Extremophiles (1998) 2:417–425

ORIGINAL PAPER

Qunxin She · Hien Phan · Roger A. Garrett
Sonja-Verena Albers · Kenneth M. Stedman
Wolfram Zillig

Genetic profile of pNOB8 from *Sulfolobus*: the first conjugative plasmid from an archaeon

Received: March 10, 1998 / Accepted: May 2, 1998

Abstract The complete nucleotide sequence of the archaeal conjugative plasmid, pNOB8, from the *Sulfolobus* isolate NOB8-H2, was determined. The plasmid is 41 229 bp in size and contains about 50 ORFs. Several direct sequence repeats are present, the largest of which is a perfect 85-bp repeat and a site of intraplasmid recombination in foreign *Sulfolobus* hosts. This recombination event produces a major deletion variant, pNOB8-33, which is not stably maintained. Less than 20% of the ORFs could be assigned putative functions after extensive database searches. Tandem ORFs 315 and 470, within the deleted 8-kb region, show significant sequence similarity to the protein superfamilies of ParA (whole protein) and ParB (N-terminal half), respectively, that are important for plasmid and chromosome partitioning in bacteria. A putative *cis*-acting element is also present that exhibits six 24-mer repeats containing palindromic sequences which are separated by 39 or 42 bp. By analogy with bacterial systems, this element may confer plasmid incompatibility and define a group of incompatible plasmids in Archaea. Although several ORFs can form putative trans-membrane or membrane-binding segments, only two ORFs show significant sequence similarity to bacterial conjugative proteins. ORF630b aligns with the TrbE protein superfamily, which contributes to mating pair formation in Bacteria, while ORF1025 aligns with the TraG protein superfamily. We infer that the conjugative mechanism for *Sulfolobus* differs considerably from known bacterial mechanisms. Finally, two transposases were detected; ORF413 is flanked by an imperfect 32-bp inverted repeat with a 5-bp direct repeat at the ends, and ORF406 is very similar in sequence to an insertion element identified in the *Sulfolobus solfataricus* P2 genome.

Communicated by G. Antranikian

Q. She · H. Phan · R.A. Garrett (✉)
Institute of Molecular Biology, Copenhagen University, Sølvgade 83H DK-1307, Copenhagen K, Denmark
Tel. +45-35-32-20-10; Fax +45-35-32-20-40
e-mail: garrett@mermaid.molbio.ku.dk

S.-V. Albers · K.M. Stedman · W. Zillig
Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany

Key words *Sulfolobus* · Thermophile · pNOB8 · Conjugation · Archaeon

Introduction

Although plasmids occur widely in the archaeal domain and several have been isolated and partially characterized, little is known about their mechanisms of maintenance, copy number control, or conjugation (Zillig et al. 1996). This contrasts with our knowledge of bacterial plasmid function for which the mechanisms of maintenance and copy number control are fairly well understood. There is also a rudimentary understanding of the diverse and complex mechanisms of conjugation, at least for some proteobacteria and gram-positive bacteria (Pansegrau and Lanka 1996).

Recently, the first archaeal conjugative plasmid, pNOB8, was isolated from *Sulfolobus* NOB8-H2 (Schleper et al. 1995), and it is of particular interest because it probably encodes the proteins responsible for maintenance, copy number control, and conjugation. Moreover, a major genetic variant, pNOB8-33, forms when pNOB8 is transferred into foreign *Sulfolobus* hosts in which it is not stably maintained (Schleper et al. 1995). Therefore, the nucleotide sequences of pNOB8 and the genetic variant pNOB8-33 were determined to make a seminal comparative study of putative gene products from pNOB8 with those of known bacterial conjugative plasmids and to analyze the regulatory mechanisms of an archaeal conjugative plasmid.

Materials and methods

Cloning and subcloning of pNOB8 plasmid

pNOB8 and pNOB8-33 DNA were prepared from *Sulfolobus* NOB8-H2 and *S. solfataricus* PH1, as described previously (Schleper et al. 1995). Nine of ten *Bam*HI fragments from this DNA were cloned into the *Bam*HI site of

pUC18, and subcloning of the *Bam*HI clones was performed using *Hin*dIII and other convenient restriction enzymes that had a compatible site within the multiple cloning region of the vector. A shot-gun library was also produced from DNA fragments of pNOB8 generated by sonication. Frayed ends of the DNA fragments were removed by treatment with mung bean nuclease (Amersham, Buckinghamshire, UK), and 0.7–2 kb fragments were recovered from agarose gels and ligated into the *Sma*I site of pUC18.

DNA sequencing and sequence analyses

All DNA clones and subclones were prepared from *E. coli* cells by the alkaline-SDS method (Sambrook et al. 1989) and purified on Jet-star columns (Genomed, Research Triangle Park, NC, USA). They were sequenced in a Vistra DNA Sequencer 725 (Amersham) using dye-primer chemistry. After primary sequencing, gaps were filled using dye-terminator chemistry in an ABI Sequencer 373A (Perkin-Elmer, Norwalk, CT). Some regions of the second strand were also analyzed using dye-terminator chemistry. Sequences were assembled by Sequencher 3.0 (Gene Code, Ann Arbor, MI, USA). Open reading frames (ORFs) were identified using GeneMarker (Borodovsky and McIninch 1993), and NCBI and EMBL databases were searched for potential homologs with BLAST and PSI-BLAST (Altschul et al. 1997) and FASTA3 (Pearson and Lipman 1988). These ORFs were also analyzed for membrane-binding and trans-membrane motifs using an EMBL server (Rost et al. 1995) and for other putative motifs by searching the Prosite (Bairoch 1993) and BLOCK databases (Henikoff and Henikoff 1994). Sequences were aligned using CLUSTAL W1.60 (Thompson et al. 1994). Repeated sequences were found by using DNA Strider (Marck 1988) and the Winseq program written by Flemming G. Hansen (personal communication).

