Molecular Characterization of the Minimal Replicon of pSCM201 Involvement in Unidirectional Theta Replication in Extremely Halophilic Archaea

Chaomin Sun¹, Meixian Zhou¹,², Yun Li¹,³, Hua Xiang¹,*

¹State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, P. R. China.
²Current address: Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, VT 05405, USA.
³Current address: Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390-9148, USA

*Corresponding author. Mailing address: Institute of Microbiology, Chinese Academy of Sciences, P. O. Box 2714, Beijing 100080, P. R. China. Tel/Fax: (+86) 10 6265 6916. E-mail: xiangh@sun.im.ac.cn.

Running title: Replication of haloarchaeal plasmid pSCM201
A 3463-bp plasmid pSCM201 was isolated from a halophilic archaeon Haloarcula sp. strain AS7094. The minimal replicon that is essential and sufficient for autonomous replication and stable maintenance in Haloarcula hispanica was determined by deletion analysis of the plasmid. This minimal replicon (~1.8 kb) consisted of only two functionally related segments: a putative origin (ori201) containing an AT-rich region and sets of repeats, and an adjacent gene encoding a putative replication initiation protein (Rep201). Electron microscopic observation and southern blotting analysis demonstrated that pSCM201 replicates via a theta mechanism. Precise mapping of the putative origin suggested that the replication initiated from a fixed site close to the AT-rich region and proceeded unidirectionally toward the downstream rep201 gene, which was further confirmed by electron microscopic analysis of the ClaI-digested replication intermediates. To our knowledge, this is the first unidirectional theta replication plasmid experimentally identified in the domain of archaea, thus providing a novel system for research on archaeal DNA replication.
Archaea are phylogenetically distinct from bacteria and exhibit significant similarities to eukaryotes in their genetic information processing (1, 3, 23, 40). Therefore, they have the potential to serve as simple models for understanding eukaryotic biology, particularly in the field of DNA replication (1, 40, 41). Although most of the archaean replication proteins are eukaryotic-like, others are more similar to the counterparts of bacteria, or are simply the archaean-specific factors, thus future work on archaean genetics may provide insights into the replication mechanism and the evolution of that process (23).

Among the domain of archaea, haloarchaea are well suited to study archaean genetics. They are easy to culture, and are the first archaea that could be efficiently transformed (46, 47). Haloarchaea are rich of plasmids, which represent more than 25% of the genetic material of the cells in some haloarchaeal strains (20), and could provide versatile replicons for investigation of archaean DNA replication. For instance, *Haloferax volcanii* contains four smaller replicons (plasmids and mini-chromosomes) ranging from 6.4 to 690 kb (9, 10), and *Haloarcula marismortui* has eight smaller replicons ranging from 33 to 410 kb (2).

Several haloarchaeal plasmids have been isolated and sequenced, some of which have been used to create haloarchaeal vectors (6, 9, 18-22, 27, 30, 36, 38, 49, 50). Interestingly, most of these plasmids (*e.g.* pGRB1, pHGN1, pHSB1, pHK2 and pNB101) are grouped into the family of rolling-circle-replicating (RCR) plasmids, as they encode homologous replication initiation proteins containing conserved motifs
The single-stranded DNA (ssDNA) forms, a hallmark of RCR plasmids, are also detected in the host strains of the plasmid pNB101 and pGRB1 (45, 50). In contrast, the plasmid pHV2 (9, 27), and the megaplasmids such as pHH1 (6, 39) and pNRC100 (36, 38), are suggested to be theta replication plasmids based on their sequence information (20). The minimal replicons of pNRC100 and pHH1 have been identified, which contain a large AT-rich region and at least one gene encoding the putative replication protein (36, 39). However, the molecular details of the replication of the haloarchaeal plasmids, especially for the theta replication plasmids that may extensively use the host eukaryotic-like replication machinery, have yet to be established.

The pSCM201 is a novel plasmid isolated from a halophilic archaeon Haloarcula sp. strain AS7094, which has little homology with those haloarchaeal plasmids reported so far. Recently, it has been used to develop a stable vector to express a halocin immunity protein in Haloarcula hispanica (48). In this study, we report the sequence features of pSCM201 and the investigation of its replication mechanism.

