JOURNAL OF BACTERIOLOGY, Oct. 2003, p. 5959–5966 Vol. 185, No. 20
0021-9193/03/$08.00+0 DOI: 10.1128/JB.185.20.5959–5966.2003

An Archaeal Chromosomal Autonomously Replicating Sequence Element from an Extreme Halophile, *Halobacterium* sp. Strain NRC-1

Brian R. Berquist and Shiladitya DasSarma*

Molecular and Cell Biology Program, University of Maryland, Baltimore, and Center of Marine Biotechnology, University of Maryland Biotechnology Institute, Baltimore, Maryland 21202

Received 27 May 2003/Accepted 31 July 2003

We report on the identification and first cloning of an autonomously replicating sequence element from the chromosome of an archaeon, the extreme halophile *Halobacterium* strain NRC-1. The putative replication origin was identified by association with the *orc7* gene and replication ability in the host strain, demonstrated by cloning into a nonreplicating plasmid. Deletion analysis showed that sequences located up to 750 bp upstream of the *orc7* gene translational start, plus the *orc7* gene and 50 bp downstream, are sufficient to endow the plasmid with replication ability, as judged by expression of a plasmid-encoded mevinolin resistance selectable marker and plasmid recovery after transformation. Sequences located proximal to the two other chromosomally carried haloarchaeal *orc* genes (*orc6* and *orc8*) are not able to promote efficient autonomous replication. Located within the 750-bp region upstream of *orc7* is a nearly perfect inverted repeat of 31 bp, which flanks an extremely AT-rich (44%) stretch of 189 bp. The replication ability of the plasmid was lost when one copy of the inverted repeat was deleted. Additionally, the inverted repeat structure near *orc7* homologs in the genomic sequences of two other halophiles, *Haloarcula marismortui* and *Haloferax volcanii*, is highly conserved. Our results indicate that, in halophilic archaea, a chromosomal origin of replication is physically linked to *orc7* homologs and that this element is sufficient to promote autonomous replication. We discuss the finding of a functional haloarchaeal origin in relation to the large number of *orc1-cdc6* homologs identified in the genomes of all haloarchaea to date.

Archaeal microorganisms are phylogenetically distinct from bacteria (38) and exhibit significant similarities to eukaryotes in their macromolecular biosynthetic machinery, including DNA replication, transcription, and translation systems (11, 31). They can therefore serve as relatively simple models for eukaryotes, although the difficulty of culturing most archaea and their limited capability for genetic manipulation in the laboratory have restricted the scope of studies of archaeal molecular biology. However, recent advances in the genomics, genetics, and biochemistry of archaea have started to open new avenues in archaeal research. This is especially true for the halophilic archaea (haloarchaea), which flourish in hypersaline conditions where NaCl concentrations can approach saturation (8). Haloarchaea are easily cultured in the laboratory and have well-developed genetic transformation systems with gene replacement and knockout methodology (6, 32). As a result, we sought to utilize a combination of genetics and genomics to initiate a study of DNA replication in haloarchaea.

The single complete haloarchaeal genome sequence obtained thus far, for *Halobacterium* strain NRC-1, revealed a dynamic 2,571,010-bp genome including 2,682 predicted genes, distributed among three replicons: a 2-Mb large chromosome and two large extrachromosomal replicons, pNRC200 (365 kb) and pNRC100 (191 kb) (30). As with other archaea, the DNA replication machinery of *Halobacterium* strain NRC-1 appears to be more eukaryotic than bacterial in nature (9, 10). Bioinformatic analysis surprisingly showed the presence of 10 homologs of the eukaryotic-type origin recognition proteins (Orc1 and Cdc6) in NRC-1 (P. Zhang, S. P. Kennedy, B. Berquist, and S. DasSarma, unpublished data), compared to one or two at most in other archaea. In addition, homologs of the eukaryotic-type replicative helicase proteins (MCM), single-stranded DNA binding proteins (RFA), processivity clamp loader proteins (RFC), processivity clamp proteins (PCNA), primase proteins, RNA primer removal proteins (Rad2 and RNase H), ATP-dependent DNA ligase, DNA polymerases (B family), and type IIB topoisomerase (Top6A and Top6B) were also found. A novel heterodimeric family D DNA polymerase gene found only in euryarchaea (4, 37) was also present, as were a few genes for bacterial protein-like replication proteins, a primase (DnaG), and type IA (TopA) and IIA (GyrA and GyrB) DNA topoisomerases.

The genome sequence of *Halobacterium* strain NRC-1 was analyzed for GC skew, (G+C)/(G−C), on a DNA strand in a sliding window (17). Typically in bacteria and other archaea, GC skew analysis will show the presence of two polarity switches, one indicating the origin of DNA replication and the other indicating the terminus. However, GC skew indicated the presence of four polarity switches for *Halobacterium* strain NRC-1, consistent with two chromosomal origins of replication. Further analysis of the polarity switch regions revealed the presence of two *orc1-cdc6* homologs located proximal to switch points, suggesting the association of the genes with replication origins (22, 27). In many bacterial species, the *dnaA* gene, encoding the replication initiation protein, is also located

* Corresponding author. Mailing address: Center of Marine Biotechnology, University of Maryland Biotechnology Institute, 701 E. Pratt St., Baltimore, MD 21202. Phone: (410) 234-8847. Fax: (410) 234-8896. E-mail: dassarma@umbi.umd.edu.

proximal to one of the polarity switches (14, 24). Our findings led us to hypothesize that multiple origins of *Halobacterium* strain NRC-1 may be used to initiate DNA replication of the chromosome, suggesting that DNA replication in this haloarchaeon may be more eukaryotic than bacterial (17).

In past studies, only a single autonomously replicating sequence (ARS) was isolated for *Halobacterium* strain NRC-1, located within the common region of the large extrachromosomal pNRC100 and pNRC200 replicons. Extensive analysis of these replicons, which exist as inversion isomers, had also shown the presence of a B family DNA polymerase, TBP, and TFB transcription factor genes, among other important and likely essential genes, suggesting minichromosome status in *Halobacterium* strain NRC-1 (7, 29). Sequence analysis of the minimal replication origin (28) showed the requirement for a unique gene, *repH*, and an AT-rich region 5′ to the gene. Elimination of either the AT-rich sequence or the *repH* gene was found to abolish the autonomous replication ability of plasmids.

In the present study, we have attempted to identify the *Halobacterium* strain NRC-1 chromosomal replication origin(s) functionally by assessing its capability to confer autonomous replication ability on nonreplicating plasmids. We have found evidence for only a single replicating sequence, located near one of the *orc1-cdc6* genes (*orc7*) on the chromosome. We further show that both the *orc7* gene and an upstream inverted repeat sequence flanking an AT-rich region are required for autonomous replication and that this region is conserved in two other distantly related haloarchaea.

