bacteriology* **179**: 2976–2986.

Ventosa A and Neito JJ (1995) Biotechnological applications and potentialities of halophilic microorganisms. *World Journal of Microbiology and Biotechnology* **11**: 85–94.

Woese CR and Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the USA* **74**: 5088–5090.

Zillig W, Arnold HP, Holz I *et al*. (1998) Genetic elements in the extremely thermophilic archaeon *Sulfolobus*. *Extremophiles* **2**: 131–140.

Further Reading

Charlebois RL (1999) Evolutionary origins of the haloarchaeal genome. In: Oren A (ed.) *Microbiology and Biogeochemistry of Hypersaline Environments*, pp. 309–317. Washington DC: CRC Press.

Charlebois RL and DasSarma S (1995) Insertion elements of halophiles. In: DasSarma S and Fleischmann EM (eds) *Archaea: A Laboratory Manual, Halophiles*, pp. 253–255. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Zeikus JG, Vieille C and Savchenko A (1998) Thermozymes: biotechnology and structure–function relationships. *Extremophiles* **2**: 179–183.