Extremophiles (2006) 10:327–335
DOI 10.1007/s00792-005-0502-z

ORIGINAL PAPER

Kan Yamashiro · Shin-ichi Yokobori · Tairo Oshima
Akihiko Yamagishi

Structural analysis of the plasmid pTA1 isolated from the thermoacidophilic archaeon *Thermoplasma acidophilum*

Received: 2 September 2005 / Accepted: 4 December 2005 / Published online: 22 February 2006

Abstract *Thermoplasma acidophilum* is a thermoacidophilic archaeon that grows optimally at pH1.8 and 56°C and has no cell wall. Plasmid pTA1 was found in some strains of the species. We sequenced plasmid pTA1 and analyzed the open reading frames (ORFs). pTA1 was found to be a circular DNA molecule of 15,723 bp. Eighteen ORFs were found; none of the gene products except ORF1 had sequence similarity to known proteins. ORF1 showed similarity to Cdc6, which is involved in genome-replication initiation in Eukarya and Archaea. *T. acidophilum* has two Cdc6 homologues in the genome. The homologue found in pTA1 is most similar to Tvo3, one of the three Cdc6 homologues found in the genome of *Thermoplasma volcanium*, among all of the Cdc6 family proteins. The phylogenetic analysis suggested that plasmid pTA1 is possibly originated from the chromosomal DNA of *Thermoplasma*.

Keywords Archaea · Cell division protein 6 · Cdc6 · Phylogenetic tree · Plasmid · pTA1 · *Thermoplasma acidophilum*

Abbreviations ORF: Open reading frame · Cdc6: Cell division control protein 6 · RBS: Ribosome binding site · NCR: Non coding region · PKD: Polycystic kidney disease · TFIIB: Transcription factor IIB · BRE: Transcription factor B recognition elemen

Communicated by F. Robb

K. Yamashiro · S. Yokobori · T. Oshima · A. Yamagishi (✉)
Department of Molecular Biology,
Tokyo University of Pharmacy and Life Science,
1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan
E-mail: yamagish@LS.toyaku.ac.jp
Tel.: +81-426-767139
Fax: +81-426-767145

Introduction

The Archaea represent one of the three major divisions of life (Woese et al. 1990). In spite of their prokaryotic morphology, the Archaea share similarities with Eukarya (Brown and Doolittle 1997). In order to study evolution of life and to harness the Archaea as a simple model for the more complex Eukarya systems, it is essential to study biochemistry and genetics in the Archaea. Identification of extra-chromosomal elements, such as plasmids and viruses can provide information needed for the development of genetic tools.

Thermoplasma acidophilum is a thermoacidophilic Archaeon isolated from a self-heating coal refuse pile. The archaeon grows optimally at pH 1.8 and 56°C, is classified in the Euryarchaeota, and has no cell wall (Darland et al. 1970). Temperature shifts cause a change in the cell shape, from irregular to spherical, suggesting the presence of an internal cytoskeleton (Searcy et al. 1981). The genome of *T. acidophilum* has been completely sequenced and was 1,564,905 bp, which is one of smallest genomes in the Archaea (Ruepp et al. 2000).

Plasmid pTA1 was found in *T. acidophilum* strains isolated from hot springs in Japan (Yasuda et al. 1995). Here we report the complete nucleotide sequence of pTA1 and the identification of eighteen open reading frames (ORFs). The product of ORF1 showed similarity to Cdc6, a genome-replication initiation protein found in Archaea and Eukarya. These results suggest that plasmid pTA1 may have originated from chromosomal DNA.

Materials and methods

Materials, strains and plasmids

The type strain of *T. acidophilum* was kindly provided by Dr. D. G. Searcy. *T. acidophilum* HO-122 was previously

isolated in our laboratory (Yasuda et al. 1995). *T. acidophilum* culture medium was prepared as described previously (Yasuda et al. 1995). All primers were obtained from Sigma Genosys (Hokkaido, Japan). 5′-[α−^{32}P]-dATP was purchased from Amersham Biosciences (Tokyo, Japan). pTA1 plasmid DNA was isolated by the alkaline lysis method from *T. acidophilum* HO-122 cells harvested after a 70-h incubation at 56°C (Sambrook et al. 1989). *Escherichia coli* was grown in LB medium (Sambrook et al. 1989).

Cloning and sequencing of pTA1

Plasmid pTA1 was digested with *Eco*RI, *Hin*dIII or *Hin*cII. Each fragment was cloned between appropriate sites of pUC118. Both strands of each fragment were sequenced by the primer walking method. Correct assembly of these fragments was checked by amplifying the regions around the restriction sites by PCR. The amplified fragments were cloned and sequenced.

Plasmid DNA was purified from *E. coli* strains with the QIAprep Miniprep Kit (Qiagen, Tokyo, Japan). Sequencing reactions were performed with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Tokyo, Japan). The products were applied to an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Tokyo, Japan).

Sequence analysis

Identification of ORFs and repeated sequences was carried out with the computer program GENETYX-MAC Ver.11.0 (Genetix Corporation, Tokyo, Japan). The nucleotide and protein sequences and conserved-domains were searched with the NCBI database using the programs BLAST and CDART, respectively.

For phylogenetic analysis, sequences were retrieved from protein databases in NCBI. The sequences of Cdc6 homologues were aligned using Clustal X version 1.81 (Thompson et al. 1997), and subsequently edited manually. The well-aligned sites (202 positions) were used for tree construction. A phylogenetic tree inferred from the maximum likelihood method was constructed using PROML with the JTT amino acid substitution matrix (Jones et al. 1992). Global rearrangements and randomized input order of sequences (10 jumbles) and a rate heterogeneity model with gamma distribution parameter alpha (1.36) calculated by treepuzzle−5.0 (Schmidt et al. 2002) were used. Bootstrap values represent a consensus of 1,000 Neighbor distance tree obtained by CONSENSUS and NEIGHBOR. The pseudo-replicates were generated with SEQBOOT and the distance was calculated by PROTDIST. The setting in Bootstrap trial was the same as those for used for PROML. PROML, SEQBOOT, PROTDIST, NEIGHBOR in the PHYLIP package version 3.62 (Felsenstein 1989) were used.

Northern analysis

Total RNA was extracted with the RNA extraction kit ISOGEN (Nippon Gene Co., Ltd. Toyama, Japan), purified with an RNeasy spin column (Qiagen, Tokyo, Japan), and treated with DNase I (Takara Bio Inc., Tokyo, Japan).