Results

Sequence of pNOB8

A combined approach of direct cloning and shotgun cloning was employed to obtain DNA fragments for sequencing. Cleavage of pNOB8 generates ten *Bam*HI fragments and their altered mobilities on agarose gels were used to characterize genetic variants of the plasmid (Schleper et al. 1995). These fragments were cloned into pUC18 and sequenced. For the largest *Bam*HI fragment, it was necessary to digest with *Bgl*II before cloning into pUC18. The primary clones were digested with different restriction enzymes and subcloned into pUC18. A shotgun library of the clones containing pNOB8 fragments was also obtained. All these clones were sequenced, using both universal and reverse primers, to generate the primary sequence of the plasmid. Remaining gaps were filled either by sequencing subclones or by dye-terminator sequencing from oligonucleotide primers, using an ABI 373A sequencer. Finally, any remaining single-strand sequences were polished using dye-terminator chemistry. Thus, the complete sequence of pNOB8 was determined on each DNA strand at least once. The plasmid contains 41 229 bp and about 50 ORFs that are organized as illustrated in Fig. 1. Searches in the DNA and protein sequence databases, using different searching tools including PSI-BLAST, the new generation of programs for detecting low homology, revealed no significant sequence similarity for 85% of the genes and gene products of pNOB8. Significant similarites were observed for nine of the larger ORFs, which are listed in Table 1. The plasmid sequence is available in the EMBL/GenBank databases (accession no. AJ010405).

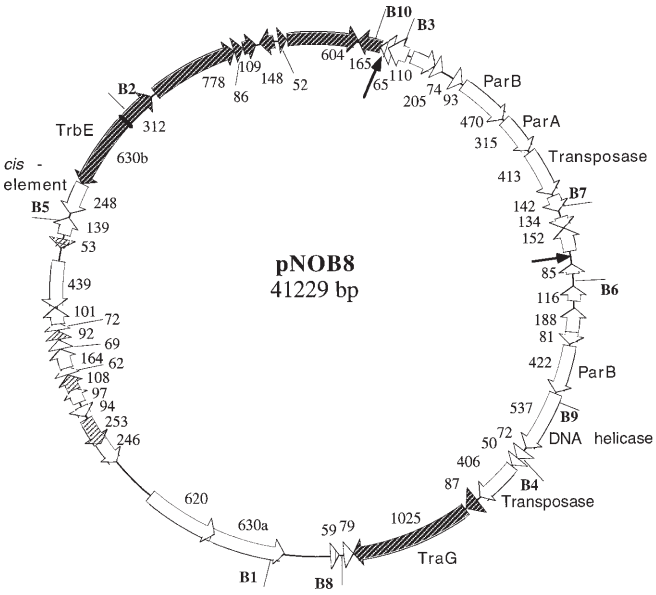

Fig. 1. ORFs identified in pNOB8. *Arrows* show the locations, orientations, and the relative sizes of the ORFs encoded by the circular plasmid. The identities of putative homologs are indicated alongside the ORFs. *Shaded* ORFs are homologs of those identified in the pING family of conjugative *Sulfolobus* plasmids (Stedman et al., in preparation); *dark* and *light shading* indicate >75% and 50%–75% amino acid identity, respectively. B1 to B10 indicate *Bam*HI cleavage sites numbered in decreasing size and clockwise from the cleavage sites (see Fig. 2A). The *filled arrows* show the limits of the 8-kb fragment deleted in pNOB8-33

Direct sequence repeats and formation of a major deletion variant

A major genetic variant of pNOB8 carrying a large deletion is observed in foreign hosts where it is not stably maintained. It was sequenced to gain insight into the mechanism of plasmid maintenance and variation. First, however, we searched for repeated sequences in pNOB8 (see Materials and methods) that might give rise, via recombination, to this and the other genetic variants which were observed earlier (Schleper et al. 1995). The direct repeats carrying up to two mismatches that are longer than 24 bp are listed in Table 2. The largest is a perfect repeat of 85 bp separated by 7942 bp.

Table 1. Identities of ORFs based on a PSI-BLAST search of databases

ORF	Putative homolog	aa-aligned (total gapped residues)	aa identity/ similarity (%)	Score (−E)
1025	TrbE family (12 VirB4, CagE, TrbE, Tra2)	781–796 (146–163)	13–18/25–32	131 (psi-6)
	TrsE of pSK41	692 (95)	17/34	106 (psi-6)
	MJECL08	562 (93)	20/36	79 (psi-6)
	TraG family (8 TraG, TraD, TaxB)	507–509 (79–81)	13–18/25–35	60 (psi-6)
630a	*S. aureus* ScdA, putative cell division	206 (41)	26/42	61 (psi-1)
630b	TrbE family (12 VirB4, CagE, TrbE, Tra2)	617–712 (81–106)	11–17/26–33	141 (psi-3)
	TrsE of pSK41	655 (109)	14/27	87 (psi-3)
537	ERCC/XPD family of DNA helicases (27 eukaryotic and bacterial DinG helicases)	706 (185)	13–15/27–28	128 (psi-4)
470	MJ1322	443 (60)	22/36	118 (spi-2)
	ParB family (10)	103–122 (12–26)	22–28/42–49	27 (psi-2)
422	Roa307 of plasmid QpH1	209 (14)	21/45	28 (spi-2)
	ParB family(10)	139–144 (7–20)	18–22/36–44	20 (psi-2)
413	Transposases (15 *Mycobacterium*, 6 *Rhizobium*, several other bacteria)	371–381 (19–29)	23–28/42–48	120 (psi-3)
406	Transposases, mainly *H. pylori*	385–426 (13–65)	20–23/39–42	120 (psi-7)
	Unassigned ORFs:			
	1. *S. solfataricus* chromosome	117–284 (1–3)	71–89/83–93	135 (blast)
	2. Eukaryotes and bacteria	385–409 (13–50)	18–23/(33–44)	118 (psi-7)
315	ParA superfamily (21)	273–302 (20–42)	16–24/33–37	73 (psi-4)

aa, amino acid

Families of the different proteins were aligned, and the number of ORFs analyzed are given in brackets. The range of aligned and gapped residues is given for these families in column 3, and the range of identities and similarities is given in column 4. The best score obtained for each family is indicated in column 5 with an −E number, e.g., $e^{-130} = 130$. Blast indicates a Blastp search, and psi-1 to -7 indicates the iteration number of the PSI-Blast search

Table 2. Direct sequence repeats in pNOB8

Length (bp)	Mismatches	Positions	Spanning length (bp)
85	0	33118–41145	7942
40	1	36287–2487	7389
35	2	39996–4790	5988
33	2	36335–2535	7396
30	2	33213–11	7995
27	0	4889–23108	18192
26	2	36428–2613	7388
26	2	4805–15937	11142
26	2	15973–23024	7025
26	2	17079–22248	5143
25	1	33583–544	8165
25	1	8293–14242	6149
24	2	9643–19520	9853