MATERIALS AND METHODS

Materials. Restriction enzymes, calf intestinal phosphatase, Vent DNA polymerase, and T4 DNA ligase were purchased from New England Biolabs, Beverly, Mass. The XL-PCR kit was purchased from Applied Biosystems, Branchburg, N.J. The digoxigenin nucleic acid labeling and detection kit was purchased from Roche Applied Science, Indianapolis, Ind. Oligonucleotides were purchased from Sigma-Genosys, The Woodlands, Tex. Gel extraction kits and plasmid purification kits were purchased from Qiagen, Valencia, Calif. Mevinolin was a gift from Merck, Rahway, N.J.

Strains and culturing. *Escherichia coli* DH5α was grown in Luria-Bertani medium supplemented with 100 μg of ampicillin/ml at 37°C. *Halobacterium* strain NRC-1 was cultured in CM^+ medium containing 4.5 M NaCl and trace metals, as well as 20 μg of mevinolin/ml where applicable, as previously described (6).

Plasmid construction. PCR amplification under standard conditions was performed on *Halobacterium* strain NRC-1 genomic DNA, which was prepared as previously described (6). PCR primers used for plasmid construction were as follows: ORC1/CDC6_6.1, 5′-CCGGAATTCAATCGCTGACGGCGGCCAGG-3′; ORC6 + 750.FOR, 5′-CCGGAATTCATCGAAGGCGGGCGCAATCC-3′; ORC6 + 500.FOR, 5′-CCGGAATTCTTCGAGGCGACGCTGCGGGA-3′; ORC6 + 250.FOR, 5′-CCGGAATTCTCGACGGCATCCACGCGGCG-3′; ORC6GENE-50.REV, 5′-CCGGAATTCCACGCCACCACCTAAAAAGC-3′; ORC6-100.REV, 5′-CCGGAATTCCGTCTGCAGCTGTGACTCCC-3′; ORC 1/CDC6_6.2, 5′-CCGGAATTCCATGGGAAGCCTTATAAGCGG-3′; ORC1/CDC6_7.1, 5′-CGCGGATCCTCTCGTGGGGAATGTGGCTC-3′; ORC7 + 750.FOR, 5′-CGCGGATCCTATAAAACCCCAGTGAGAGG-3′; ORC7 + 500.FOR, 5′-CGCGGATCCTAGAACTAGACAACTACCGT-3′; ORC7 + 250.FOR, 5′-CGCGGATCCCATCCCTCCCCCCTCTAACC-3′; ORC7GENE-50.REV, 5′-CGCGGATCCCTCCTCGACCGGAACCGACA-3′; ORC7-100.REV, 5′-CGCGGATCCGCTGAGCAGGTCATCGAAGA-3′; ORC1/CDC6_7.2, 5′-CGCGGATCCAACCCCGGCGTATAACTCGT-3′; ORC1/CDC6_8.1, 5′-CCGGAATTCGTAATCGAGGAAGTCGGCGT-3′; ORC8 + 750.FOR, 5′-CCGGAATTCTCGCTGCCGTGCTCGCCGTC-3′; ORC8 + 500.FOR, 5′-CCGGAATTCCACGTCGTGTTGGCGGTGGT-3′; ORC8 + 250.FOR, 5′-CCGGAATTCGGCTCTCCCACCCACCCCGT-3′; ORC8GENE-50.REV, 5′-CCGGAATTCGGCGTCCGCGATGGCCTCCT-3′; ORC8-100.REV, 5′-CCGGAATTCAGGGCGTTGGCGTACTGGTC-3′; and ORC1/CDC6_8.2, 5′-CCGGAATTCACCGCGTCGTTCAAGCCACG-3′. The PCR products were then gel purified and digested with either *Eco*RI or *Bam*HI. These products were then cloned into either the *Eco*RI or the *Bam*HI site of pNGMEV101 (28). Plasmids were transformed into *E. coli* DH5α by electroporation and recovered by alkaline lysis, as described previously (33). A description of all plasmid constructs is found in Table 1.

Transformation of *Halobacterium* strain NRC-1 and autonomous plasmid replication assay. DNA-mediated transformation of *Halobacterium* strain NRC-1 was performed by the EDTA-polyethylene glycol procedure, as described previously (6). Mevinolin-resistant (Mev^r) transformants were selected by plating on CM^+ solid medium containing 20 μg of mevinolin/ml. Transformation cultures were also spotted onto CM^+ solid medium containing 20 μg of mevinolin/ml. Plasmid DNA was prepared by an alkaline sodium dodecyl sulfate procedure, as described previously (6). Plasmid DNA was prepared from at least three separate Mev^r colonies. This DNA was fractionated by electrophoresis on 0.5% agarose, denatured, transferred to a nylon membrane (Roche), and probed according to the manufacturer's instructions (Roche) with a digoxigenin-labeled PCR product of the *bla* gene as a plasmid-specific probe.

DNA sequence analysis. Sequence data for *Haloarcula marismortui* were obtained from the *H. marismortui* genome web page at http://zdna2.umbi.umd.edu/~haloweb/hma.html. Sequence data for *Haloferax volcanii* were downloaded from http://zdna2.umbi.umd.edu/~haloweb/hvo.html. To identify the conserved inverted repeat in *H. marismortui* and *H. volcanii*, BLAST analysis was performed on an SGI O_2 workstation with GCG computer software (Genetics Computer Group Inc.). Alignment of the conserved ARSs was done using CLUSTAL_X 1.81. Analysis of GC percentages in a sliding window was performed using the EMBOSS program DAN. Phylogenetic analysis of archaeal Orc1-Cdc6 proteins was performed using CLUSTAL_X for multiple sequence alignment and Treeview for neighbor joining tree construction (36).

RESULTS

Cloning and analysis of a chromosomal ARS element. To identify ARS elements from the large chromosome of *Halobacterium* strain NRC-1, a directed approach was used based on previous studies implying a physical association between archaeal replication origins and the *orc1-cdc6* gene (22, 27) and GC skew data obtained for the large chromosome (17). Selected regions of the genome proximal to the three chromosomal *orc1-cdc6* genes (*orc6*, *orc7*, and *orc8* [Fig. 1]) were cloned into an *E. coli* plasmid vector which was not able to replicate in haloarchaea and assayed for endowment with autonomous replication ability. The plasmids were constructed to initially contain both 1 kb 5′ to the respective *orc* gene start and 1 kb 3′ to the *orc* gene stop (pBBori*n* series, with *n* being 6, 7, and 8 for *orc6*, *orc7*, and *orc8*, respectively). Plasmid derivatives were constructed which deleted the entire downstream 3′ region 50 bp beyond the *orc* gene translational stop codon and deleted either 0, 250, 500, or 750 bp from 1 kb 5′ toward the *orc* gene translational start codon (pBB*n*G1000, -750, -500, or -250). Finally, constructs were made containing the 750, 500, or 250 bp 5′ to the *orc* gene, but with the deletion of the respective *orc* open reading frame (pBB*n*750, -500, or -250).