DNA fragments for each target region were amplified by PCR with each ORF-specific primer pair and pTA1 as template. Each fragment was cloned into the *Sma*I site of pSPT18 and used for the preparation of each RNA probe. RNA probes were labeled with digoxigenin by using a DIG RNA Labeling Kit SP6/T7 (Roche Diagnostics, Tokyo, Japan). Total RNA from the *T. acidophilum* type strain or strain HO-122 was separated on 1.5% agarose gels containing 6.7% formaldehyde, and then transferred onto Hybond-N+ membranes (Amersham Pharmacia Biotech, Tokyo, Japan). DIG Easy Hyb solution (Roche Diagnostics, Tokyo, Japan) was used for the hybridization at 58°C, except for ORF6 and ORF16, which were hybridized at 55°C. Hybridization signals were detected with a DIG Nucleic Acid Detection Kit (Roche Diagnostics, Tokyo, Japan) according to the manufacturer's standard protocol.

Primer extension analysis of ORF1

The 5′-end of the ORF1 transcript was determined with a Primer Extension System-AMV Reverse Transcriptase kit (Promega, Tokyo, Japan) using 5′-[α−^{32}P]-dATP-labeled oligonucleotides as primers. Primers used were pe1 (CCTTAAGGTTGTTCCCGTTGTCCT), pe2 (CC AGTTGTTCCAGGATAGCGTTGA) and pe3 (CTC -GTCAGTTCGTCCAGGTTGTTT). After purification with a chroma spin+TE-10 (Clontech, Tokyo, Japan), all samples were separated on a 8% polyacrylamide sequencing gel containing 7 M urea, and analyzed with an imaging plate and a BAS 3000 system (Fuji Film Ltd., Japan).

Results

Structural features of the pTA1 sequence

The complete sequence of the pTA1 plasmid from *T. acidophilum* HO-122 was determined (Fig. 1) (GeneBank accession no. AB190359). It was a circular plasmid with 15,723 bp of DNA and a G+C content of 46.7%, which is similar to that of *T. acidophilum* chromosomal DNA (46.0%) (Ruepp et al. 2000). The sequence was numbered from the first base of a *Dra*I site. The overall DNA sequence of pTA1 had no regions related to the other archaeal plasmids or to genome sequences of *T. acidophilum* or *T. volcanium* except ORF1 and 2; thus pTA1 appeared to be a novel plasmid.

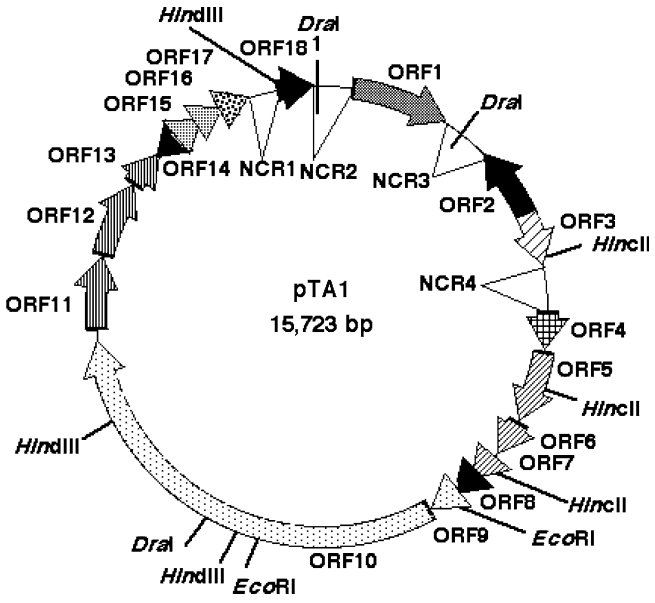


Fig. 1 Map of pTA1. Eighteen ORFs are shown by *arrows*. NCR1, 2 and 3 represent intergenetic regions near ORF1. Restriction recognition sites used in subcloning and the *Dra*I site at base one are indicated. Putative ORFs in the same multicistronic operons are indicated by *arrows* with the same pattern

Open reading frames in pTA1

Sequence analysis identified eighteen ORFs that were more than 100 amino acids in size (Fig. 1). The consensus sequence of the TATA box in Archaea is TTTAt/aATA (Reiter et al. 1988). Sequences resembling the TATA box were found upstream between −31 and −276 bp of all ORFs except for ORFs 13 and 16 (Fig. 2a). In all of these promoters, the last four nucleotides were highly conserved (t/aATA). Stem-loop structures and thymidine-rich regions, which are potential transcription termination regions (Brown et al. 1989), were found downstream of ORFs 1, 2, 3, 4, 7, 8, 10, 13, 14, 16, 17 and 18. Some ORFs had a short overlap or a very short spacer region.

ORF1 seemed to encode a 41.5-kDa protein that has a motifs of ''ATPases associated with various cellular activities'' and ''winged-helix'', and showed sequence similarity to cell division control protein 6 (Cdc6) (Fig. 3). ORF1 has a class II winged-helix motif, suggesting direct DNA binding (Capaldi and Berger 2004; Singleton et al. 2004). Cdc6 is known as a genome replication initiator in Archaea and Eukarya (Giraldo 2003).

In general, Rep protein plays a role as the replication initiator for plasmids. A rolling-circle replication mechanism has been postulated for many plasmids in Archaea as well as in Bacteria (Erauso et al. 1996). However, there was no homologue of Rep protein on pTA1. Accordingly, ORF1 of pTA1 is likely to be a replication initiator.

Northern analysis of the transcripts

Since some ORFs were overlapping or had short intergenic sequences, they may constitute polycistronic operons (Fig. 2c). Depending on the location of the terminators, ORFs 5–7, 9,10, 11–13 and 15–16 may constitute multicistronic operons.

In order to investigate expression of ORFs in pTA1, northern analysis was carried out. No signal was detected with any of the probes when total RNA from the type strain, which has no plasmid, was used as the target (Fig. 4a).