The first nucleotide of pNOB8 was assigned to the one immediately after the largest repetitive sequence in the *Bam*HI fragment 7 (see Fig. 1) such that pNOB8 and pNOB8-33 have the same numbering systems. The data exclude the 24-mer repeats in the putative *cis*-acting element and several additional perfect direct repeats that occur in the 15 to 23-bp range. Direct repeats were found using DNA Strider (see Materials and methods)

The major genetic variant, pNOB8-33, is about 8 kb smaller than the wild-type plasmid. When digested with *Bam*HI, fragments B3, B7, and B10 were absent and B10* appeared (Fig. 2A). Sequencing revealed that fragments B10, B3, and B7 are contiguous in the genome (Fig. 1) such that a single recombination event could produce the variant (Fig. 2B). Sequence analysis revealed that the largest repeat was present in fragments B10 and B7. These fragments border the deleted region, which includes all of B3 and parts of B10 and B7 (Fig. 1). Fragment B10* from pNOB8-33 was cloned into pUC18 and sequenced and shown to contain the remainders of fragments B10 and B7 and only one copy of the 85-bp sequence, thereby confirming that the deletion resulted from recombination at the 85-bp repeat (Fig. 2B).

Maintenance of pNOB8 and pNOB8-33

When present in a foreign recipient, such as *Sulfolobus solfataricus* PH1 (Schleper et al. 1995), pNOB8 is not stably maintained. Immediately after conjugation, the plasmid replicates rapidly and its copy number rises to more than 30 copies per chromosome (Schleper et al. 1995). Thereafter, replication is inhibited and the copy number decreases (data not shown). Upon further growth the culture is cured of the plasmid. In the parent *Sulfolobus* strain, NOB8-H2, curing was never observed and the copy number was usually low, indicating that a plasmid maintenance system operates. For another parent *Sulfolobus* strain, NOB8-H1, isolated from the same enrichment culture, pNOB8 was always present in low copy number and no curing was observed.

On repeated transfer and growth of cultures of *S. solfataricus* PH1 transformed with pNOB8, the deletion variant pNOB8-33 appears frequently whereas it has not been observed in the parent strain NOB8-H2, probably because it is not stably maintained in the latter. Moreover, *S. solfataricus* PH1 containing pNOB8 grows very slowly and exhibits a decreased plating efficiency (36% compared

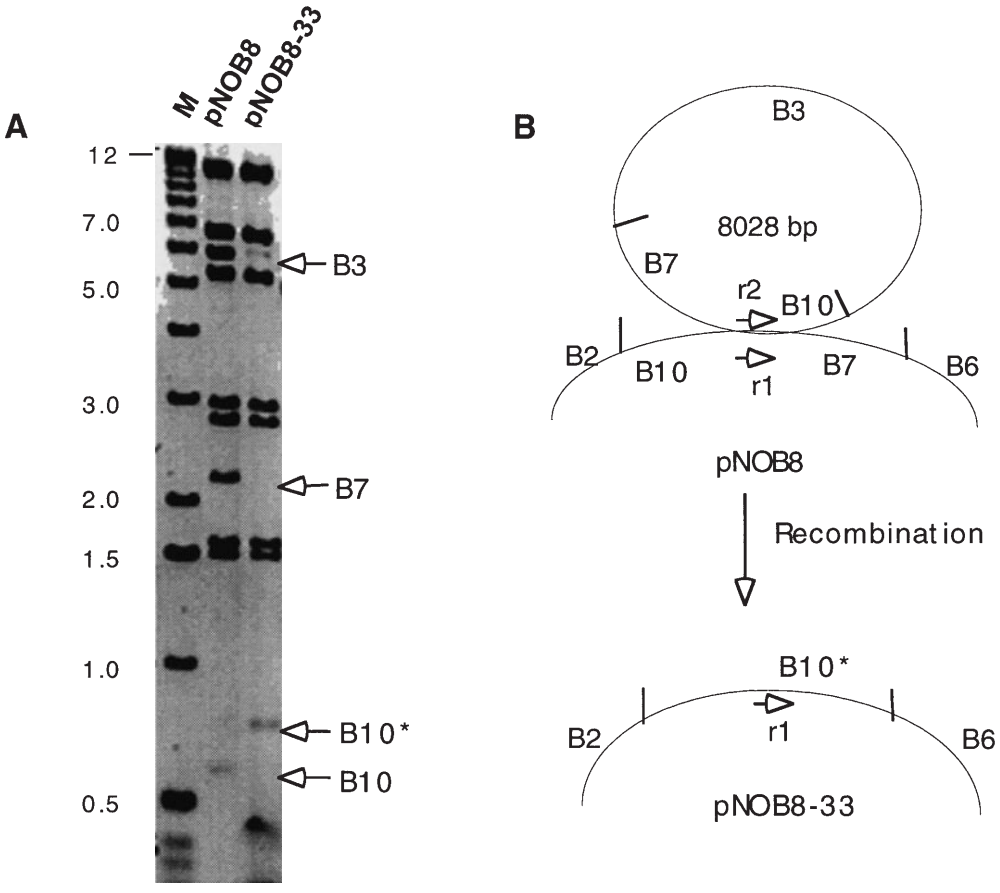

Fig. 2A,B. Formation of the major variant pNOB8-33. **A** Comparison of pNOB8 and pNOB-33 after digestion with *Bam*HI and electrophoretic separation on a 0.6% agarose gel. *B3*, *B7*, and *B10* indicate fragments exclusive to pNOB8 while *B10** is present only in pNOB8-33. *Track M* contains DNA markers with sizes indicated in kbp. **B** Illustration of the deletion event leading to the formation of the variant pNOB8-33; the altered *Bam*HI fragments (shown in **A**) are indicated, and *r1* and *r2* denote the 85-bp direct repeats

with 90% for strain NOB8-H1). These deleterious effects on growth rate and plating efficiency of transcipients of strain PH1 are less strong for cells containing pNOB8-33. This correlates with the observation that when the copy number of pNOB8 increases in NOB8-H2, the growth rate decreases (Schleper et al. 1995; data not shown).