The minimal replicon containing a replication origin (ori201) and an adjacent gene (rep201) was determined. The replication initiation point (RIP) at the vicinity of an AT-rich region in ori201 was identified, and the unidirectional theta replication model of pSCM201 was established. To our knowledge, this is the first unidirectional theta-type replication plasmid experimentally identified in the domain of archaea, which provides a novel system for archaeal genetic research.
MATERIALS AND METHODS

Strains, plasmids and primers. The halophilic archaeon Haloarcula sp. strain AS7094 is the natural host of pSCM201; The Har. hispanica strain ATCC33960 was used as recipient for pSCM201 derivatives. These haloarchaeal strains were cultivated at 37℃ in AS-168 medium as described previously (29). When necessary, mevinolin was added to a final concentration of 3-5 µg/ml. Escherichia coli strain JM109 was used as the host for all cloning and sequencing experiments, and cultured in Luria-Bertani (LB) medium (42). When needed, ampicillin was added to a final concentration of 100 µg/ml for selection. The cloning vector pUCm-T (Sangon, Shanghai, China) was used to clone PCR products. All primers used in this study are listed in Table 1, and the strains and plasmids are listed in Table 2.

Isolation of the cryptic plasmid pSCM201 and DNA sequencing. The plasmid pSCM201 was isolated from haloarchaeal strain AS7094 by the alkaline lysis method (37). For DNA sequencing, the pSCM201 DNA was digested at either the HindIII or the ClaI site to generate two kinds of linear plasmid DNA molecules, which were inserted into pBluescript II KS- (Stratagene) at the corresponding sites to yield two different recombinant clones, and sequenced by primer walking with the dideoxynucleotide termination cycle sequencing method. The software package GCG (University of Wisconsin, Madison, WI) was used to assemble and analyze the sequences. The FASTA and BLAST programs of the GCG package were used for homology searches in the GenBank and SWISS protein databases. Helix-turn-helix (HTH) motif was identified by software GYM 2.0
Determination of the minimal replicon. In order to identify the minimal replicon of pSCM201, different regions of this plasmid were amplified by PCR and ligated into the cloning Vector (pUCm-T) to generate a series of derivative plasmids. Then the mevinolin-resistance gene ($Mev^r$) was cut as an EcoRI-KpnI fragment from pUBP2 (6) and introduced into the same sites of above plasmids. The resulting plasmids were then transformed into *Har. hispanica* ATCC33960 as described by Cline et al. (12).

Copy number determination and plasmid stability assays. To measure the copy number of pSCM201 per chromosome equivalent, a comparison was made between the hybridization signals from the single-copy chromosome-encoded $cop1$ gene (the accession number is D31880) and the pSCM201-encoded $rep201$ gene in total DNA extracts of AS7094. Restriction enzyme SalI was selected to digest the total DNA of AS7094 and to linearize the pSCM201, resulting in a clear discrimination between chromosome and pSCM201-borne signals. A $rep201$-specific probe (562 bp) was amplified by PCR using the primers orf1F and orf1R, and a $cop1$-specific probe (562 bp) was amplified with the primers cRF and cRR (Table 1). The two PCR products had equal G+C content of 60.5%. The hybridized signal levels of $rep201$ gene and $cop1$ gene were quantified by using a PhosphorImager (Amersham Biosciences) with the ImageQuant 5.2 software. The plasmid copy number was estimated by comparing the signals generated by the plasmid and the chromosomal $cop1$ gene. The relative copy numbers of other pSCM201 derivatives were evaluated.
by comparing the relative intensity of the plasmid bands as described by Emond et al. (14), and their stability was assayed by plasmid maintenance in the absence of selection pressure as described by Ng et al. (36). Experiments were done in triplicate.

**Isolation of plasmid replication intermediates (RIs).** The RIs of pSCM201 from AS7094 and pSCM2011 from *E. coli* were isolated as described by Mojica et al. (33) and Santamaria et al. (43), respectively, except alkaline and RNase treatments were avoided. For replication initiation point (RIP) mapping assay, the RIs of plasmids containing single-stranded DNA were enriched by BND cellulose (Sigma) chromatography as described elsewhere (26). For electron microscopy, the RIs of pSCM201 were further separated by electrophoresis in low melting temperature agarose gel. The DNA from the gel region between the supercoiled plasmid and high-molecular-weight chromosomal bands was isolated, which included all possible plasmid RIs.

**Detection of RIs by electron microscopy.** Samples for electron microscopy were prepared as described by Burkardt et al. (8). Micrographs were recorded using a Hitachi H-700A electron microscope. pBR322 was treated in a similar way and was used as the size marker.