The results are summarized in Table 1. For ORF1, the length of the signal detected was similar to the length of ORF1 (Fig. 4a). The ORF2 specific signal was shorter than ORF2. Therefore, the actual start codon of ORF2 may be located downstream of the predicted start. Additional ATG and GTG codons are present about 100 bp downstream of the predicted initial codon. However, no promoter sequences were found upstream of the alternative first codons. Accordingly, the true initiation codon of ORF2 is not clear. Specific signals for ORFs 3, 8, 9, 10, 11, 12, 13, 14, 16 and 18 were not detected. These ORFs are scarcely transcribed under the conditions used in our analysis. Two signals each were detected for ORFs 4 and 17. When hybridization stringency was increased, all of the signals disappeared. The longer signals are long enough to encode ORFs 4 and 17, respectively, and the shorter signal may result from degradation products or an alternative transcript of the longer mRNA. Since a 1.9 kbp signal was detected by ORF5, 6 and 7 specific probes, these ORFs are likely to constitute a polycistronic operon. Although the position of promoters and terminators suggested that ORFs 15 and 16 might constitute a polycistronic operon, specific signal was detected only with the ORF15-specific probe. Accordingly, the operon structure of ORF15 and 16 is not clear.

The northern analysis suggests that promoters in front of ORF1, 2, 4, 5, 15 and 17 are active under the conditions. According to the alignment of putative promoters (Fig. 2a), it is likely that more than six conserved bases in putative promoter are required for transcription. However, the failure of detection of transcripts with well conserved promoter sequence (e. g. ORF10 and 14) suggests the involvement of other factors in transcription.

Primer extension analysis of ORF1

In the 5′-region of ORF1, there are two candidates for the first codon (Fig. 4b). We determined the 5′-terminus of ORF1 by primer extension analysis. A specific band corresponding to the G complementary to the *C is visible at the position indicated by the arrow in Fig. 4c. This transcription start site was located 2 bp upstream of the second ATG. A putative promoter and RBS were

a

```
ORF1   :(-31)  AACACGTAAAAGAA aTTAAATA CAACAATGACATTAATCCAATCAATGATAAG
ORF2   :(-103) GTCAGGTCTGGCGG aTgATATA CTCTGGGGGCGGTCCTGTTGTCCTGGCCATC
ORF3   :(-127) GCACATCCAACCGC TcacTATA AGCATTCCTCACATCGATGAACTGATAGGGC
ORF4   :(-276) CAAAATATGAATAC TTTtAATA AAAGTATGCCTTCACTATCCAAATGAAGAAC
       :(-29)  CGGGCCCAGCCCTA cTcATATA TTACCAGAAGGAAAAGGGAAGATGGACCAGG
ORF5   :(-71)  AGGCAAACAGAAAC aaTATATA CATAGCGATCACAATACAAGGGAAGCAATAT
ORF6   :(-136) TCCATGGACGCACC TgcATATA AACCCAATGGGCTACATCCCATCACGCAGGA
ORF7   :(-22)  GAGCCACAATAAAA ggTATATc AGGAGGGGATAAAGATGATAGGAATCACAGG
ORF8   :(-31)  AGTGGGGATCGAAA gagATATA CTGGGAGGAGAAGAAGATCTGTAATGGAGGA
ORF9   :(-235) CCGAGATCCTCCAA TcTATaTA TTCTTTCATCAACAGCATAATGGCCATTTTT
ORF10  :(-17)  CAGTAGTTGATACT TTTATATA GGTGATATTATGAAAAAAGCAAAACTAATAA
ORF11  :(-91)  AGCCGTATTCATCA TcatTATA GGATTCATAGCCCTATACAGGATGAACATCA
ORF12  :(-67)  TCGCAGACATCCCG aTatTATA CCTGGGAGTGAGCGGGCTGATCAAGACCCAC
ORF14  :(-131) AATGTTCGAATTCT TTTATATt TTTGCTTTCAGCGATCCATGTCGGGATCTTC
ORF15  :(-197) AGACATAATAGTAT gaTAAATA TGAAAAGGGATATTGCAATGGCCGACCCAAA
ORF17  :(-96)  GGACCTCGAAAACA TcTATAaA CGCATGAGAATCCTATACTACTGCCGGAATT
ORF18  :(-73)  ACCATTCTTATTAT TaTtAATA AGAACAACTCTTAATATTAATAAGAAACCAA
CON    :                   TTTAWATA
```

b

ORF1
taaTCCCCCGATACCGAAATAAAGGGGGATACCGC
ATTACACGGCATAAATCATCATTCCACACACAACC
ACGAATACGGAAAACCCAAATTCCAA***TTTTATATA TGATAAATTT***AGA

ORF2
taaGTGATTCTCTTTTTCTCAT***TTTCTCTTTCTCT TTCC***AGTGTGG

ORF3
tgaGAAGACTATCTTATATCCATCAAAGGCCATGA
GAAGATAA***TTCATC***AATTGAGACTTACCGGACCCG

ORF4
tgaATGAACGATAAACCATGCAAAGACCGCATAAA
GAATCGAGGAAGAGCAAGGGAAAGCGGTTTTTCAT
CGGTGTAGCATTCGCCGTAA***TTTTT***GTGATCG

ORF7
taaTGGAGGACCATCATGGCCAGCTTCACAACCGA
GATCCTCCAA***TCTATTTATTCTTTCATC***AACAGCA

ORF8
tgaACCCCCTGAACTGGATCAAGGACATATCGCAG
GCAACCATCAACCAGATCCAGAATGCCA***TCCTATA CCTTTTCC***AGGAATT

ORF10
taaTGGGGAAAATAGTAGATACAGCAATAGCCGTA
TT***CATCATCATTAT***AGGATTCATA

ORF13
tgaAGGAGCAGAACGAACATTATCCAAAGTAATAA
AAAACGGCAAAAGCCGAAAACGGAGGGCCGGATCC
GCCACCAGGGCAGGCCTCCAAAAACGGCAAACAAC
GAACCCAA***TATTTTT***AACAACC