The three plasmid deletion series were used to transform a *Halobacterium* strain NRC-1 culture, and experiments were repeated to confirm transformation efficiencies. Transformation efficiencies are shown in Table 2, expressed as CFU per microgram of plasmid DNA. Several plasmids containing the *orc7* gene, pBBori7, pBB7G1000, and pBB7G750, showed high transformation efficiencies ($>10^4$ CFU/μg of transforming DNA), while those lacking more than 500 bp of the upstream region, pBB7G500 and pBB7G250, and those lacking the *orc7* gene, pBB7750, pBB7500, and pBB7250, as well as all the others derived from the *orc6* and *orc8* gene regions, had trans-

TABLE 1. Plasmids used in this study

Plasmid name	Description	Reference
pNGMEV101	pTZ19 vector with the *mev* gene from *Haloferax volcanii*	28
pNG168	pNGMEV101 with the minimal origin of replication from pNRC100	28
pBBori6	pNGMEV101 with a 3-kb fragment containing the *orc6* gene plus 1,000 bp 5′ to the gene start and 1,000 bp 3′ to the gene stop	This work
pBB6G1000	pNGMEV101 with a 2-kb fragment containing the *orc6* gene plus 1,000 bp 5′ to the gene start	This work
pBB6G750	pNGMEV101 with a 1.75-kb fragment containing the *orc6* gene plus 750 bp, 5′ to the gene start	This work
pBB6G500	pNGMEV101 with a 1.5-kb fragment containing the *orc6* gene plus 500 bp 5′ to the gene start	This work
pBB6G250	pNGMEV101 with a 1.25-kb fragment containing the *orc6* gene plus 250 bp 5′ to the gene start	This work
pBB6750	pNGMEV101 with a 750-bp fragment containing 750 bp 5′ to the gene start	This work
pBB6500	pNGMEV101 with a 500-bp fragment containing 500 bp 5′ to the gene start	This work
pBB6250	pNGMEV101 with a 250-bp fragment containing 250 bp 5′ to the gene start	This work
pBBori7	pNGMEV101 with a 3.5-kb fragment containing the *orc7* gene plus 1,000 bp 5′ to the gene start and 1,000 bp 3′ to the gene stop	This work
pBB7G1000	pNGMEV101 with a 2.5-kb fragment containing the *orc7* gene plus 1,000 bp 5′ to the gene start	This work
pBB7G750	pNGMEV101 with a 2.25-kb fragment containing the *orc7* gene plus 1,000 bp 5′ to the gene start	This work
pBB7G500	pNGMEV101 with a 2-kb fragment containing the *orc8* gene plus 750 bp 5′ to the gene start	This work
pBB7G250	pNGMEV101 with a 1.75-kb fragment containing the *orc8* gene plus 500 bp 5′ to the gene start	This work
pBB7750	pNGMEV101 with a 750-bp fragment containing 750 bp 5′ to the gene start	This work
pBB7500	pNGMEV101 with a 500-bp fragment containing 500 bp 5′ to the gene start	This work
pBB7250	pNGMEV101 with a 250-bp fragment containing 250 bp 5′ to the gene start	This work
pBBori8	pNGMEV101 with a 3-kb fragment containing the *orc8* gene plus 1,000 bp 5′ to the gene start and 1,000 bp 3′ to the gene stop	This work
pBB8G1000	pNGMEV101 with a 2-kb fragment containing the *orc8* gene plus 1,000 bp 5′ to the gene start	This work
pBB8G750	pNGMEV101 with a 1.75-kb fragment containing the *orc8* gene plus 1,000 bp 5′ to the gene start	This work
pBB8G500	pNGMEV101 with a 1.5-kb fragment containing the *orc8* gene plus 750 bp 5′ to the gene start	This work
pBB8G250	pNGMEV101 with a 1.25-kb fragment containing the *orc8* gene plus 500 bp 5′ to the gene start	This work
pBB8750	pNGMEV101 with a 750-bp fragment containing 750 bp 5′ to the gene start	This work
pBB8500	pNGMEV101 with a 500-bp fragment containing 500 bp 5′ to the gene start	This work
pBB8250	pNGMEV101 with a 250-bp fragment containing 250 bp 5′ to the gene start	This work

formation efficiencies 2 orders of magnitude lower, similar to background level. These results are consistent with autonomous replication of plasmids containing the *orc7* gene plus either 1,000 or 750 bp of 5′-flanking sequence and loss of replication ability when containing only 500 or 250 bp of 5′-flanking sequence or when the *orc7* gene was deleted. These results show that the *orc7* gene plus sequence elements between 500 and 750 bp upstream are required for autonomous replication.

For a qualitative demonstration of plasmid autonomous replication ability, nonselected cultures containing transformants were spotted in 50-μl aliquots on selective CM^+ agar plates containing mevinolin (Fig. 2). Patches of growth on plates provided confirmation that cultures containing plasmids with the *orc7* gene plus 750 bp or more 5′ to the gene transformed the host with high efficiency. In contrast, only background colonies were evident when only 500 or 250 bp upstream of the *orc7* gene was included. These background colonies are also seen for the other two *orc* gene constructs. The small amount of background growth could be attributed to two events: (i) spontaneous mutation to mevinolin resistance, by either increased gene expression or gene copy number of 3-hydroxy-3-methylglutaryl coenzyme A reductase (18), or (ii) acquisition of mevinolin resistance by plasmid integration.