**RIP mapping.** RIP mapping was performed as described by Gerbi and Bielinsky (17). Briefly, Vent (exo-) polymerase was used to extend from a labeled primer to the DNA/RNA junctions of the nascent strand templates in the RIs, in which the nicked contaminating DNA had been removed by λ-exonuclease treatment. For the primer extension experiments of both the pSCM201 RIs from *Haloarcula* sp.
AS7094 and pSCM201 (Table 2) RIs from *E. coli*, the primers RIPF (located about 250 bp upstream of AT-rich region) and RIPR (located about 170 bp downstream of AT-rich region) (Table 1 and Fig. 5A) were respectively applied with the following procedure: 3 min, 94°C; then 30 cycles of 1 min at 94°C, 1 min at 70°C and 1.5 min at 72°C. Sequencing reactions were performed in parallel by fmol DNA cycle sequencing system (Promega) according to the protocols of the manufacturer with the same primers, and all samples were electrophoresed and analyzed on a 6% acrylamide gel containing 8 M urea and 1×TBE. The primers used in RIP mapping and DNA sequencing were labeled by T4 Polynucleotide Kinase with \[^\gamma-32P\] ATP (5,000 Ci/mmol).

**Nucleotide sequence accession numbers.** The nucleotide sequences of pSCM201 and the 16S rRNA gene of the *Haloarcula* sp. AS7094 were deposited in GenBank under the accession numbers AY443099 and DQ334803, respectively.

**RESULTS**

**Isolation of plasmid pSCM201 from an archaeon *Haloarcula* sp. AS7094.** To find novel plasmids in the extremely halophilic archaea, we have screened more than 20 haloarchaeal strains isolated from the salt lakes or marine salterns in P.R.China. An endogenous plasmid about 3.5 kb in size, named pSCM201, was isolated in a halophilic archaeon strain AS7094. To identify the host strain, the 16S rRNA gene of AS7094 was amplified with the specific haloarchaeal primers (16sF and 16sR in Table 1) and the sequence was determined. The sequence BLAST in NCBI database showed
that AS7094 was closely related to *Haloarcula argentinensis* (homology of the 16S rRNA gene is about 99%), thus the host strain of pSCM201 was nominated *Haloarcula* sp. AS7094.

**Sequence analysis of pSCM201.** DNA sequence analysis showed that pSCM201 is a circular molecule of 3463 bp (Fig. 1) with a G+C content of 59.9%, which is in accordance with the high G+C content of the genomes or plasmids of haloarchaea (6, 18, 19, 22, 27, 36, 38, 50). Examination of the six possible phases of the plasmid sequence revealed the presence of three ORFs (ORF1, ORF2, ORF3) in the two strands, which cover about 70% of the entire sequence (Fig. 1). These ORFs were supposed to encode three polypeptides with 399, 135 and 212 amino acids, respectively. No significant homology was found in the NCBI protein databases for the deduced proteins encoded by ORF2 and ORF3. However, a putative leucine zipper (LZ) motif (residue 88-108), a helix-turn-helix (HTH) motif (residue 129-150) and an ATPase domain (residue 148-309) were detected in the ORF1-encoded protein (Fig. 2). Since most replication initiation proteins harbour DNA-binding motifs and possess ATPase activities (28), and leucine zipper (LZ) motif was also found in Rep proteins of many iterons-controlled plasmids (11, 13), we presumed that the protein encoded by ORF1 was the replication initiation protein of pSCM201, named Rep201 (44.7 kDa). The Rep201 is highly acidic, with a calculated pI of 4.74, which has been suggested as a major adaptive mechanism of most haloarchaeal proteins to function in nearly saturating salinity (24).

Upstream of the *rep201* gene, a 24-bp AT-rich region (70.9% A+T) located about
550 bp 5′ to the rep201 start codon was detected (Fig. 2), which is flanked by a perfect inverted repeat IR-1 (5′-CTACACCC-3′). Downstream of this region, there are sets of perfect inverted and direct repeats, followed by 7 copies of 9-bp iteron-like direct repeat with consensus sequence of GGAGCGAGC (Fig. 2). Since these features resemble the replication origins of many theta replication plasmids, we presumed the above region as the replication origin of pSCM201, and named ori201 (Fig. 1 and Fig. 2).

**Determination of the minimal replicon of pSCM201.** In order to determine the minimal replicon of pSCM201, the different regions of this plasmid plus the mevinolin-resistance gene were cloned into the *E. coli* cloning vector pUCm-T without any haloarchaeal replicons (see Materials and Methods), and tested for their abilities to replicate and maintain in *Har. hispanica* ATCC33960. The minimal replicon (~1.8 kb) was demonstrated to contain only the putative replication origin region (ori201) and the adjacent rep201 gene, as deletion of the ORF2, ORF3 and the partial non-coding sequence downstream of the ORF3, could not interfere with the replication (Fig. 3). The pSCM201-derived vectors are very stable in the transformed *Har. hispanica* cells, and the minimal replicon is sufficient for stable maintenance of the derived plasmids. For instance, after 100 generations of growth for even without selection pressure, the pSCM201 derivatives (pSCM204, pSCM212, pSCM216) were sustained in more than 98% cells of the transformed *Har. hispanica* (Fig. 3), and the copy numbers are equal to that of pSCM201 (about 10 ± 2 per chromosome).