ORF14
tgaATTTCTCGTGTTATCCTTCCTTGGGGGATGGA
ATATATCGTATAGTGGCG***TCTTTGTGTCTTCCT***GG

ORF16
ttaCAAGAACCCCCACCAAAAAGAAATTGATGGAC
CATTAA***TTTTTTCCCC***AACCAT

ORF17
ttaTATGCCATACTATTA***TTATATCTAT***GAAAAAT

ORF18
tgaAGATACACTTTATCGCAAAAACGAAATAAAAT
TAA***TTTTTTT***CAAACGGAACCAAGA

c

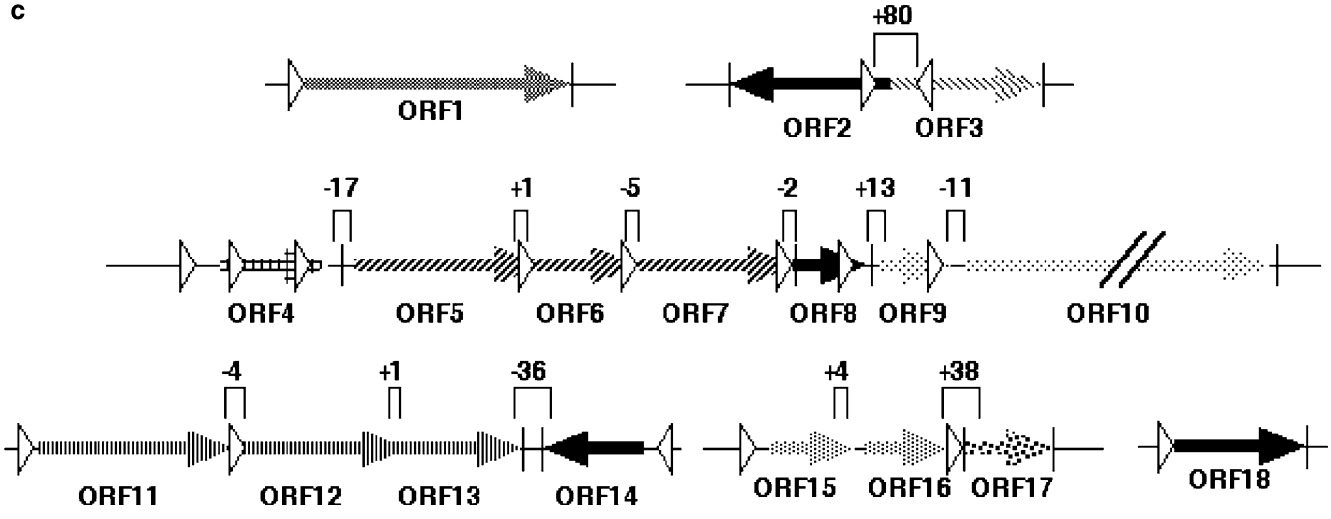


Fig. 2 a-c Sequences of putative promoters and terminators in pTA1. **a** Putative promoter sequences. A *box* indicates putative promoter TATA box sequences in ORFs in pTA1. Putative first codons are in *bold*. *Figures at the far left* indicate the promoter position of the first base numbered from the first codon. *CON* is the consensus promoter sequence of the archaeal TATA box (Reiter et al. 1988). Mismatches are indicated in *lowercase letters*. Base C at the transcriptional initiation site that was determined by primer extension analysis of ORF1 is *double underlined*. ORF4-2 shows an alternative promoter found near the second codon. **b** Putative terminator sequences. *Italic bold letters* indicate the pyrimidine-rich regions. The inverted repeats are underlined. *Lowercase letters* are stop codons of ORFs. **c** Map of promoters and terminators. Overlaps and gaps between ORFs are shown by positive (+) and negative (−) figures, respectively. *Arrow* ORF; *White triangle* promoter; *Vertical line* terminator. Putative ORFs in the same multicistronic operons are indicated by *arrows* with the same pattern