Plasmid maintenance may be controlled by ORFs 315 and 470

There are three large ORFs in the 8-kb deleted region of ORFs 315, 470, and 413 where the latter is a transposase (see Table 1). ORF 315 shows sequence similarity to the ParA superfamily of ATPases involved in partitioning of bacterial plasmids and chromosomes (Table 1). An alignment was made for the whole protein and representative members of the ParA superfamily in which we included the proteins that have been well characterized functionally (Fig. 3). The best whole protein alignment was with RepA of *Agrobacterium tumefaciens* plasmid pTiB6S3, which showed 20% identity and 36% similarity. Alignments are shown for sections of the N-terminal half (Fig. 3) that contain the type I ATP and type II ATP/GTP-binding motifs and, also, motifs 2 and 4, which are considered to be important for the interaction of ParA and ParB in bacteria (Motallebi-Veshareh et al. 1990). We infer, therefore, that ORF 315 is a ParA homolog.

This supposition is reinforced by the observed similarity of the N-terminal region of ORF 470 (positions 1–140) to bacterial ParB proteins (Table 1). The best alignment was with RepB encoded by *A. tumefaciens* plasmid pTiB6S3, which has 14% identity and 36% similarity for amino acids 1–266. The sequence includes the conserved motifs 1 and 2, which are aligned for a selection of divergent ParB proteins in Fig. 4A. These motifs may be involved in interaction with ParA and unknown host factors in bacteria (Hanai et al. 1996). A "helix-turn-helix" DNA-binding motif also occurs in the center of ORF 470 (Fig. 4B). Although this motif belongs to a different family of DNA-binding motifs from that found in the ParB proteins, it is located at the same position in the sequence relative to motifs 1 and 2. Finally, at least one acidic domain occurs in the C-terminal half of the protein that aligns well with those found in eukaryotes (Fig. 4C). These acidic domains are important for protein–protein interactions of several eukaryotic DNA-binding proteins, such as the Myb family (reviewed by Lipsick 1996). Protein–protein and protein–DNA interactions are also a functional property of ParB proteins. Because the RepA and RepB proteins of agrobacterial and rhizobial plasmids control maintenance, as do the ParA and ParB proteins for other bacterial plasmids, it is likely that ORFs

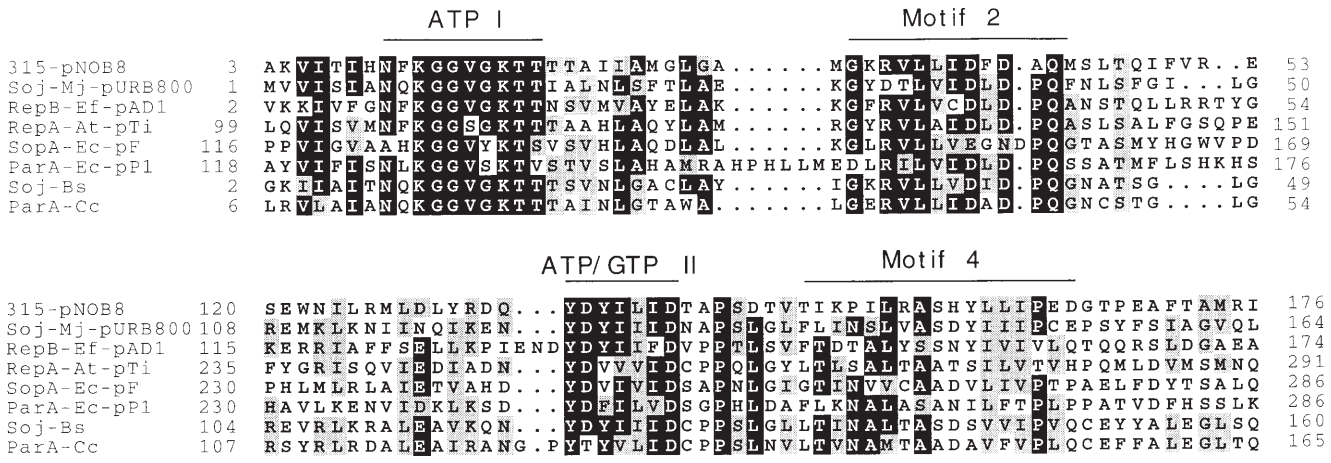

Fig. 3. Partial alignment of ORF315 and the ParA superfamily. Alignments of sections of the N-terminal half of ORF 315 and ParA proteins that contain recognizable motifs. The *ATP* motifs correspond to type I ATP and type II ATP/GTP binding sites (Motallebi-Veshareh et al. 1990). *Motifs 2* and *4* are of unknown function (Lin and Mallavia 1995). *Black background*, residues identical to ORF 315; *shaded background*, residues similar to ORF 315. *Soj-Mj-pURB800*, Soj homolog of the large extrachromosomal element of *Methanococcus jannaschii* (Bult et al. 1996); *RepB-Ef-pAD1*, RepB protein of *Enterococcus faecalis* plasmid pAD1 (Weaver et al. 1993); *RepA-At-pTiB6S3*, RepA of *Agrobacterium tumefaciens* plasmid pTiB6S3 (Tabata et al. 1989); *SopA-Ec-pF*, SopA of *E. coli* plasmid F (Mori et al. 1986); *ParB-Ec-pP1*, ParA of *E. coli* plasmid P1 (Abeles et al. 1985); *Soj-Bs*, Spo0J of *Bacillus subtilis* (Ogasawara and Yoshikawa 1992); *ParA-Cc*, ParA of *Caulobacter crescentus* (Mohl and Gober 1997)