The rep201 gene is necessary for this plasmid replication. Interruption of this
gene at the ClaI site resulted in loss of replication ability of the derivative pSCM203 (Table 2) in *Har. hispanica*. Removing the last 112 amino acids from Rep201 also abolished the replication (pSCM214, see Fig. 3). Moreover, when deletions extended into the inverted repeat (nucleotide 1317) downstream of *rep201*, which we considered to be the transcriptional terminator of *rep201* (Fig. 2), the derivative pSCM213 became unstable. About 40% of *Har. hispanica* cells transformed with pSCM213 lost the plasmids after 100 generations without selection pressure, and the copy number of the pSCM213, if existed, decreased to 5 ± 2 per chromosome. Likewise, the *ori201* region was also required for the plasmid replication. A shuttle vector named pSCM217, which was only short of the AT-rich region in contrast to pSCM216, lost replication ability in *Har. hispanica* (Fig. 3). Even interruption of this region at the NcoI site by insertion of ~6.2 kb fragment including the pUCm-T sequence and *Mev* gene, the resultant plasmid pSCM202 (Table 2) could not replicate in haloarchaea.

Taking these results together, we conclude that the putative replication origin *ori201* combined with its adjacent gene (*rep201*) are essential and sufficient for plasmid replication in halophilic archaea, which constitute the minimal replicon of pSCM201.

**pSCM201 replicates via a theta-replication mechanism.** To determine whether pSCM201 replicates via a RCR or a theta-replication mechanism, we first checked if single-stranded DNA (ssDNA) intermediates of this plasmid were generated, as it is the hallmark of a RCR plasmid (25, 45, 50). The plasmid replication intermediates
(RIs) were separated by agarose gel electrophoresis and blotted onto a nylon membrane. Then the DNA of pSCM201 was detected with a rep201-derived probe that was amplified by primers orf1F and orf1R (Table 1) and labeled with $\alpha^{-32}$P-dCTP, and the known RCR-type plasmid pNB101 (50) was used as the positive control, which was detected with a probe amplified by primers P66 and P67 (Table 1). No ssDNA of pSCM201 was detected in this assay even after prolonged exposure of the film, while under the same condition the ssDNA of pNB101 was detected (Fig. 4A). The absence of ssDNA RIs suggested that pSCM201 might not replicate via the RCR mechanism.

To explore whether pSCM201 replicates via a theta-replication mechanism, we then enriched the RIs and subjected them to electron microscopic analysis. The characteristic bubble of theta-type replication was observed in the pSCM201 RIs (Fig. 4B). In addition, when the ClaI-cleaved RIs of pSCM201 were subjected to neutral/neutral two dimension (N/N 2D) agarose gel electrophoresis, an arc supposed to be formed from the replication bubble-carrying molecules was also detected (data not shown). These results demonstrate that pSCM201 is indeed a theta-replication plasmid.

**RIP mapping of pSCM201.** To further characterize the plasmid, we determined the potential RIP within the ori201 by RIP mapping (17). This technique can map the replication initiation site with a nucleotide resolution, and has been successfully applied to determine the precise RIPS of eukaryotic and archaeal chromosome recently (5, 32, 41). The principle of RIP mapping is to extend from a labeled primer
to the DNA/RNA junctions of RNA-primed nascent strands with Vent (exo-) polymerase, and the RIP is usually defined as the transition point between the leading and lagging strands. This technique would be competent to determine the RIP of a theta-type plasmid, as in which the replication also initiates with RNA primers.

To investigate if the RIP of pSCM201 is located within the putative ori201 and to precisely identify its position, the vicinity of the AT-rich region in ori201 was primarily taken into account, and the oligonucleotides RIPF and RIPR (Table 1 and Fig. 5A) were subjected to the primer extension reactions (see Materials and Methods). It was clearly revealed that the primer extension from RIPR, which is located downstream of the AT-rich region with the reverse direction of rep201 transcription, gave one prominent band with the nascent DNA templates of pSCM201 from Haloarcula sp. AS7094 (Fig. 5B, bottom strand, RH). In contrast, primer extension with the samples of pSCM2011 containing the complete sequence of pSCM201 isolated from E. coli did not give rise to such a specific extension product using the same primer (Fig. 5B, bottom strand, RE). This is consistent with the notion that the ori201 would function in Haloarcula sp. AS7094 but not in E. coli. Furthermore, when using the same nascent DNAs as the templates, no apparent and specific product was generated with RIPF, the primer located upstream of the AT-rich region with the direction of rep201 transcription (Fig. 5B, top strand).