To discriminate between plasmid integration and autonomous replication, total DNA was extracted from transformants for all of the plasmids. This DNA was then used in Southern blot analysis with a *bla* gene-specific probe for plasmid DNA

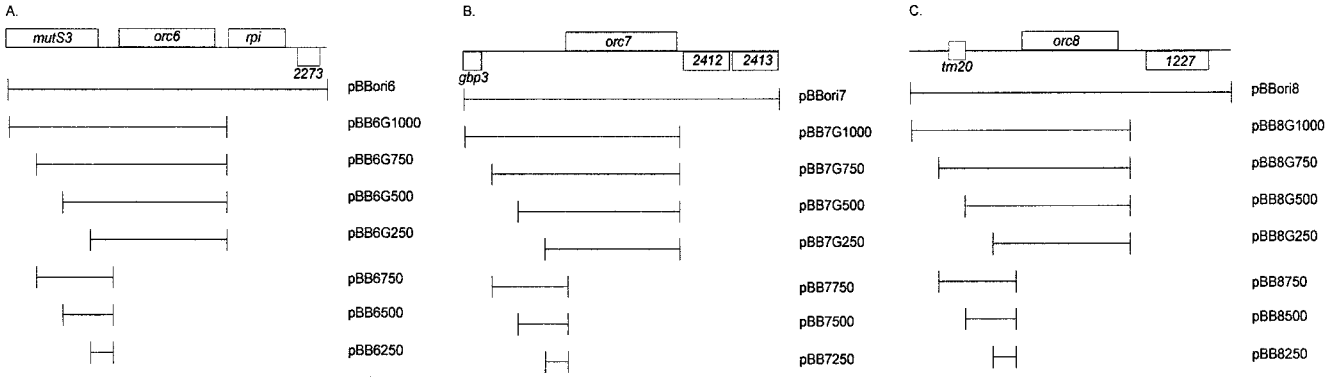

FIG. 1. Chromosomal *orc1-cdc6* loci and plasmid constructs for assaying ARS activity. The figure shows the *orc6* (A), *orc7* (B), and *orc8* (C) genetic loci and schematics of DNA inserted into plasmid pNGMEV101 to assay for autonomous replication ability. The extent of regions cloned into pNGMEV101 is shown by lines below the genetic map with names of constructed plasmids indicated.

TABLE 2. Transformation efficiencies for plasmids containing DNA sequences proximal to the *orc7* gene[a]

Plasmid name	Transformation efficiency (10^2 CFU/μg of DNA transformed)
pBBori7	500
pBB7G1000	260
pBB7G750	370
pBB7G500	5
pBB7G250	5
pBB7750	12
pBB7500	8
pBB7250	2
pNG168	250
pNGMEV101	1

[a] Transformation efficiencies for all other plasmid constructs listed in Table 1 ranged from 2×10^2 to 12×10^2 CFU/μg of DNA transformed.
[b] Transformation efficiencies were averaged from the results of two to three separate experiments.

(Fig. 3, lanes 1 to 5). Plasmid bands consistent with replicative supercoiled and nicked forms were observed in the blot for plasmid DNA for pBBori7, pBB7G1000, and pBB7G750, but not for pBB7G500 or pBB7G250, or other plasmids, where hybridization is to high-molecular-weight DNA, consistent with plasmid integration (Fig. 3). For comparison, the replicative supercoiled and nicked plasmid forms for pBBori7, pBB7G1000, pBB7G750, pBB7G500, and pBB7G250 purified from *E. coli* were visualized with ethidium bromide (Fig. 3, lanes 6 to 10). This confirms that the bands seen in lanes 1 through 3 are the replicative plasmid forms of pBBori7, pBB7G1000, and pBB7G750. Faint bands seen at lower-molecular-weight positions in lanes 4 and 5, for pBB7G500 and pBB7G250, are very likely the result of sequential plasmid integration and excision events, occurring due to large regions of homology (>1 kb) between the plasmids. This has been previously reported for other plasmids (28, 32). Little or no integration is observed for plasmids without extended regions of homology, e.g., pNGMEV101 (Table 2 and Fig. 2) (28). These results are consistent with the presence of an ARS element in pBBori7, pBB7G1000, and pBB7G750, where no evidence for integration could be observed. For *orc6* and *orc8* plasmid constructs, a majority of hybridization for Mev^r transformants is seen to be associated with chromosomal DNA, consistent with plasmid integration (data not shown). This analysis clearly shows that the *orc7* gene region, including 750 bp of upstream DNA, is the only one for which autonomous replication can be demonstrated.

Analysis of the ARS element DNA sequence. Based on the experimental replication evidence for *orc7* constructs, a closer examination of the DNA sequence located upstream of the *orc7* gene was undertaken. The GC percentage of the 1 kb 5′ of the *orc7* gene was calculated and plotted with a sliding window of 70 bp with 50-bp steps (Fig. 4). Immediately striking, from this analysis, was a stretch of extreme AT richness located ~300 to 500 bp from the *orc7* translational start. The GC content of this 189-bp region is only 56%, corresponding to a substantially higher AT content than the chromosomal average GC content of 68%. Also striking was the finding of a 31-bp nearly perfect (with one mismatch) inverted repeat flanking the AT-rich stretch. Significantly, autonomous replication ability is abolished when the upstream half of the inverted repeat is deleted in pBB7G500 (see above; Table 2 and Fig. 2).

Conservation of the ARS element in other distantly related halophiles. The availability of partial genome sequences for *H. marismortui* and *H. volcanii* allowed for comparative analysis to be performed with other distantly related halophilic archaea. The 1 kb upstream plus the *orc7* gene from *Halobacterium* strain NRC-1 was used as an initial query to identify a contig containing the *orc7* homolog and its upstream sequence in both *H. marismortui* and *H. volcanii*. Once the contigs were located, the 31-bp inverted repeat sequence from *Halobacterium* strain NRC-1 was used to search for any sequence similarity in comparable regions in *H. marismortui* and *H. volcanii*. A nearly identical inverted repeat was identified upstream of the *orc7* homologs in both organisms (Fig. 5). The inverted repeat in *H. marismortui* is 28 bp in length, containing two mismatches in sequence, and is separated by 249 bp. The GC content of the sequence between the inverted repeat is 48% compared to an average chromosomal GC content of 61% in *H. marismortui*. In *H. volcanii*, the inverted repeat is a perfect 28 bp in length, with a 2-bp insertion in the second half. The GC content of the sequence between the inverted repeat is 50% compared to a genomic average GC content of 65%. This analysis indicates that a putative chromosomal origin, containing a ~30-bp inverted repeat flanking a ~200-bp AT-rich region, is conserved in halophilic archaea.

Phylogenetic analysis of archaeal chromosomally encoded Orc1-Cdc6 proteins. The partial genomic sequences of *H. marismortui* and *H. volcanii* were examined to determine whether they contained a plurality of *orc1-cdc6* genes, similar to that found in the completed genome of *Halobacterium* strain NRC-1. These two other haloarchaea were found to contain a greater number of distinct *orc1-cdc6* genes than even *Halobac-*

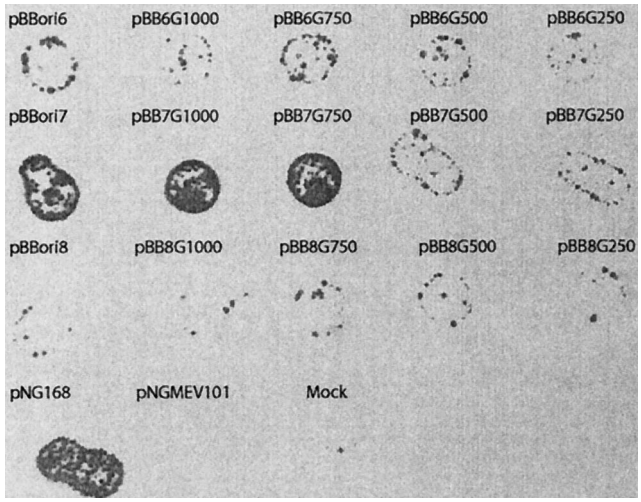

FIG. 2. ARS assay for *Halobacterium* strain NRC-1 chromosomal elements. *Halobacterium* strain NRC-1 cultures were transformed with plasmid constructs, containing DNA sequences proximal to and including each of the chromosomal *orc* genes, and spotted in 50-μl aliquots onto CM$^+$-mevinolin plates to test the replication abilities of those particular plasmids. Regions containing growth of transformants are seen as dense patches, and background is observed as individual colonies.