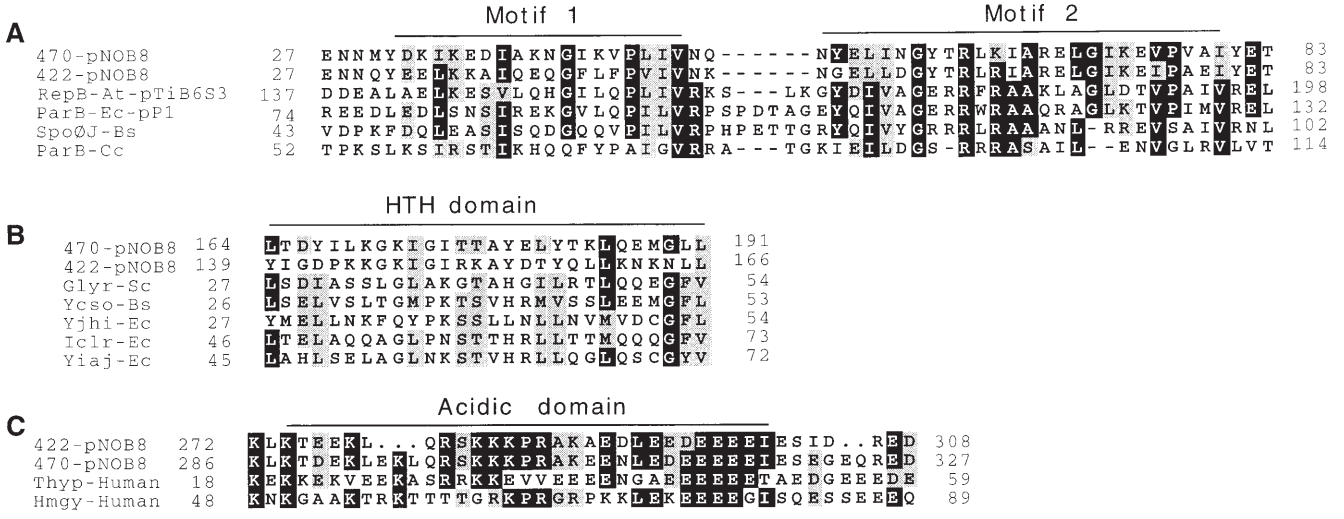

Fig. 4A–C. Partial alignment of ORF470 with ORF422 and ParB proteins. Sequence regions containing the recognizable motifs are aligned. *Black background*, residues identical to ORFs 470 and 422; *shaded background*, residues similar to ORFs 470 and 422. **A** Alignment of *motifs 1* and *2* within the N-terminal regions of ORFs 470 and 422 with ParB homologs. The selected sequences are *ParB* of *E. coli* plasmid P1 (Abeles et al. 1985), *RepB* of *A. tumefaciens* plasmid pTiB6S3 (Tabata et al. 1989), *Spo0J* of *B. subtilis* (Ogasawara and Yoshikawa 1992), and *ParB* of *C. crescentus* (Mohl and Gober 1997). *Motifs 1* and *2* are of unknown function (Lin and Mallavia 1995). **B** Alignment of the "helix-turn-helix" (*HTH*) DNA-binding domain with those of bacterial regulatory proteins. *Glyr-Sc*, regulator of glycerol operon of *Streptomyces coelicolor* (Smith and Chater 1988); *Ycso-Bs*, hypothetical transcriptional regulator in *mtld* 3′-region of *B. subtilis* (accession no., P42968); *Yjhi-Ec*, hypothetical transcriptional regulator in *feci-fimb* intergenic region of *E. coli* (Burland et al. 1995); *Iclr-Ec*, *iclr* gene product from *E. coli* (Sunnarborg et al. 1990); *Yiaj-Ec*, hypothetical transcriptional regulator in *avta-selb* region of *E. coli* (Sofia et al. 1994). The similarity was found by a BLOCK search with ORF470. **C** Alignment of the "*acidic domain*" of the C-terminal domain of ORFs 470 and 422 with higher eukaryotic DNA-binding proteins. *Thyp-Human*, parathymosin (Clinton et al. 1989); *Hmgy-Human*, high mobility group protein (Eckner and Birnstiel 1989). The similarity was found by BLOCK search and using the Multiple Sequence Alignments program at the EMBL server (Sander and Schneider 1991)

315 and 470 function similarly to the bacterial Par system in controlling the segregation and maintenance of pNOB8.

Furthermore, it is likely that two different ParBs are encoded by pNOB8 because ORF 422 aligns with both ORF 470 and the N-terminal half of the ParB protein family (Table 1; Fig. 4A); it also contains motifs similar to those found in ORF 470 (Fig. 4A–C). The alignment of ORFs 422 and 470 shows 38% identity and 53% similarity over their

Table 3. ORFs containing putative membrane-spanning and -binding segments

ORF (aa)	Location of segments
Trans-membrane	
620	264–275
604	11–28, 45–62, 80–97, 110–128, 154–171, 183–200, 216–240, 261–281, 301–318, 336–357, 361–379, 564–579
537	37–47, 383–393
439	266–275, 388–403
312	108–125
253	80–96
248	54–71, 135–152, 157–178, 213–230
246	63–83
205	62–79, 140–158
164	149–156
152	132–147
148	58–70
139	9–16
109	16–33, 59–77, 89–107
Membrane binding	
630a	265–274
315	100–111

All analyses were performed using the e-mail service of an EMBL server (Rost et al. 1995)

whole amino acid sequence, with the latter ORF carrying one main insertion, clearly indicating that they are homologs. We conclude that the presence of one ParB protein, ORF 422, and no ParA protein in pNOB8-33 accounts for its defective maintenance.

A putative *cis*-acting element may be responsible for segregation and incompatibility

ORF 248 contains a 24-nucleotide sequence that is repeated six times with a constant spacing of 39 or 42 bp and has a 10-bp imperfect palindrome in its center (boldface):

CTTTCAA**TTCTATAGTA**GATTATC

This structure resembles *cis*-acting elements such as *sopC* that cause incompatibility for some bacterial plasmids (Herman and Schneider 1992) and may, therefore, have a similar function for pNOB8. The central palindrome may resemble the inverted repeat of the *sopC* element providing a binding site for the putative ParB homologs ORFs 470 and 422. Although amino acid sequence analyses reveal ORF 248 to be one of the lower probability ORFs in pNOB8, it can generate trans-membrane fragments (Table 3), and the possibility remains that the eight amino acid repeats (FQFYSRLS) in the putative gene product may be of functional significance.

Search for gene products involved in conjugation

Bacterial conjugation systems all require several gene products for both mating pair formation and DNA transfer. Although 16 of the archaeal plasmid ORFs contain putative membrane-binding motifs, 14 of which exhibit trans-membrane helices with putative inner and outer membrane segments (Table 3), only ORFs 630b and 1025 show significant sequence similarity to any of the bacterial conjugative proteins (see Table 1). ORF 630b showed similarity to the TrbE superfamily of proteins (Table 1). The best whole protein alignment was obtained with VirB4 of *A. tumefaciens* plasmid pTiC58, which yielded 16% identity and 35% similarity for 432 amino acids at the C-terminal end. Alignments with two of the most conserved regions of divergent members of the TrbE family are illustrated in Fig. 5A, one of which is a type I ATP-binding motif. The remaining N-terminal region of the protein yielded 10% identity and 33% similarity to the VirB4 protein.