These results are significantly different from those of bidirectional theta replication, but support the argument that pSCM201 replicates unidirectionally. As to a unidirectional replication origin, only the primer orientated on the opposite of
replication (RIPR in the case of pSCM201) can produce a prominent extension product, which points to the DNA/RNA junction of the leading strand and thus represents the RIP. Characteristically, the primer from the other direction upstream of the RIP could hardly give rise to any extension products, since almost no RNA-primed nascent DNA in that direction would be generated in such a unidirectional theta replication origin. Our results suggested that pSCM201 replicated from a fixed site 14-bp downstream of the AT-rich region between a conserved inverted repeat IR-1(Fig. 5B, C and D), and progressed unidirectionally toward the downstream rep201 gene.

Electron microscopic analysis of replication direction of pSCM201. To further confirm the unidirectional replication of pSCM201, the single ClaI restriction site was used to linearize the plasmid, thus introducing a convenient reference point in electron microscopic analysis. Measurements of about 20 bubble-carrying molecules (Fig. 6A and B) revealed that the sizes of the long unreplicated arms were constant (2400±100 bp). Measurements of about 20 double Y-shaped molecules showed that the size of one fork was constant (1100±100 bp), while the other was changeable but not more than 2400±100 bp (Fig. 6C). For X-shaped molecules (Fig. 6D), the longer and shorter forks were usually about 2400 and 1100 bp, respectively. These data were summarized in Fig. 6E. In both the bubble-carrying and double Y-shaped replication intermediates, the length from the ClaI site to the left-hand branch point (L) or right-hand branch point (R) was calculated. It was revealed that position of left-hand branch point did not move with increasing extent of replication, whereas the position
of the right-hand branch point depended linearly on the extent of replication (Fig. 6E),
thus confirming the unidirectional replication mode.

This left-hand branch point (~2400 bp downstream or 1100bp upstream of the
ClaI site, indicated by white arrows in Fig. 6A-D) was in accordance very well with
the replication initiation point (RIP) pinpointed by RIP mapping, as the distance from
the designated RIP (Fig. 5C) to the downstream ClaI site was exactly 1103 bp.

DISCUSSION

In this manuscript we characterized a novel haloarchaeal plasmid pSCM201
(3463 bp), and showed that this small, naturally occurring plasmid replicated via a
unidirectional theta-replicating mechanism, initiating at the vicinity of an AT-rich
region within the *ori201*. The evidences for this replication mechanism were based on
the following observations: (i) The minimal replicon of pSCM201 containing the
putative *ori201* region and *rep201* gene showed structural similarities with theta
replication plasmids but not RCR plasmids; (ii) Unlike RCR plasmids, pSCM201 did
not generate ssDNA replication intermediates (RIs); (iii) Electron microscopy studies
showed the presence of replication bubble of RIs; (iv) RIP mapping of pSCM201
determined that the location of RIP was at the vicinity of the AT-rich region in the
*ori201*, and revealed that the replication progressed only in one direction, which was
further confirmed by electron microscopy. To our knowledge, pSCM201 is the first
unidirectional theta replication plasmid experimentally identified in the domain of
archaea.
The pSCM201 is very stable. The vectors derived from the complete minimal replicon of pSCM201 had a constant copy number of about 10 for at least 100 generations in the absence of selective pressure. So we concluded that most efficient control elements were within this minimal replicon. In contrast to the stable plasmid pSCM212, the unstable and low-copy-number plasmid pSCM213 only lacked the putative terminator of rep201 (Fig. 3), indicating that Rep201 may participate in the control of the plasmid replication initiation as well as its copy number. Rep201 contains a putative LZ motif, thus has the potential to form a dimer. Indeed such a dimer was detected in the native PAGE gel with the purified His-tagged Rep201 protein (data not shown), suggesting that Rep201 may be involved in control of plasmid copy number via handcuffing like many bacterial theta replication plasmids (11). Interestingly, Rep201 had an ATPase domain (residues 148-309), which is different from those of bacterial theta-replication plasmids that usually lack the ATPase activities and might need the ATPase activities of DnaA to melt the origin (11). Therefore, Rep201 may have some novel functions not yet been observed in its bacterial counterparts.