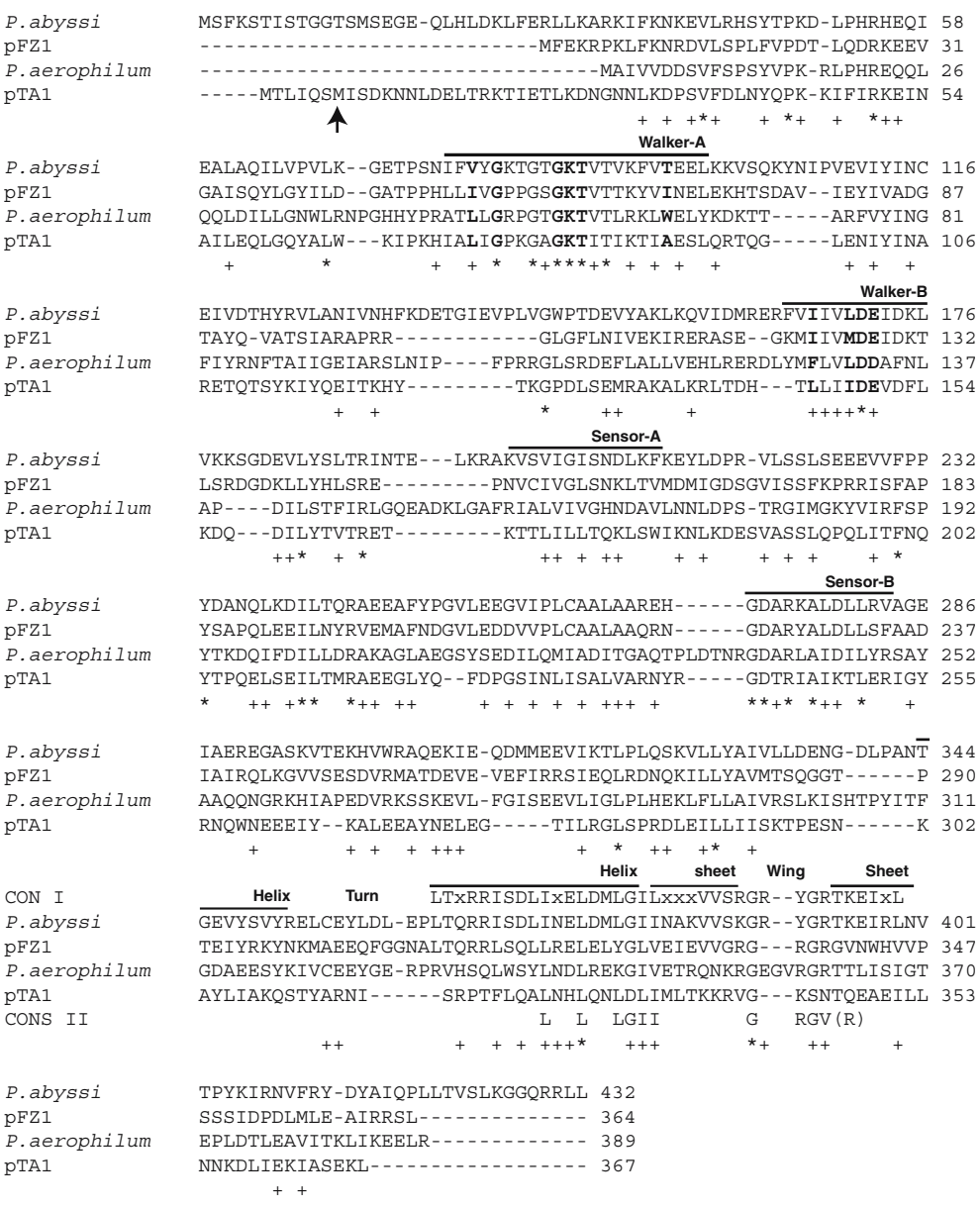


Fig. 3 Alignment of Cdc6. ATP-binding motifs and a winged-helix motif are indicated by *upper lines* (Liu et al. 2000; Singleton et al. 2004). An *arrow* shows the first methionine after the transcription initiation site determined by primer extension analysis. CON I and CON II are consensus sequences in class I and class II winged-helix motifs, respectively (Singleton et al. 2004). Upper two and lower two homologues of Cdc6 have class I and class II winged-helix motifs, respectively. *Numbers* at the right refer to amino acid numbers. Accession numbers of the sequences are *Pyrococcus abyssi* (PAB2265), *Pyrobaculum aerophilum* (PAE0737) and plasmids pFZ1 (NP 039767) from *Methanobacterium thermoformicicum*

found at −31 bp and at +2 bp of the start codon, respectively (Fig. 4b). Based on this information, the second ATG likely to be the real start codon.

Phylogenetic analysis of Cdc6 in pTA1

To estimate the relationship between ORF1 of pTA1 and Cdc6, a phylogenetic tree was derived from an alignment of Cdc6 homologues from Archaea and Eukarya. We retrieved Cdc6 homologues from the NCBI database through similarity searches with BLASTP using reported homologues as the query. We omitted some homologues that were very short or lacked the conserved domain. In the Thermoplasmaceae, two and three homologues were found in *T. acidophilum* and in *T. volcanium*, respectively.

For construction of our tree, the sequences of the Cdc6 homologues in the Archaea were aligned and five representative from Eukarya was included as the outgroup (Fig. 5). Some points are noted for the phylogenetic tree of Cdc6. First, the tree has two large branches (Group A, B). Group A contains Cdc6 homologues of three *Pyrococcus* species and a plasmid pFZ1. No Group B homologues were found in these species. Homologues were found in both groups for the species that have more than one homologue. However, Cdc6 of *P. aerophilum*, which contains only one homologue, was a member of Group B. Second, Eukarya branched at the border of the two groups. It is likely that group A and B diverged in early evolutionary history, since these groups are rooted by the Eukaryal Cdc6 branch. Third, this branching is concordant with the classification based on winged-helix motif sequence (Singleton et al. 2004):

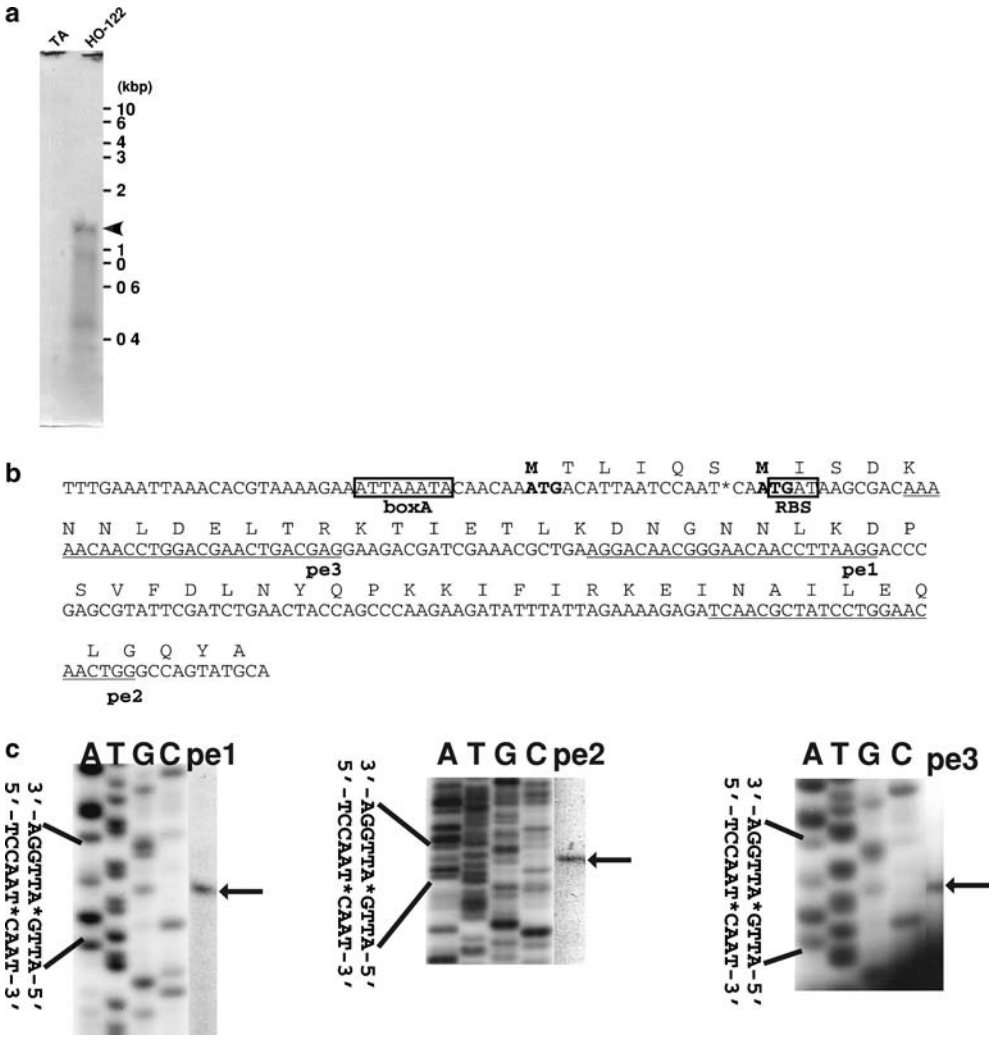


Fig. 4 **a**–**c** Northern analysis and 5′-end mapping of ORF1 transcript by primer extension analysis. **a** Northern analysis of ORF1. Total RNA from type strain or strain HO-122 was analyzed in lane TA or HO-122, respectively. Marker positions are shown on the *right*. *Arrow* indicates signal position. **b** 5′-teminal region of ORF1. The RBS and TATA box are shown by *boxes*. The 3′-sequence of 16s rRNA in *T. acidophilum* is 5′-TGGATCACCTCC-3′ (Ree et al. 1989). *Underlines* indicate primers (pe1, 2 and 3, see Materials and methods) used in primer extension analysis. *Asterisks* indicate the transcriptional initiation site. **c** Autoradiogram of the electrophoresis gel loaded with primer extension products. A, T, G and C: DNA sequencing ladders. pe1, 2 and 3: primer extension reaction products. Transcription initiation sites are indicated by *asterisks*. Possible start codons are in *bold*. *Arrows* indicate primer extension signals