ORF 1025 showed significant sequence similarity to the TraG superfamily proteins (Table 1), and the best whole protein alignment was to TrwB of *E. coli plasmid* pR388, with 18% identity and 36% similarity. Alignments of sections of the sequence with divergent representatives of the TraG superfamily, which include type I ATP and type II ATP/GTP-binding motifs, as well as a highly conserved but functionally undefined motif 3, are illustrated in Fig. 5B. Although there are 13% gapped residues in the alignment, two main gaps accounted for most of them. We infer that ORFs 630b and 1025 are ATPases and are, at least within their C-terminal halves, homologs of the TrbE and TraG proteins, respectively. Intriguingly, the two ORFs align with each other (592 amino acids aligned; 15% identity and 32% similarity) and, moreover, the type I ATP-binding motif (Fig. 5) occurs in approximately the same position in each protein.

ORF 537 shows significant sequence similarity to a yeast Chl protein that has been identified as a DNA helicase (Table 1). However, although such activity is essential for unwinding double-stranded DNA during bacterial conjugation before DNA transfer, ORF 537 belongs to a different class of DNA helicases than the bacterial enzymes.

Insertion elements

ORF 406 is a transposase that is very similar in sequence (>75% identity) to the Ro2 elements identified in the genome of *S. solfataricus* P2 (Sensen et al. 1996) (see Table 1). However, it is not flanked by inverted repeats that could facilitate transposition. ORF 413, which is located in the 8-kb fragment that is absent from pNOB8-33, is a putative homolog of the IS256 family of bacterial transposases, with highest similarity to the transposases from *Mycobacterium* and *Rhizobium* (Table 1), and it is flanked by a 32-bp inverted repeat with six mismatched base pairs; a juxtapositioned 5-bp direct repeat may denote the insertion site.

Discussion

Given the genetic diversity among conjugative plasmids of the Bacteria, the probability of detecting homologs in pNOB8 from the archaeon *Sulfolobus* was not high. In fact,

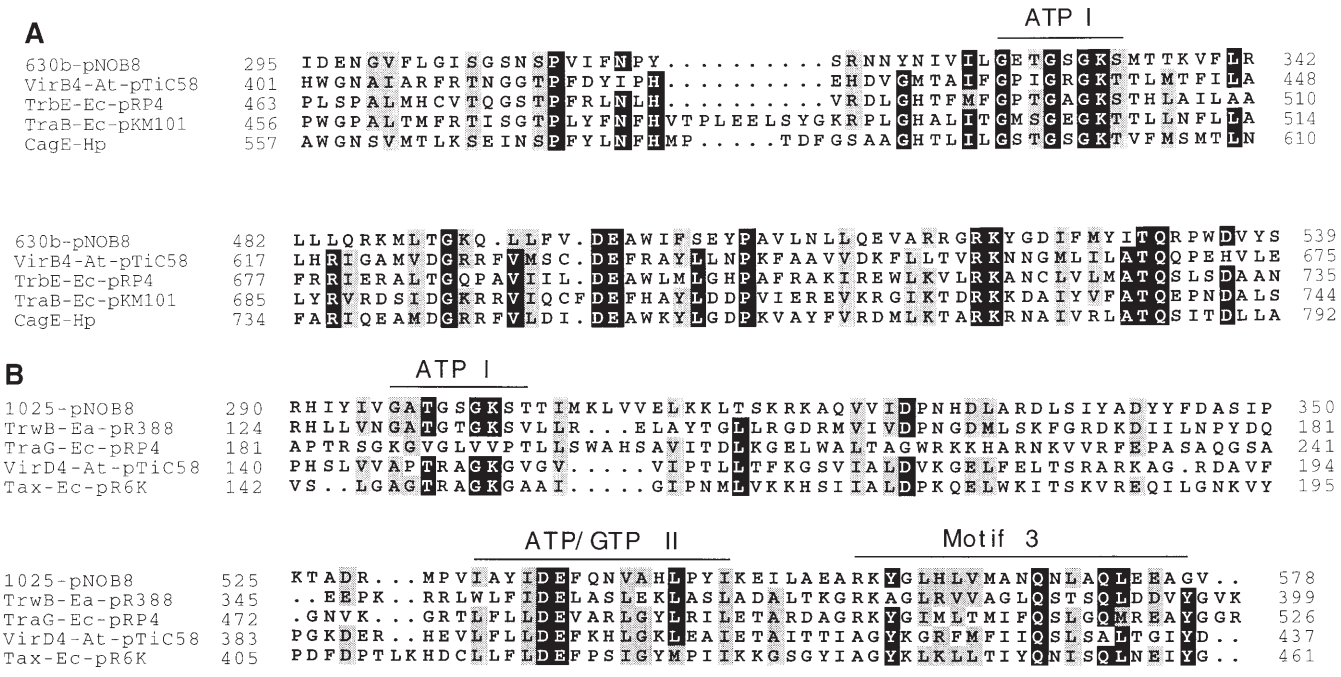

Fig. 5A,B. Partial alignment of ORFs 630b and 1025 with the TrbE and TraG superfamilies, respectively. Alignments of sections of the C-terminal halves of each protein that contain recognizable motifs or highly conserved regions. Type I ATP-binding motifs are indicated for both ORFs. ORF 1025 also carries a type II ATP/GTP site and *motif 3*, which constitutes the most conserved region of the TraG proteins but is of unknown function. *Black background*, residues identical to ORFs 630b and 1025; *shaded background*, similar residues. **A** ORF 630b is aligned with *VirB4* from *A. tumefaciens* pTiC58 plasmids (Kuldau et al. 1990), *TrbE* of IncP plasmid RP4 (Lessl et al. 1992a), *TraB* of the IncN plasmid pKM101 (Pohlman et al. 1994), and *CagE* from the island of *Helicobacter pylori* (Censini et al. 1996). **B** ORF 1025 is aligned with *TrwB* of *E. coli* plasmid pR388 (Llosa et al. 1994), *TraG* of *E. coli* plasmid pRP4 (Lessl et al. 1992b), *VirD4* from *A. tumefaciens* plasmid pTiC58 (Rogowsky et al. 1990), and *TaxB* of *E. coli* plasmid pR6K (Nunez et al. 1997)

only about 15% of the putative ORFs were assigned putative functions, including ORFs 315, 406, 413, 470, 537, 630b, 1025, and a *cis*-acting element, which enabled us to draw some important preliminary conclusions about plasmid maintenance and conjugation in the archaeon.