It is noteworthy that the inverted repeat IR-1 (5'-CTACACCC-3') flanking the AT-rich region in the ori201, is similar to the inner sequence (5'-CCCCACCC-3') of putative origin recognition boxes (ORBs) of Halobacterium sp. strain NRC-1 (Fig. 5D). This inner sequence has been speculated as the binding sites of the initiation proteins of archaea, the Orc1/Cdc6, or other DNA replication initiation proteins (4, 41). Similar structures containing inverted repeat flanking an AT-rich region were
found to be highly conserved upstream of \emph{orc7} gene (coding for the putative replication initiation protein) in the chromosomes of three other halophiles, \emph{Haloarcula marismortui}, \emph{Haloferax volcanii} (4) and \emph{Natronomonas pharaonis} (15) (Fig. 5D). In addition, chromosomes of two other archaea, \emph{Pyrococcus} (32, 34) and \emph{Sulfolobus} (41), are also reported to contain inverted repeat\textendash AT-rich region upstream of their \emph{orc7} homologs. This structure is reminiscent of a duplex unwinding element, a feature found in many bacteria and eukaryotic origins of replication and thought to facilitate melting of DNA at the origin (35). Such a structure in pSCM201 would contribute to its theta replication, as the RIP of pSCM201 was just located between the inverted repeat (IR-1) and the nearby AT-rich region. Moreover, there is a pair of 7-bp direct repeat (TCAATCA) within the AT-rich region, which might be the entry site of helicase or its loader, like that of \emph{E.coli} plasmid pSC101 (7). Downstream of the AT-rich region, there are sets of inverted repeats and direct repeats including 7 putative iterons. These repeats might be the binding sites of Rep201 or other involved proteins.

Theta-type replication of plasmids, including many iteron-controlled bacterial plasmids (31), is usually unidirectional (13). In the well-studied unidirectional theta replication plasmid pBR322, it is found that the orientation of the DnaA boxes is critical for the efficiency of the unidirectional replication (44). It is suggested that DnaA may interact with one of the DNA strands more strongly, on which helicase loading would be more productive for unidirectional replication (16). For pSCM201, it will be interesting to investigate the possible correlation of replication direction and
the structural asymmetry of the ori201 in the future.

ACKNOWLEDGEMENTS

We would like to thank Prof. Peijin Zhou for providing several haloarchaeal strains for plasmid screening, and Dr. Juan Wang at University of Pennsylvania for helpful comments during preparation of this manuscript. This work was supported by grants from the National Basic Research Program (“973” program) of China (2004CB719603), the National “863” Program of China (2004AA626010), the National Natural Science Foundation of China (30570029), and the Chinese Academy of Sciences.

REFERENCES


sequence of the arcaebacterial plasmid pHSB from *Halobacterium*, strain SB3. Nucleic Acids Res. **16**:4158.


replicon in a halophilic archaeon: megaplasmid or minichromosome? Genome Res. 8:1131-1141.


47. **Soppa, J.** 2006. From genomes to function: haloarchaea as model organisms. Microbiology 152:585-590.


**FIGURE LEGENDS**

**FIG. 1.** Map of pSCM201. The three open reading frames, named ORF1 (*rep201*), ORF2, and ORF3, are indicated with closed arrows. *Ori201* is indicated by a box, and the replication direction is indicated by an open arrow. Stem-loop represents the putative terminator of *rep201*. The restriction sites for appropriate enzymes are indicated, with their positions in the parentheses. Base 1 is arbitrarily defined by the first base of the HindIII recognition site.

**FIG. 2.** Sequence analysis of the replicon of pSCM201. The AT-rich region of *ori201* is boxed and direct repeats within this AT-rich region are indicated with dashed arrows. The inverted repeats (IR-1 to IR-4) are shown with arrows and the two sets of direct repeats (DR-1 and DR-2) with half arrows. The putative iterons are boxed with pentagons. The putative TATA-box for
rep201 is double underlined. The putative transcriptional terminator of rep201 is indicated with open arrows. The restriction sites (HindIII, NcoI, ClaI) are underlined. Several primers’ positions are indicated with bent arrows. The amino acid sequence of the Rep201 (399 amino acids) encoded by rep201 gene is indicated below the DNA sequence. The numbers on the right indicate the nucleotide number of pSCM201, while the numbers on the left indicate amino acid number of Rep201. The putative leucine zipper (LZ) motif, the helix-turn-helix (HTH) motif and the ATPase domain of Rep201 are also indicated.

FIG. 3. Determination of the minimal replicon of pSCM201. A linear restriction map is shown on the top of the diagram. Putative ORFs (black arrows), ori201 (rectangle) and terminator (stem-loop) shown on the restriction map were inferred from sequence analysis. Thick lines under the linear map represent the fragments cloned for autonomous replication assays. The names of primers (indicated as grey block) are placed under the ends of each fragment. The names of recombinant plasmids containing these fragments are indicated on the left. The ability of the derivatives to replicate (repl.) in Haloarcula hispanica is indicated as “+” (yes) or “−” (no). Stability (stab.) indicates the ratio of cells (%) harboring the plasmids after 100 generations of growth without selective pressure.