Table 1 Northern analysis of pTA1 ORF transcripts

ORF	Promoter[a]	Terminator[a]	ORF length (bp)	Signal (kbp)[b]	Operon length (bp)[c]
1	-31	+3	1,101	1.4	1,101
2	-103	+2	834	0.74	834
3	-126	+12	600	ND	600
4	-276 (−29)[d]	+15	453 (399)[d]	1.0, 0.68	453 (399)[d]
5	-71		813	1.9	1,623
6	-136		459	1.9	
7	-22	+5	351	1.9	
8	-31	+14	291	ND	291
9	-235		306	ND	5,154
10	-17	+18	4,848	ND	
11	-91		831	ND	2,112
12	-67		837	ND	
13		+57	441	ND	
14	-131	+14	333	ND	333
15	-197		306	0.8, 0.6	543
16		+21	237	ND	
17	-97	+1	375	0.45, 0.69	375
18	-32	+5	441	ND	441

ND Specific signal was not detected

[a] The position of putative promoters and terminators from the first and the last codon, respectively

[b] Length of the signal detected by northern analysis of whole RNA from *T. acidophilum* HO-122

[c] Expected length of polycistronic operons. The length between the first codon of the first ORF and the stop codon of the last ORF are given

[d] Promoter found near second ATG, and ORF and operon length expected from the second ATG