The genetic variant pNOB8-33 and a Par-like segregation system

Direct repeat sequences often lead to genome recombination and instability. Plasmid pNOB8 is rich in such sequences (see Table 2), which may explain why recombination occurs frequently during prolonged growth of transcipients (Schleper et al. 1995; Prangishvili et al., in press). Although the wild-type form of the plasmid is apparently favored in the parent strain, the most commonly observed variant in transcipients, pNOB8-33, lacks an 8-kb fragment as a result of recombination at the 85-bp direct repeat (see Fig. 2B). Clearly, the deleted sequence does not carry genes that are essential for plasmid replication and conjugation because pNOB8-33 is propagated and conjugated efficiently (Schleper et al. 1995). Copy number control also seems to be unaffected in pNOB8-33 as the same pattern of rapid replication followed by down-regulation occurs for both pNOB8-33 and pNOB8 after conjugation into *S. solfataricus* PH1. The decrease in copy number presumably requires an unidentified plasmid product, encoded by both pNOB8 and pNOB8-33, that is expressed during the rapid replication phase.

Because pNOB8 causes strong growth inhibition and seems to provide no competitive advantage for its host (Schleper et al. 1995), there must be an active plasmid maintenance system; otherwise, the plasmid would be lost rapidly. The presence of ParA and ParB homologs in the pNOB8 sequence, and the appearance of pNOB8-33 in *S. solfataricus* PH1, leads to the following hypothesis for a maintenance function. In parent strains, NOB8-H2 and NOB8-H1, the Par gene products, are functional such that even at a low copy number pNOB8 is stably maintained. In NOB8-H2, the deletion variant pNOB8-33 is not observed because it fails to partition properly and is rapidly lost from the culture. On the other hand, in *S. solfataricus* PH1, partitioning is defective, even for the wild-type pNOB8, presumably because a host target protein fails to recognize the Par proteins. Thus, in *S. solfataricus* PH1, neither plasmid is correctly partitioned and curing occurs. Presumably, the deletion variant pNOB8-33 is positively selected in this host because it is smaller and poses less of a burden on the cellular metabolism, leading to a slightly faster growth rate and higher plating efficiency of the transcipient (Schleper et al. 1995).

A gene homologous to the *parA* superfamily, *soj*, is also present in the archaeal plasmid pURB800 of

Methanococcus jannaschii (Bult et al. 1996). This may indicate that the Par segregation system is not confined to the Crenarchaeota but is used generally in Archaea for plasmid and, possibly, chromosome segregation.

pNOB8 conjugation

The diverse conjugative systems of bacteria share the following features: (a) adhesion of donor and recipient cells via an extracellular filamentous structure for some proteobacteria or by a fibrillar "adhesion substance" for some gram-positive bacteria; (b) covalent association of a relaxase protein to the leading 5′-end of the DNA that initiates DNA replication and transfer via a strand- and site-specific cleavage event; and (c) transfer of single-stranded DNA that is generated by rolling-circle-type replication (Pansegrau and Lanka 1996). Trans-membrane or membrane-binding segments are important for facilitating cell–cell contacts and forming membrane pores during conjugation. Although 16 ORFs from pNOB8 were predicted by topological analyses to contain such segments (see Table 3), the only potential homologs detected were ORFs 630b and 1025, which align, most significantly in their C-terminal halves, with the TrbE and TraG families of ATPases that are involved in bacterial conjugation and virulence (Fig. 5). Both ORFs contain ATP-binding motifs and may, therefore, generate energy for the conjugation process. Moreover, the two ORFs also align partially with one another, suggesting that the TrbE and TraG families may also have a common evolutionary origin. We conclude that since the TrbE and TraG proteins are the most conserved of the bacterial conjugative proteins (Lessl et al. 1992a,b; Censini et al. 1996), the bacterial and archaeal conjugation systems have evolved, at least in part, from a common ancestral system.

Common features of the *Sulfolobus* plasmids

A few *Sulfolobus* plasmids have now been isolated and partially characterized. Two small cryptic ones, pRN1 and pRN2, have been sequenced (Keeling et al. 1996) and at least two families of larger conjugative plasmids, pING and pSOG, have been described (Prangishvili et al., in press). The pING family shows a high level of genetic instability and has been sequenced in our laboratories (Stedman et al., in preparation). The results indicate that there is a significant sequence similarity of sections of the parent genome to pNOB8. The homologous ORFs, shaded in Fig. 1, include ORFs 630b and 1025 (described earlier) but not the ParA homolog ORF 315 or the putative *cis*-acting element within ORF 248 (see Table 1). Comparative sequence analyses of these plasmid families will facilitate identification of the gene products involved in conjugation and in other regulatory processes of archaeal plasmids.

Acknowledgments The Copenhagen laboratory was supported by EU grants BIO4-CT96-0488 and BIO4-CT96-0270 and by a Novo-Nordisk Foundation grant, and the Martinsried laboratory received EU grant BIO4-CT96-0488 and support from the Deutsche Forschungsgemeinschaft. K.M.S. was supported by a Marie Curie Fellowship from the European Commission. We thank Anja Schweier for help with the gel in Fig. 2A and Flemming Hansen for the Winseq program.