FIG. 4. Analyses of replication intermediates (RIs) of pSCM201 by Southern blotting (A) and electron microscopy (B). (A) Detection of the ssDNA RIs of pNB101 (left panel) and pSCM201 (right panel). (B) Uncut replication bubble-shaped molecule of pSCM201. The bar corresponds to 0.5 kb.

FIG. 5. Mapping of replication initiation point (RIP) of pSCM201. (A) Outline of RIP mapping for unidirectional theta replication of pSCM201 (not to scale). Nascent DNAs (leading
strand and Okazaki fragments) initiated by a small RNA primer (black rectangles) were used as the templates in the primer extension reaction to determine the RIP. Ovals indicate the primers used in RIP mapping. The directions of the pSCM201 replication, rep201 transcription and the primer extension are indicated by arrows. (B) Identification of replication initiation point by RIP mapping. Primer extension was performed with oligonucleotide RIPF for the top strand (left panel), and RIPR for the bottom strand (right panel). Lanes contain dideoxy sequencing reactions (GATC) and the products of primer extension of pSCM201 RIs from Haloarcula sp. strain AS7094 (RH) and pSCM2011 RIs from E. coli (RE). Black arrow indicates the RIP position of pSCM201. The open arrows indicate the positions of IR-1. The dashed lines indicate the positions of AT-rich region. (C) Nucleotide sequence of a region containing the IR-1, AT-rich region and RIP of pSCM201. (D) Conserved inverted repeats in ori201 and the chromosomal replication origins of other four haloarchaea. The black arrows indicate the long inverted repeat flanking the AT-rich region in the four haloarchaeal chromosomal replication origins. The open arrows indicate IR-1 of ori201, which shares high homology with the inner sequence of those chromosomal long inverted repeats. NRC-1: Halobacterium sp. strain NRC-1; H. ma: Haloarcula marismotui; N. ph: Natronomonas pharaonis; H. vo: Haloferax volcanii.

FIG. 6. Analysis of ClaI-cleaved RIs of pSCM201 by electron microscopy. (A and B) Replication bubble-shaped molecules of pSCM201. (C) Double Y-shaped molecule of pSCM201. (D) X-shaped molecule of pSCM201. The bars correspond to 1.0 kb. The white arrows indicate the replication initiation point. (E) Analysis of the branch-point positions of the replication forks in the pSCM201 RIs. Left panel: The pSCM201 RIs were schematically ordered by increasing extent of replication (one line: unreplicated region; two lines: replicated region). The positions of
left-hand and right-hand branch points were measured as the “L” and “R”, respectively, relative to the same ClaI site. Right panel: Positions of the left-hand (white circles) or right-hand (black circles) branch points are plotted versus the extent of replication.
### TABLE 1. Primers used in this study

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16sF</td>
<td>ATTCCGGTTGTACCTAGCCGAG</td>
</tr>
<tr>
<td>16sR</td>
<td>AGGAGGTGATCCAGCCAGAG</td>
</tr>
<tr>
<td>orf1F</td>
<td>AAGCGACTAAGCCACATC</td>
</tr>
<tr>
<td>orf1R</td>
<td>CCTTCTCTGCCTATATCT</td>
</tr>
<tr>
<td>cRF</td>
<td>GCAGAATACCTCATCACAG</td>
</tr>
<tr>
<td>cRR</td>
<td>GCAGGATGATACCGAAGC</td>
</tr>
<tr>
<td>204F</td>
<td>TCCCAATCCTGGCTGAC</td>
</tr>
<tr>
<td>204R</td>
<td>GCAGGGAACATAAACACC</td>
</tr>
<tr>
<td>205F</td>
<td>ACGAAGTGACCGAAGG</td>
</tr>
<tr>
<td>209R</td>
<td>CCAGACCGAACACCAC</td>
</tr>
<tr>
<td>212R</td>
<td>GGCTACGGGAAATGCTT</td>
</tr>
<tr>
<td>213R</td>
<td>GCCTATCTACACACAAG</td>
</tr>
<tr>
<td>214R</td>
<td>ACCGCTCCTGCTCTCT</td>
</tr>
<tr>
<td>216F</td>
<td>GACCGCTACACCACAG</td>
</tr>
<tr>
<td>P66</td>
<td>CGCGAGAAAGCGCTACAC</td>
</tr>
<tr>
<td>P67</td>
<td>TCACGCCACTTGCTAACG</td>
</tr>
<tr>
<td>RIPF</td>
<td>ACCTATGACAGCAAGCAAGCCTAC</td>
</tr>
<tr>
<td>RIPR</td>
<td>GACGAACGAACAGCGGATAAACAG</td>
</tr>
</tbody>
</table>
## TABLE 2. Strains and plasmids used in this study.