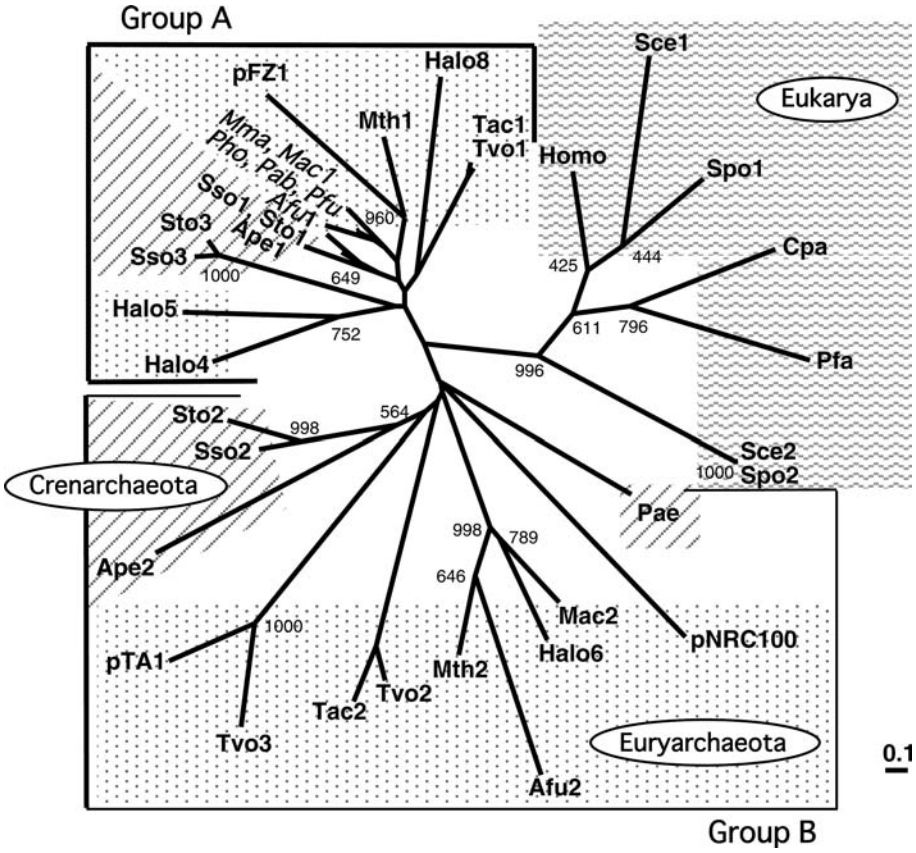


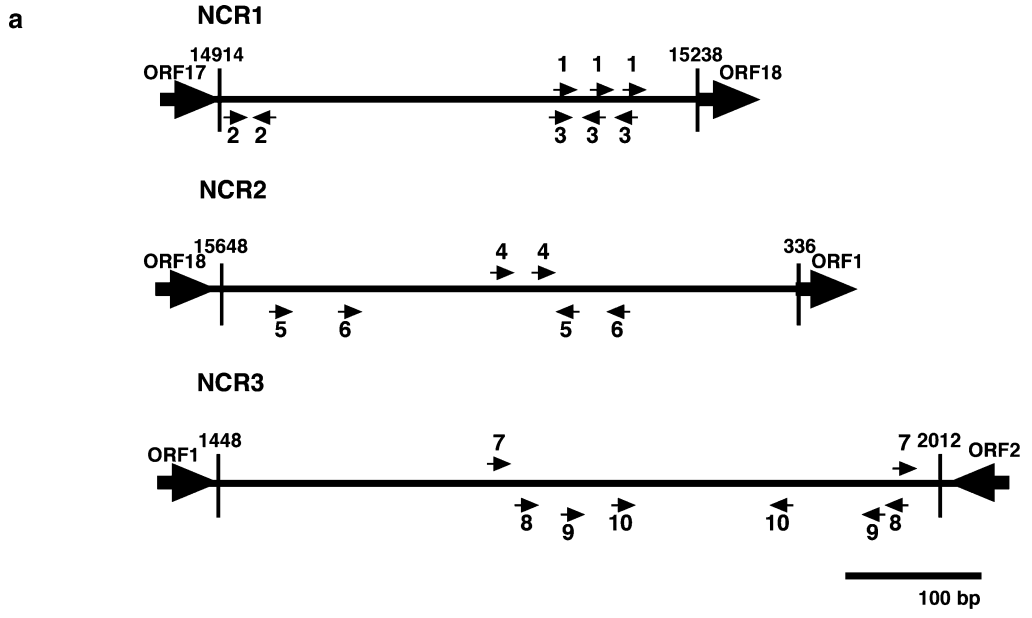
Fig. 5 Phylogenetic tree of Cdc6 and Cdc6 homologues. A *scale bar* represents 0.1 substitutions per residue. The numbers next to the nodes refer to the bootstrap values (out of 1,000) that support the branching pattern. Bootstrap values lower than 400 are not shown. Sequence accession numbers: Homo, *Homo sapiens* NP 001245; *Sce*1 and 2, *Saccharomyces cerevisiae* P09119 (Cdc6) and AAC49129 (Orc1); *Spo*1 and 2, *Schizosaccharomyces pombe* CAA18425 (Cdc6) and CAA22443 (Orc1); *Pfa*, *Plasmodium falciparum* PFL0150w; *Cpa*, *Cryptosporidium parvum* EAK87413; *Afu*1 and 2, *Archaeoglobus fulgidus* AF0244 and AF0695; *Ape*1 and 2, *Aeropyrum pernix* APE0475 and APE0152; Halo4, 5, 6 and 8, *Halobacterium* sp. NRC-1 VNG6363G, VNG6272G, VNG2271G, VNG1224G; *Mth*1 and 2, *Methanothermobacter thermautotrophicus* MTH1412 and MTH1599; *Mac*1 and 2, *Methanosarcina acetivorans* MA0001 and MA0085; *Mma*, *Methanosarcina mazei* MM1314; *Pab*, *Pyrococcus abyssi* PAB2265; *Pae*, *Pyrobaculum aerophilum* PAE0737; *Pfu*, *Pyrococcus furiosus* PF0017; *Pho*, *Pyrococcus horikoshii* PH0124; *Sso*1, 2 and 3, *Sulfolobus solfataricus* SSO0257, SSO0771 and SSO2184; *Sto*1, 2 and 3, *Sulfolobus tokodaii* ST0305, ST0471 and ST2165; *Tac*1 and 2, *Thermoplasma acidophilum* TA0636 and TA0451m; *Tvo*1, 2 and 3, *Thermoplasma volcanium* TVN0758 and TVN1044, TVN1477; plasmids pFZ1 NP 039767 from *Methanobacterium thermoformicicum*

Group A and B homologues, respectively, have Class I and Class II consensus winged-helix motif sequences. Fourth, in general the branch length of group A was shorter than that of group B. Fifth, the subdivision between euryarchaeota and crenarchaeota groups was not clearly supported. Finally, the Cdc6 homologue of pTA1 was closely related to that of *T. volcanium Tvo*3 with a high bootstrap value.

Estimation of the replication origin of pTA1

Repeated sequences are often targeted by multimeric proteins, or they can form secondary structures, such as a stem-loop. In some Archaeal species, *ori*C sequences with some repeats were reported (Lopez et al. 1999; Matsunaga et al. 2003). Non coding regions near ORF1 were named NCR 1 (n.t. 14914 to 15238), NCR 2 (n.t. 15648 to 366), NCR 3 (n.t. 1448 to 2012) and NCR 4 (n.t. 3414 to 3912) (Fig. 1). pTA1 contained three groups of direct repeats (1, 3, 4) and seven inverted repeats (2, 5–10) in the NCRs. These repeated sequences are more than 13 bp in length and have homology higher than 90% within each group (Fig. 6). Although there was no direct or inverted repeat, a 19-bp palindrome (ATACTTTTAa/tTAAAAGTAT) was found in NCR 4.

NCR 1 has three tandem direct and inverted repeats (group 1 and 3). Because of the palindromic sequence of these repeats, these multiple repeats overlapped. Multiple repeats appear in the origin of well-known plasmids in some bacteria (Kues and Stahl 1989). Plasmid pNRC100 and the chromosomal replicons of halophilic Archaea also contain tandem repeats (Mojica et al. 1995; Ng and DasSarma 1993). Some groups of repeated sequence have mutual similarity between groups. Groups 1, 3 and 4 shared an AT-rich 10 bp sequence, and Groups 6 and 9 share a similar 10 bp sequence. Con-



b

1:	15159-	TATTATTATTAAtAA	-15173	**2:**	14907-	CAAAGATATAATA	-14919
	15184-	TAaTATTAaTAAGAA	-15198		14943-	CAtAGATATAATA	-14931
	15214-	TATTATTATTAAGAA	-15228				
3:	15155-	TTCTTATTAtTATTA	-15169	**5:**	15678-	TAAAATTAATTTTT	-15691
	15198-	TTCTTATTAATATTA	-15184		200-	TAAAATTAtTTTTT	-187
	15228-	TTCTTAaTAATAaTA	-15214				
4:	129-	ATTAATATATTAA	-141	**7:**	1643-	CTGAAAAGAGAAAGAGAAA	-1661
	172-	ATTAATATAaTAA	-184		1975-	CTGgAAAGAGAAAGAGAAA	-1993
6:	17-	AcAGGGGGGACATTTAACGACA	-38	**8:**	1682-	CTCTTTCCAGTAT	-1694
	252-	AGAGGGGGGACATTTtACcACA	-231		1984-	CTCTTTCCAGTgT	-1972
9:	1722-	GGGGAAATTTGCGACGG	-1739				
	1973-	GtGGAAATTTGCGACGG	-1957	**10:**	1758-	TTGCTTCGGTAGAGGAC	-1774
					1897-	TcGCTTCGGTAGAGGgC	-1881

Fig. 6 **a, b** Direct and inverted repeats found in NCR1, 2 and 3 and the genomes of *T. acidophilum* and *T. volcanium*. **a** Direct and inverted repeats in NCR1, 2 and 3. Repeats of 13 bp and longer are shown by *arrows*. *Figures near the arrows* indicate group numbers of repeats. *Arrows shown above and below the lines* indicate direct and inverted repeats, respectively. *Numbers* at both ends of intergenic regions represent nucleotide positions. **b** Sequences of repeats, where group numbers at the left correspond to those in **a**. Figures at the end of the sequences represent positions in the plasmid pTA1. Mismatches are shown in *lowercase*. Conserved nucleotides among different groups are *underlined*

served repeated sequences are often recognized by DNA binding proteins. The presence of a 13 bp consensus repeat (t/cTa/cCAg/cTGGAAAT) has been postulated in the *ori*C of Archaea (Lopez et al. 1999). Recently, a 34 bp long inverted repeat sequence including the 13 bp consensus repeat was found in the replication initiation site of *Pyrococcus* strains (Matsunaga et al. 2003). In pTA1 NCR regions, 7 bp in Group 8 matched with this consensus repeat (TTCCAGT).