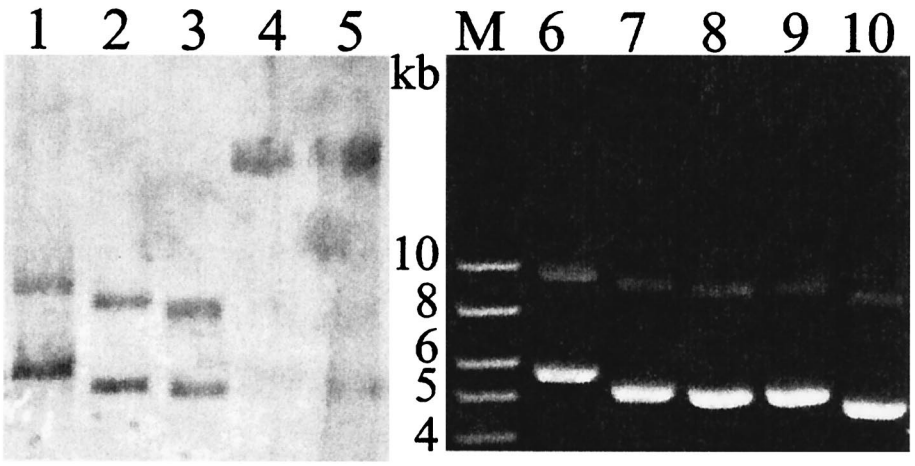

FIG. 3. Replication assay for *orc7* plasmids in *Halobacterium* strain NRC-1. (Left) Southern blot analysis of *orc7*-containing plasmids with a *bla* gene probe; (right) agarose gel of purified plasmids. Lanes 1 to 5 contain *Halobacterium* strain NRC-1 DNA transformed with plasmids (lane 1, pBBori7; lane 2, pBB7G1000; lane 3, pBB7G750; lane 4, pBB7G500; and lane 5, pBB7G250). Lanes 6 to 10 contain plasmid DNA purified from *E. coli* (lane 6, pBBori7; lane 7, pBB7G1000; lane 8, pBB7G750; lane 9, pBB7G500; and lane 10, pBB7G250) visualized with ethidium bromide. Lane M contains DNA ladder markers.

terium strain NRC-1, 15 each in *H. marismortui* and *H. volcanii*. By using the three NRC-1 chromosomal Orc1-Cdc6 proteins as queries, the most similar Orc6, Orc7, and Orc8 homologs were identified in the two other halophile sequences. To examine the relationships between chromosomally encoded archaeal Orc1-Cdc6 proteins, an alignment was made of these Orc1-Cdc6 proteins and a neighbor joining tree was constructed (Fig. 6). The tree was rooted using *Saccharomyces cerevisiae* Orc1p (1) and Cdc6p (19) along with the *Schizosaccharomyces pombe* homologs Orp1p (26) and Cdc18p (16), respectively. The neighbor joining tree displays three classes of Orc1-Cdc6, corresponding to the three Orc1-Cdc6 proteins encoded on the *Halobacterium* strain NRC-1 chromosome: Orc6, Orc7, and Orc8. Other archaeal organisms have Orc1-Cdc6 homologs located within either one or both of the clades that contain the haloarchaeal Orc6 and Orc7 homologs. Interestingly, only haloarchaea contain homologs in the Orc8 clade.

DISCUSSION

Previously, our laboratory reported the cloning of a putative replication origin from a library of the pNRC100 and pNRC200 minichromosomes of *Halobacterium* sp. strain NRC-1 by its property of conferring autonomous replication ability on nonreplicating plasmids (28). Now, we have cloned a putative replication origin of the third replicon, the large chromosome from strain NRC-1, also by virtue of its autonomous replication ability. Whereas the minichromosomal ARS contained the *repH* gene, similar to other haloarchaeal extrachromosomal replicons, the large chromosomal ARS contains the *orc7* gene, a relative of the eukaryotic ORC initiator gene family (1, 12, 13, 35). As a result of our genetic approach, we now have in hand a haloarchaeal system with well-defined ARS elements from each replicon.

Bioinformatic analysis of the *Halobacterium* strain NRC-1 genome by GC skew had suggested multiple replication origins of the large chromosome, coincident with two of the three chromosomal *orc1-cdc6* genes (17). Subsequent analysis by the Z-curve method confirmed this result (39). Therefore, we cloned segments of DNA surrounding the chromosomal *orc1-cdc6* genes into a plasmid which does not replicate in this haloarchaeon and tested for autonomous replication ability. Our results showed that a chromosomal region around the *orc7* gene, including an upstream inverted repeat flanking an AT-rich region plus the gene itself, is sufficient to endow plasmids with replication ability. Interestingly, however, the regions surrounding two other chromosomal *orc1-cdc6* genes, *orc6* and *orc8*, were not capable of autonomous replication. These results suggest that the *Halobacterium* sp. strain NRC-1 large chromosome may have a single replication origin associated with *orc7*, similar to other archaea and bacteria (21, 22, 23, 27). However, the possibility exists that other functional chromo-

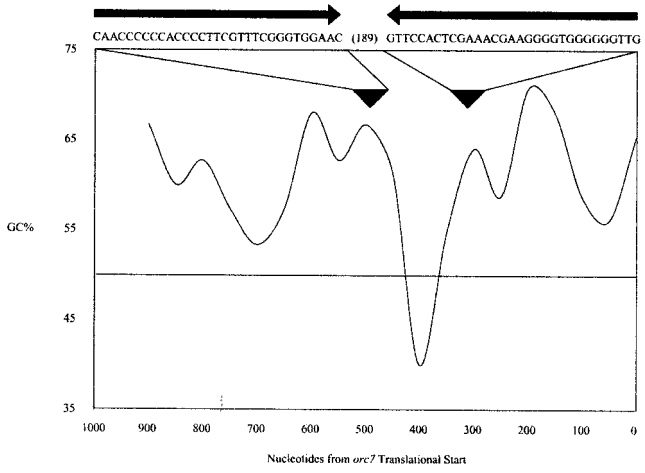

FIG. 4. GC percentage plot of the 1 kb upstream of the *orc7* gene. Plotted are GC contents of 70-mers in a 50-bp sliding window for the DNA sequence 1 kb 5′ to the *orc7* gene translational start. Shown above is the 31-bp inverted repeat sequence flanking a very AT-rich segment of DNA. Indicated by arrows are the positions of the inverted repeats. The horizontal line represents 50% GC.