<table>
<thead>
<tr>
<th>Strains and plasmids</th>
<th>Genotype or description</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Haloarcula</em> sp. strain AS7094</td>
<td>Natural host of pSCM201</td>
<td>IMCAS</td>
</tr>
<tr>
<td><em>Haloarcula hispanica</em> strain ATCC33960</td>
<td>Recipient of pSCM201 derivatives</td>
<td>ATCC</td>
</tr>
<tr>
<td><em>Escherichia coli</em> JM109</td>
<td><em>recA1 supE44 endA1 hisD17 gyrA96 relA1 thi</em> Δ*(lac-proAB)* F* [traD36 proAB’ lacI4 lacZΔ M15]*</td>
<td>42</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pUCm-T</td>
<td>2.77-kb Vector for cloning PCR products; Amp&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Sangon, China</td>
</tr>
<tr>
<td>pUCmT</td>
<td>2.77-kb replication type of pUCm-T</td>
<td>This work</td>
</tr>
<tr>
<td>pBluescript II KS-</td>
<td>2.96-kb cloning vector; Amp&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Stratagene</td>
</tr>
<tr>
<td>pUBP2</td>
<td>12.3-kb shuttle vector; Amp&lt;sup&gt;+&lt;/sup&gt;, Mev&lt;sup&gt;+&lt;/sup&gt;</td>
<td>6</td>
</tr>
<tr>
<td>pBR322</td>
<td>4361-bp plasmid; Amp&lt;sup&gt;+&lt;/sup&gt;, Tet&lt;sup&gt;+&lt;/sup&gt;</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>pSCM201</td>
<td>A 3463-bp natural plasmid of <em>Haloarcula</em> sp. strain AS7094; GenBank Accession No. AY443099</td>
<td>This work</td>
</tr>
<tr>
<td>pSCM2011</td>
<td>3463-bp HindIII fragment from pSCM201 cloned into pBluescript II KS-; Amp&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pSCM2012</td>
<td>3463-bp ClaI fragment from pSCM201 cloned into pBluescript II KS-; Amp&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pSCM202</td>
<td>3463-bp NcoI fragment from pSCM201 cloned into pUCmT, then inserted 3.5-kb Mev&lt;sup&gt;+&lt;/sup&gt; fragment; Amp&lt;sup&gt;+&lt;/sup&gt;, Mev&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pSCM203</td>
<td>3.5-kb Mev&lt;sup&gt;+&lt;/sup&gt; fragment cloned into pSCM2012; Amp&lt;sup&gt;+&lt;/sup&gt;, Mev&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pSCM204</td>
<td>3389-bp PCR product (204F-204R) cloned into pUCm-T, then inserted 3.5-kb Mev&lt;sup&gt;+&lt;/sup&gt; fragment; Amp&lt;sup&gt;+&lt;/sup&gt;, Mev&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pSCM212</td>
<td>2143-bp PCR product (204F-212R) cloned into pUCm-T, then inserted 3.5-kb Mev&lt;sup&gt;+&lt;/sup&gt; fragment; Amp&lt;sup&gt;+&lt;/sup&gt;, Mev&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pSCM213</td>
<td>2069-bp PCR product (204F-213R) cloned into pUCm-T, then inserted 3.5-kb Mev&lt;sup&gt;+&lt;/sup&gt; fragment; Amp&lt;sup&gt;+&lt;/sup&gt;, Mev&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pSCM214</td>
<td>1730-bp PCR product (204F-214R) cloned into pUCm-T, then inserted 3.5-kb Mev&lt;sup&gt;+&lt;/sup&gt; fragment; Amp&lt;sup&gt;+&lt;/sup&gt;, Mev&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pSCM216</td>
<td>1999-bp PCR product (216F-209R) cloned into pUCm-T, then inserted 3.5-kb Mev&lt;sup&gt;+&lt;/sup&gt; fragment; Amp&lt;sup&gt;+&lt;/sup&gt;, Mev&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pSCM217</td>
<td>1885-bp PCR product (205F-209R) cloned into pUCm-T, then inserted 3.5-kb Mev&lt;sup&gt;+&lt;/sup&gt; fragment; Amp&lt;sup&gt;+&lt;/sup&gt;, Mev&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This work</td>
</tr>
</tbody>
</table>
pSCM201
3463 bp

ori201

NcoI(3224)

HindIII(3463,1)

ClaI(654)

ORF1 (rep201)

ORF2

ORF3