References

Abeles AL, Friedman SA, Austin SJ (1985) Partition of unit-copy miniplasmids to daughter cells. III. The DNA sequence and functional organization of the P1 partition region. J Mol Biol 185:261–272

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman, DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Borodovsky M, McIninch J (1993) GeneMark: parallel gene recognition for both DNA strands. Comput Chem 17:123–133

Bairoch A (1993) The PROSITE dictionary of sites and patterns in proteins, its current status. Nucleic Acids Res 21:3097–3103

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NSM, Venter JC (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 273:1058–1073

Burland V, Plunkett G III, Sofia HJ, Daniels DL, Blattner FR (1995) Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.8 through 100 minutes. Nucleic Acids Res 23:2105–2119

Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, Borodovsky M, Rappuoli R, Covacci A (1996) Cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. Proc Natl Acad Sci USA 93:14648–14653

Clinton M, Frangou-Lazaridis M, Panneerselvam C, Horecker BL (1989) The sequence of human parathymosin deduced from a cloned human kidney cDNA. Biochem Biophys Res Commun 158:855–862

Eckner R, Birnstiel ML (1989) Cloning of cDNAs coding for human HMG I and HMG Y proteins: both are capable of binding to the octamer sequence motif. Nucleic Acids Res 17:5947–5959

Hanai R, Liu R, Benedetti P, Caron PR, Lynch AS, Wang JC (1996) Molecular dissection of a protein SopB essential for *Escherichia coli* F plasmid partition. J Biol Chem 271:17469–17475

Henikoff S, Henikoff JG (1994) Protein family classification based on searching a database of blocks. Genomics 19:97–107

Herman ND, Schneider TD (1992) High information conservation implies that at least three proteins bind independently to F plasmid *incD* repeats. J Bacteriol 174:3558–3560

Keeling P, Klenk H-P, Singh RK, Feeley O, Schleper C, Zillig W, Doolittle WF, Sensen C (1996) Complete nucleotide sequence of the *Sulfolobus islandicus* multicopy plasmid pRN1. Plasmid 35:141–144

Kuldau GA, De Vos G, Owen J, McCaffrey G, Zambryski P (1990) The virB operon of *Agrobacterium tumefaciens* pTiC58 encodes 11 open reading frames. Mol Gen Genet 221:256–266

Lessl M, Balzer D, Pansegrau W, Lanka E (1992a) Sequence similarities b*et*ween the RP4 Tra2 and the Ti VirB region strongly support the conjugation model for T-DNA transfer. J Biol Chem 267:20471–20480

Lessl M, Pansegrau W, Lanka E (1992b) Relationship of DNA-transfer systems: essential transfer factors of plasmids RP4, Ti and F share common sequences. Nucleic Acids Res 20:6099–6100

Lin Z, Mallavia LP (1995) The partition region of plasmid QpH1 is a member of a family of two trans-acting factors as implied by sequence analysis. Gene 160:69–74

Lipsick JS (1996) One billion years of Myb. Oncogene 13:223–235

Llosa M, Bolland S, de la Cruz F (1994) Genetic organization of the conjugal DNA processing region of the *IncW* plasmid R388. J Mol Biol 235:448–446

Marck C (1988) "DNA Strider": a "C" program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res 16:1829–1836

Mohl DA, Gober JW (1997) Cell cycle-dependent polar localization of chromosome partitioning proteins in *Caulobacter crescentus*. Cell 88:675–684

Mori H, Kondo A, Ohshima A, Ogura T, Hiraga S (1986) Structure and function of the F plasmid genes essential for partitioning. J Mol Biol 192:1–15

Motallebi-Veshareh M, Rouch DA, Thomas CM (1990) A family of ATPases involved in active partitioning of diverse bacterial plasmids. Mol Microbiol 4:1455–1463

Nunez B, Avila P, de la Cruz F (1997) Genes involved in conjugative DNA processing of plasmid R6K. Mol Microbiol 24:1157–1168

Ogasawara N, Yoshikawa H (1992) Genes and their organization in the replication origin region of the bacterial chromosome. Mol Microbiol 6:629–634

Pansegrau W, Lanka E (1996) Enzymology of DNA transfer by conjugative mechanisms. Prog Nucleic Acids Res Mol Biol 54:197–251

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444–2448

Pohlman RF, Genetti HD, Winans SC (1994) Common ancestry between IncN conjugal transfer genes and macromolecular export systems of plant and animal pathogens. Mol Microbiol 14:655–668

Prangishvili D, Albers S, Holz I, Stedman K, Arnold HP, Klein T, Singh H, Hiort J, Schweier A, Kristjansson JU, Zillig W (1998) A family of conjugative plasmids from strains of the archaeon *Sulfolobus islandicus*. Plasmid, in press

Rogowsky PM, Powell BS, Shirasu K, Lin TS, Morel P, Zyprian EM, Steck TR, Kado CI (1990) Molecular characterization of the vir regulon of *Agrobacterium tumefaciens*: complete nucleotide sequence and gene organization of the 28.63-kbp regulon cloned as a single unit. Plasmid 23:85–106

Rost B, Casadio R, Fariselli P, Sander C (1995) Prediction of helical transmembrane segments at 95% accuracy. Protein Sci 4:521–533

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sander C, Schneider R (1991) Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins 9:56–68

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

Sensen CW, Klenk HP, Singh RK, Allard G, Chan CC, Liu QY, Penny SL, Young F, Schenk ME, Gaasterland T, Doolittle WF, Ragan MA, Charlebois RL (1996) Organizational characteristics and information content of an archaeal genome: 156 kb of sequence from *Sulfolobus solfataricus* P2. Mol Microbiol 22:175–191

Smith CP, Chater KF (1988) Structure and regulation of controlling sequences for the *Streptomyces coelicolor* glycerol operon. J Mol Biol 204:569–580

Sofia HJ, Burland V, Daniels DL, Plunkett G III, Blattner FR (1994) Analysis of the *Escherichia coli* genome. V. DNA sequence of the region from 76.0 to 81.5 minutes. Nucleic Acids Res 22:2576–2586

Sunnarborg A, Klumpp D, Chung T, LaPorte DC (1990) Regulation of the glyoxylate bypass operon: cloning and characterization of iclR. J Bacteriol 172:2642–2649

Tabata S, Hooykaas PJ, Oka A (1989) Sequence determination and characterization of the replicator region in the tumor-inducing plasmid pTiB6S3. J Bacteriol 171:1665–1672

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Weaver KE, Clewell DB, An F (1993) Identification, characterization, and nucleotide sequence of a region of *Enterococcus faecalis* pheromone-responsive plasmid pAD1 capable of autonomous replication. J Bacteriol 175:1900–1909

Zillig W, Prangishvilli D, Schleper C, Elferink M, Holz I, Albers S, Janekovic D, Gotz D (1996) Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic Archaea. FEMS Microbiol Rev 18:225–236