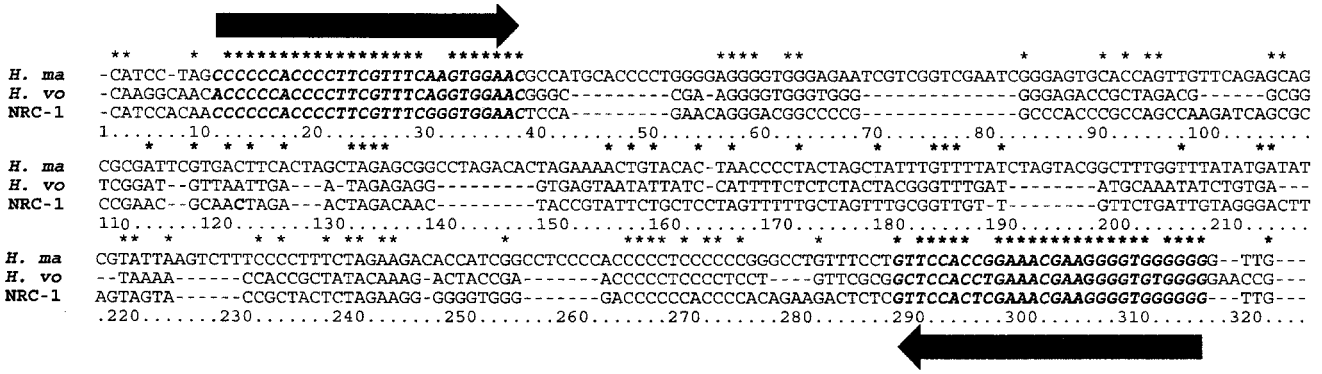

FIG. 5. Multiple sequence alignment of large inverted repeats upstream of *orc7* genes in *Halobacterium* strain NRC-1, *H. marismortui*, and *H. volcanii*. Shown are the aligned conserved ARS element sequences from three haloarchaea. In boldface italics and with large arrows above are the conserved large inverted repeats flanking the AT-rich region. Also shown in boldface, at position 121, is the endpoint for the insertion into pBB7G500.

somal replication origins may not confer the autonomous replication ability on plasmids or may not be associated with *orc1-cdc6* genes, and therefore, we cannot rule out the use of more than a single replication origin in vivo based on our present studies.

We genetically analyzed the *orc7* gene region by deletion analysis. While the downstream region was dispensable for autonomous replication, the upstream region and the gene itself were found to be essential. The upstream region contained a 31-bp inverted repeat flanking a 189-bp AT-rich region. When a 500-bp region including one copy of the inverted repeat was deleted, replication ability was lost, showing the importance of this region for function. Moreover, similar inverted repeat–AT-rich structures were found to be highly conserved upstream of *orc7* homologs in two other distantly related halophiles, *H. marismortui* and *H. volcanii*, consistent with a function in replication initiation for this sequence in haloarchaea. With the conservation of a large inverted repeat flanking an AT-rich region, it is tempting to speculate that these repeats serve as landing pads for archaeal Orc1-Cdc6 and other DNA replication initiation and possibly elongation proteins, with the AT-rich region acting as an initial DNA-unwinding element. All together, our results strongly support the idea that the *orc7* gene and upstream region function in chromosomal replication initiation in haloarchaea.

Although DNA replication origins have been well defined for bacteria and simple eukaryotes, they have remained relatively unstudied in the archaeal domain. In general, bacterial origins of replication contain multiple DnaA boxes and an AT-rich region proximal to the *dnaA* gene itself. The *E. coli oriC* contains five such DnaA boxes and an AT-rich region over 260 bp, while the *Streptomyces oriC* contains 19 DnaA boxes and five short AT-rich stretches over 600 bp (25). Among eukaryotic organisms, *S. cerevisiae* and *S. pombe* are the best studied. *S. cerevisiae* origins are ~100 bp in size and contain a conserved and essential AT-rich ARS consensus sequence, along with other essential B elements, which are less well conserved (2, 15). In *S. pombe* replication origins are much larger, ~1,000 bp, with only short (20- to 50-bp) AT-rich stretches of variable sequence being essential for origin function (2, 5). The haloarchaeal *orc7*-associated replication origin appears to be quite distinct from these examples.

Among archaea, in vivo identification of an early replicating chromosomal segment of *Pyrococcus abyssi* indicated that this organism initiates bidirectional DNA replication from a single region, located 5′ to an *orc1-cdc6* gene (27). Our phylogenetic analysis indicates that the *Pyrococcus* gene is most closely re-