Discussion

The sequence of plasmid pTA1, which was isolated from the thermoacidophilic Archaeon *T. acidophilum*, was determined. According to similarity searches, ORF1 had high sequence similarity to the replication initiator Cdc6 protein, which suggests a close correlation between pTA1 and the genome. Because many Archaeal plasmids use the rolling circle replication mechanism (Erauso et al. 1996), there is an alternative possibility that pTA1 is replicated by a rolling mechanism using a novel Rep protein. However, single stranded intermediates of rolling circle replication were not detected by Southern analysis (data not shown).

By phylogenetic analysis, homologues of Cdc6 were separated into two groups rooted by a branch from Eukarya (Fig. 5). The Cdc6 homologues present in species that contained more than two homologues typically had representatives of both groups. Most species that contained only one homologue had a group A member, except *P. aerophilum*. The Cdc6 homologue

encoded on pTA1 belongs to group B. Recently, Robinson et al. (2004) analyzed the biochemical characteristics of Cdc6 homologues found in *S. solfataricus*. They proposed that two of them, *Sso*1 and *Sso*3, are positive initiators, and *Sso*2 is a negative regulator of replication (Robinson et al. 2004). These homologues belong to different groups: *Sso*1 and *Sso*3 are in group A, while Sso2 is a member of group B. The Cdc6 homologues in groups A and B may have different characteristics and roles in archaeal replication systems. However, the situation may not be so simple. Only one Cdc6 homologue was found in *P. aerophilum* and it is a member of group B. No obvious Cdc6 homologues were found in the genomes of *Methanocaldococcus jannaschii* or *Methanopyrus kandleri*. Accordingly, the roles of the multiple Cdc6 homologues present in archaeal genomes will have to be examined. The Cdc6 homologue from pTA1 may contribute to the study of archaeal replication systems.

T. acidophilum and *T. volcanium* have two and three homologues of Cdc6, respectively. *Tac*1 and *Tvo*1 represent deep branches in group A in the phylogenetic tree. The Cdc6 homologues of *Thermoplasmatales* are separated into two branches in group B. *Tvo*3 is on the same branch as the pTA1 Cdc6. The branching pattern suggests that the third Cdc6 homologue in the chromosomal DNA was incorporated into the plasmid pTA1 in *T. acidophilum* along with the initiation sequence.

No sequences similar to the known origin of Archaea could be found in the intergenic regions near ORF1 except one short sequence. However, the presence of some direct and inverted repeats in the intergenic regions suggests that one or more of these intergenic regions may function as a putative replication origin for plasmid pTA1.

A genetic manipulation system for *T. acidophilum*, including plasmid vectors and transformation methods, has not been developed. Here, we report a plasmid sequence and a putative replication origin. This plasmid may be useful for construction of vectors. The knowledge presented here can be used to develop genetic tools for *T. acidophilum* that can be used for functional analysis of the genes identified by the whole genome sequence analysis.

References

Brown JR, Doolittle WF (1997) Archaea and the prokaryote-to-eukaryote transition. Microbiol Mol Biol Rev 61:456–502

Brown JW, Daniels CJ, Reeve JN (1989) Gene structure, organization, and expression in archaebacteria. Crit Rev Microbiol 16:287–338

Capaldi SA, Berger JM (2004) Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon *Methanothermobacter thermoautotrophicus*. Nucleic Acids Res 32:4821–4832

Darland G, Brock TD, Samsonoff W, Conti SF (1970) A thermophilic, acidophilic mycoplasma isolated from a coal refuse pile. Science 170:1416–1418

Erauso G, Marsin S, Benbouzid-Rollet N, Baucher MF, Barbeyron T, Zivanovic Y, Prieur D, Forterre P (1996) Sequence of plasmid pGT5 from the archaeon *Pyrococcus abyssi*: evidence for rolling-circle replication in a hyperthermophile. J Bacteriol 178:3232–3237

Felsenstein J (1989) PHYLIP-Phylogeny Inference Package (Version 3.2). Cladistics 5:164–166

Giraldo R (2003) Common domains in the initiators of DNA replication in Bacteria, Archaea and Eukarya: combined structural, functional and phylogenetic perspectives. FEMS Microbiol Rev 26:533–554

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282

Kues U, Stahl U (1989) Replication of plasmids in gram-negative bacteria. Microbiol Rev 53:491–516

Liu J, Smith CL, DeRyckere D, DeAngelis K, Martin GS, Berger JM (2000) Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. Mol Cell 6:637–648

Lopez P, Philippe H, Myllykallio H, Forterre P (1999) Identification of putative chromosomal origins of replication in Archaea. Mol Microbiol 32:883–886

Matsunaga F, Norais C, Forterre P, Myllykallio H (2003) Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. EMBO Rep 4:154–158

Mojica FJ, Ferrer C, Juez G, Rodriguez-Valera F (1995) Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. Mol Microbiol 17:85–93

Ng WL, DasSarma S (1993) Minimal replication origin of the 200-kilobase *Halobacterium* plasmid pNRC100. J Bacteriol 175:4584–4596

Ree HK, Cao KM, Thurlow DL, Zimmermann RA (1989) The structure and organization of the 16S ribosomal RNA gene from the archaebacterium *Thermoplasma acidophilum*. Can J Microbiol 35:124–133

Reiter WD, Palm P, Zillig W (1988) Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eukaryotic pol II promoters. Nucleic Acids Res 16:1–19

Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R, Bell SD (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. Cell 116:25–38

Ruepp A, Graml W, Santos-Martinez ML, Koretke KK, Volker C, Mewes HW, Frishman D, Stocker S, Lupas AN, Baumeister W (2000) The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. Nature 407:508–513

Sambrook J, Fritsch EF, Messing J (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504

Searcy DG, Stein DB, Searcy KB (1981) A mycoplasma-like archaebacterium possibly related to the nucleus and cytoplasms of eukaryotic cells. Ann N Y Acad Sci 361:312–324

Singleton MR, Morales R, Grainge I, Cook N, Isupov MN, Wigley DB (2004) Conformational changes induced by nucleotide binding in Cdc6/ORC from *Aeropyrum pernix*. J Mol Biol 343:547–557

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 87:4576–4579

Yasuda M, Yamagishi A, Oshima T (1995) The plasmids found in isolates of the acidothermophilic archaebacterium *Thermoplasma acidophilum*. FEMS Microbiol Lett 128:157–162