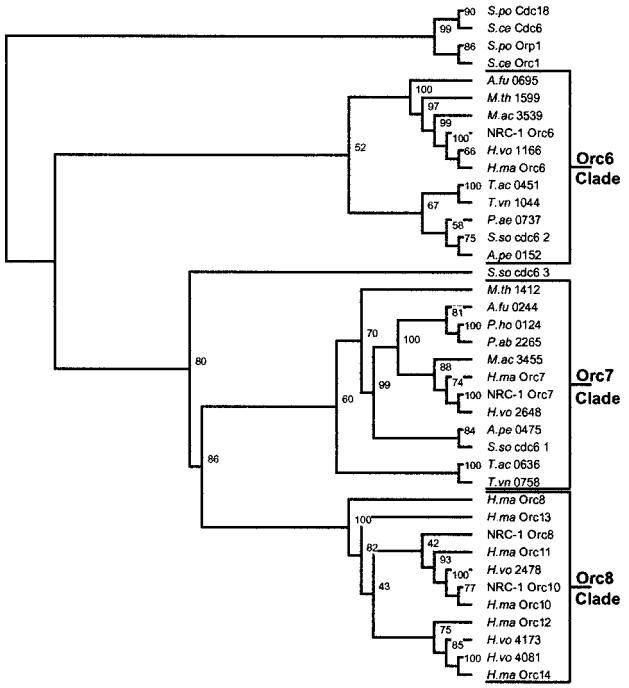

FIG. 6. Neighbor joining tree of Orc1-Cdc6 homologs found in sequenced archaeal genomes. Included in the tree are chromosomally encoded Orc1-Cdc6 homologs from all sequenced archaeal genomes to date and from the incomplete sequences of *H. marismortui* and *H. volcanii*. Homologs of Orc1, Cdc6, and Orp1 and Cdc18, from *S. cerevisiae* and *S. pombe*, respectively, are included as outgroups. Families of Orc1-Cdc6 genes are shown with the *Halobacterium* strain NRC-1 homolog name. Organism abbreviations are as follows: *S.ce*, *S. cerevisiae*; *S.po*, *S. pombe*; *A.fu*, *Archaeoglobus fulgidus*; *M.th*, *M. thermoautotrophicus*; *M.ac*, *Methanosarcina acetivorans*; NRC-1, *Halobacterium* strain NRC-1; *H.vo*, *H. volcanii*; *H.ma*, *H. marismortui*; *T.ac*, *Thermoplasma acidophilum*; *T.vn*, *Thermoplasma volcanium*; *P.ae*, *Pyrobaculum aerophilum*; *S.so*, *S. solfataricus*; and *A.pe*, *Aeropyrum pernix*.

lated to the *orc7* gene family. This origin region is conserved in three *Pyrococcus* species and contains a number of small repeats, which are also present in the genome of *Methanothermobacter thermoautotrophicus* (formerly *Methanobacterium thermoautotrophicum*) (20, 22). Recently, *Pyrococcus* species have also been reported to contain a large inverted repeat flanking an AT-rich region upstream of their *orc7* homologs (23). However, to date, there is no report of autonomous replication of these regions or genetic analysis (3). Based on our present knowledge, it appears that a large inverted repeat and an AT-rich region are general features of archaeal replication origins and that origins in archaea are conserved in structure but not always in sequence.

In *Halobacterium* strain NRC-1, 10 *orc1-cdc6* genes are carried within the genome, including four located on the large (2-Mb) chromosome (one is interrupted by an insertion sequence element), five on the pNRC200 replicon, and one on the pNRC100 replicon. This is in stark contrast to most other archaea, which typically carry one or two *orc1-cdc6* homologs, with the exception of *Sulfolobus solfataricus*, which carries three (34). In systems with a few Orc1-Cdc6 homologs, these have been proposed to act as origin recognition factors, initially binding to origin DNA sequences and serving to locally unwind the DNA and nucleate formation of a replisome, akin to DnaA in bacteria and the six-membered ORC complex in eukaryotes (13). For those containing two, it would seem plausible that one might act to bind origin sequences, i.e., ORC-DnaA function, and the other might act as Cdc6 does in eukaryotes and DnaC in bacteria, namely, recruiting the replicative helicase to the origin.

For halophilic archaea, the extremely large size of the *orc1-cdc6* family has been maintained throughout evolution. In addition to *Halobacterium* strain NRC-1, the two other halophiles examined here, *H. marismortui* and *H. volcanii*, carry at least 15 distinct *orc1-cdc6* homologs each (Zhang et al., unpublished). Our phylogenetic analysis shows the existence of several classes of Orc1-Cdc6 proteins from archaeal organisms (Fig. 6). In particular two of these classes contain representatives from nonhalophilic and halophilic archaea (Orc6 and Orc7 clades). It is likely that the Orc7 clade represents the origin binding proteins in archaeal organisms, based on the requirement of this gene for autonomous replication and its association with the origin of replication in *Pyrococcus* sp. (22, 23, 27). The Orc6 clade may possibly represent replicative helicase-recruiting proteins for organisms which possess both, based on its homology to Cdc6, the eukaryotic replicative helicase loader, and conservation in most sequenced archaeal organisms to date. For organisms containing a single Orc1-Cdc6 protein, this protein may act for both functions. Alternatively, other yet unknown proteins may act in the process of replication initiation.

For haloarchaea, the other Orc1-Cdc6 homologs may also aid in the replication initiation process, perhaps forming multisubunit complexes akin to the six-membered eukaryotic ORC complex or recruiting the replicative helicase and other replication-associated proteins. Alternatively, the other Orc1-Cdc6 homologs may play physiological roles different from replication initiation. An expansion of *orc1-cdc6* genes has occurred and is being maintained in halophilic lineages. It is possible that this may be an adaptation to living in saturating salt and containing a cytoplasm with an equally high salt concentration. This may be related to the speed or processivity of DNA polymerase or unknown activities required for life in saturating salt.

The results presented here provide a significant advance in our understanding of genetic elements essential for initiation of chromosomal DNA replication in the archaeal domain of life. We provide the first report of cloning of an ARS element from an archaeal chromosome and display genetically the DNA regions necessary to provide autonomous replication ability. We also show the sequence and structure conservation of this chromosomal origin in two other distantly related halophilic archaea. *Halobacterium* strain NRC-1 provides an excellent model system to study the characteristics of archaeal replication origins, and the results presented here provide the groundwork for further studies of essential elements for archaeal DNA replication initiation.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation grants MCB-0296017, -0135595, and -0196502.

REFERENCES

1. **Bell, S. P., and B. Stillman.** 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature **357:**128–134.
2. **Bell, S. P., and A. Dutta.** 2002. DNA replication in eukaryotic cells. Annu. Rev. Biochem. **71:**333–374.
3. **Bernander, R.** 2003. The archaeal cell cycle: current issues. Mol. Microbiol. **48:**599–604.
4. **Cann, I. K., and Y. Ishino.** 1999. Archaeal DNA replication: identifying the pieces to solve a puzzle. Genetics **152:**1249–1267.
5. **Clyne, R. K., and T. J. Kelly.** 1995. Genetic analysis of an ARS element from the fission yeast *Schizosaccharomyces pombe*. EMBO J. **14:**6348–6357.
6. **DasSarma, S., and E. M. Fleischman.** 1995. Archaea: a laboratory manual, vol. 1. Halophiles. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
7. **DasSarma, S., S. P. Kennedy, B. Berquist, W. V. Ng, N. S. Baliga, J. L. Spudich, M. P. Krebs, J. A. Eisen, C. H. Johnson, and L. Hood.** 2001. Genomic perspective on the photobiology of *Halobacterium* species NRC-1, a phototrophic, phototactic, and UV-tolerant haloarchaeon. Photosynth. Res. **70:**3–17.
8. **DasSarma, S., and P. Arora.** 2002. Halophiles, p. 458–466. *In* Encyclopedia of life sciences, vol. 8. Nature Publishing Group, London, United Kingdom.
9. **DasSarma, S.** Genome sequence of an extremely halophilic archaeon. *In* C. M. Fraser, T. Read, and K. E. Nelson (ed.), Microbial genomes, in press. Humana Press, Inc., Totowa, N.J.
10. **Edgell, D. R., and W. F. Doolittle.** 1997. Archaea and the origin(s) of DNA replication proteins. Cell **89:**995–998.
11. **Gaasterland, T.** 1999. Archaeal genomics. Curr. Opin. Microbiol. **2:**542–547.
12. **Giraldo, R., and R. Diaz-Orejas.** 2001. Similarities between the DNA replication initiators of Gram-negative bacterial plasmids (RepA) and eukaryotes (Orc4p)/archaea (Cdc6p). Proc. Natl. Acad. Sci. USA **98:**4938–4943.
13. **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.
14. **Grigoriev, A.** 1998. Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res. **26:**2286–2290.
15. **Huang, R. Y., and D. Kowalski.** 1993. A DNA unwinding element and an ARS consensus comprise a replication origin within a yeast chromosome. EMBO J. **12:**4521–4531.
16. **Kelly, T. J., G. S. Martin, S. L. Forsburg, R. J. Stephen, A. Russo, and P. Nurse.** 1993. The fission yeast *cdc18*$^{+}$ gene product couples S phase to START and mitosis. Cell **74:**371–382.
17. **Kennedy, S. P., W. V. Ng, S. L. Salzberg, L. Hood, and S. DasSarma.** 2001. Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. Genome Res. **11:**1641–1650.
18. **Lam, W. L., and W. F. Doolittle.** 1992. Mevinolin-resistant mutations identify a promoter and the gene for a eukaryote-like 3-hydroxy-3-methylglutaryl-coenzyme A reductase in the archaebacterium *Haloferax volcanii*. J. Biol. Chem. **267:**5829–5834.
19. **Lisziewicz, J., A. Godany, D. V. Agoston, and H. Kuntzel.** 1988. Cloning and characterization of the *Saccharomyces cerevisiae* CDC6 gene. Nucleic Acids Res. **16:**11507–11520.

20. **Lopez, P., H. Philippe, H. Myllykallio, and P. Forterre.** 1999. Identification of putative chromosomal origins of replication in Archaea. Mol. Microbiol. **32:**883–886.
21. **Maisnier-Patin, S., L. Malandrin, N. K. Birkeland, and R. Bernander.** 2002. Chromosome replication patterns in the hyperthermophilic euryarchaea *Archaeoglobus fulgidus* and *Methanocaldococcus* (*Methanococcus*) *jannaschii*. Mol. Microbiol. **45:**1443–1450.
22. **Matsunaga, F., P. Forterre, Y. Ishino, and H. Myllykallio.** 2001. In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. Proc. Natl. Acad. Sci. USA **98:**11152–11157.
23. **Matsunaga, F., C. Norais, P. Forterre, and H. Myllykallio.** 2003. Identification of short ′eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. EMBO Rep. **4:**154–158.
24. **McLean, M. J., K. H. Wolfe, and K. M. Devine.** 1998. Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. J. Mol. Evol. **47:**691–696.
25. **Messer, W.** 2002. The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. FEMS Microbiol. Rev. **26:**355–374.
26. **Muzi-Falconi, M., and T. J. Kelly.** 1995. Orp1, a member of the Cdc18/Cdc6 family of S-phase regulators, is homologous to a component of the origin recognition complex. Proc. Natl. Acad. Sci. USA **92:**12475–12479.
27. **Myllykallio, H., P. Lopez, P. Lopez-Garcia, R. Heilig, W. Saurin, Y. Zivanovic, H. Philippe, and P. Forterre.** 2000. Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. Science **288:**2212–2215.
28. **Ng, W. L., and S. DasSarma.** 1993. Minimal replication origin of the 200-kilobase *Halobacterium* plasmid pNRC100. J. Bacteriol. **175:**4584–4596.
29. **Ng, W. V., S. A. Ciufo, T. M. Smith, R. E. Bumgarner, D. Baskin, J. Faust, B. Hall, C. Loretz, J. Seto, J. Slagel, L. Hood, and S. DasSarma.** 1998. Snapshot of a large dynamic replicon in a halophilic archaeon: megaplasmid or minichromosome? Genome Res. **8:**1131–1141.
30. **Ng, W. V., S. P. Kennedy, G. G. Mahairas, B. Berquist, M. Pan, H. D. Shukla, S. R. Lasky, N. S. Baliga, V. Thorsson, J. Sbrogna, S. Swartzell, D. Weir, J. Hall, T. A. Dahl, R. Welti, Y. A. Goo, B. Leithauser, K. Keller, R. Cruz, M. J. Danson, D. W. Hough, D. G. Maddocks, P. E. Jablonski, M. P. Krebs, C. M. Angevine, H. Dale, T. A. Isenbarger, R. F. Peck, M. Pohlschroder, J. L. Spudich, K. W. Jung, M. Alam, T. Freitas, S. Hou, C. J. Daniels, P. P. Dennis, A. D. Omer, H. Ebhardt, T. M. Lowe, P. Liang, M. Riley, L. Hood, and S. DasSarma.** 2000. Genome sequence of *Halobacterium* species NRC-1. Proc. Natl. Acad. Sci. USA **97:**12176–12181.
31. **Olsen, G. J., and C. R. Woese.** 1997. Archaeal genomics: an overview. Cell **89:**991–994.
32. **Peck, R. F., S. DasSarma, and M. P. Krebs.** 2000. Homologous gene knockout in the archaeon *Halobacterium salinarum* with *ura3* as a counterselectable marker. Mol. Microbiol. **35:**667–676.
33. **Sambrook, J., and D. W. Russell.** 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
34. **She, Q., R. K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M. J. Awayez, C. C. Chan-Weiher, I. G. Clausen, B. A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P. M. Gordon, I. Heikamp-de Jong, A. C. Jeffries, C. J. Kozera, N. Medina, X. Peng, H. P. Thi-Ngoc, P. Redder, M. E. Schenk, C. Theriault, N. Tolstrup, R. L. Charlebois, W. F. Doolittle, M. Duguet, T. Gaasterland, R. A. Garrett, M. A. Ragan, C. W. Sensen, and J. Van der Oost.** 2001. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. Proc. Natl. Acad. Sci. USA **98:**7835–7840.
35. **Spradling, A. C.** 1999. ORC binding, gene amplification, and the nature of metazoan replication origins. Genes Dev. **13:**2619–2623.
36. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
37. **Uemori, T., Y. Sato, I. Kato, H. Doi, and Y. Ishino.** 1997. A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization. Genes Cells **2:**499–512.
38. **Woese, C. R., and G. E. Fox.** 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA **74:**5088–5090.
39. **Zhang, R., and C. T. Zhang.** 2003. Multiple replication origins of the archaeon *Halobacterium* species NRC-1. Biochem. Biophys. Res. Commun. **302:**728–734.