Extremophiles (2004) 8:253–258
DOI 10.1007/s00792-004-0385-4

ORIGINAL PAPER

Ren Zhang · Chun-Ting Zhang

Identification of replication origins in the genome of the methanogenic archaeon, *Methanocaldococcus jannaschii*

Received: 29 September 2003 / Accepted: 10 February 2004 / Published online: 9 April 2004

Abstract *Methanocaldococcus jannaschii* has been notorious as an archaeon in which the replication origins are difficult to identify. Although extensive efforts have been exerted on this issue, the locations of replication origins still remain elusive 7 years after the publication of its complete genome sequence in 1996. Ambiguous results were obtained in identifying the replication origins of *M. jannaschii* based on all theoretical and experimental approaches. In the genome of *M. jannaschii*, we found that an ORF (MJ0774), annotated as a hypothetical protein, is a homologue of the Cdc6 protein. The position of the gene is at a global minimum of the *x* component of the *Z* curve, i.e., RY disparity curve, which has been used to identify replication origins in other Archaea. In addition, an intergenic region (694,540–695,226 bp) that is between the *cdc6* gene and an adjacent ORF shows almost all the characteristics of known replication origins, i.e., it is highly rich in AT composition (80%) and contains multiple copies of repeat elements and AT stretches. Therefore, these lines of evidence strongly suggest that the identified region is a replication origin, which is designated as *oriC1*. The analysis of the *y* component of the *Z* curve, i.e., MK disparity curve, suggests the presence of another replication origin corresponding to one of the peaks in the MK disparity curve at around 1,388 kb of the genome.

Keywords Archaea · *cdc6* gene · *Methanocaldococcus jannaschii* · Replication origin · *Z* curve

Communicated by G. Antranikian

R. Zhang
Department of Epidemiology and Biostatistics,
Tianjin Cancer Institute and Hospital, 300060 Tianjin, China

C.-T. Zhang (✉)
Department of Physics, Tianjin University,
300072 Tianjin, China
E-mail: ctzhang@tju.edu.cn
Fax: +86-22-27402697

Introduction

Methanocaldococcus jannaschii is an autotroph that lives under an extreme condition, which "not long ago, no one would have believed you if you'd told them such organisms existed on Earth" (Morell 1996). *M. jannaschii* grows at pressures of more than 200 atm and at temperatures up to 94°C (Jones et al. 1983). The completed sequencing of the *M. jannaschii* genome in 1996 prompted a flurry of excitement in the scientific community (Bult et al. 1996; Morell 1996), not only because of its unusual living environment, but also (most importantly) because it was the first archaeon that was sequenced. The availability of the complete genome sequence of *M. jannaschii* led to an unprecedented opportunity to investigate the evolutionary relationship between Archaea and the other two domains of life—Bacteria and Eukaryotes (Bult et al. 1996; Woese and Fox 1977).

M. jannaschii has been notorious as an archaeon in which the replication origins are difficult to identify. Although extensive efforts have been exerted on this issue, the locations of replication origins still remain elusive 7 years after the publication of its complete genome sequence in 1996 (Bult et al. 1996). Ambiguous results were obtained in identifying the replication origins of *M. jannaschii* based on all in silico attempts, which usually assess the biases in nucleotides, oligomer usage, or codon usage (Lopez et al. 1999; Rocha et al. 1999; Salzberg et al. 1998). Recently, a technique called marker frequency analysis has been successfully applied in vivo to identify the location of the replication origin of the archaeon *Archaeoglobus fulgidus*; it failed, however, in the case of *M. jannaschii* (Maisnier-Patin et al. 2002). Nevertheless, an intriguing alternative explanation is that the negative result of the marker frequency analysis in *M. jannaschii* may be due to the existence of multiple replication origins (Maisnier-Patin et al. 2002).

We have used the *Z*-curve method to identify replication origins of the archaea *Methanosarcina mazei*, *Halobacterium* species NRC-1, and *Sulfolobus* *solfataricus* (Zhang and Zhang 2002, 2003). One of the two predicted replication origins in *Halobacterium* species NRC-1 was later confirmed by experimental evidence (Berquist and DasSarma 2003). Multiple replication origins of *S. solfataricus* suggested based on the *Z*-curve analysis are also consistent with recent experimental results (Robinson et al. 2004). The *Z* curve is a three-dimensional curve that constitutes a unique representation of a DNA sequence, i.e., for the *Z* curve and the given DNA sequence, each can be uniquely reconstructed from the other. Based on the *Z* curve, it has been shown that a DNA sequence is uniquely described by three independent distributions, i.e., those of the bases of purine/pyrimidine, amino/keto, and weak/strong H-bonds, respectively (Zhang and Zhang 1991, 1994). In this paper, the *Z*-curve method was used to identify the replication origins in the genome of *M. jannaschii*. Our results strongly suggest that we have identified a replication origin. In addition, the existence of another replication origin is also suggested based on the *Z*-curve analysis.

Materials and methods

The genomic sequence of *Methanocaldococcus jannaschii* was downloaded from http://www.ncbi.nlm.nih.gov.

The *Z* curve is composed of a series of nodes $P_0, P_1, P_2, \ldots, P_N$ (N = length of the DNA sequence), whose coordinates x_n, y_n, and z_n ($n = 0, 1, 2, \ldots$) are uniquely determined by the formulas (Zhang and Zhang 1991, 1994):

$$\begin{cases} x_n = (A_n + G_n) - (C_n + T_n) \equiv R_n - Y_n, \\ y_n = (A_n + C_n) - (G_n + T_n) \equiv M_n - K_n, \\ z_n = (A_n + T_n) - (C_n + G_n) \equiv W_n - S_n, \end{cases} \tag{1}$$
$$n = 0,\ 1,\ 2, \ldots,\ N,\ x_n,\ y_n,\ z_n \in [-N,\ N],$$

where A_n, C_n, G_n, and T_n are the cumulative occurrence numbers of A, C, G, and T, respectively, in the subregion from the first to the *n*th base in the sequence. The *Z* curve is defined as the connection of the nodes P_0, P_1, P_2, . . . , $P_{N,}$ one by one, sequentially with straight lines. The three components x_n, y_n, and z_n of the *Z* curve display the distributions of purine/pyrimidine (R/Y), amino/keto (M/K), and strong-H bond/weak-H bond (S/W) bases along the sequence, respectively. The *x* and *y* components are termed RY disparity and MK disparity curves, respectively. The RY and MK disparity curves, as well as AT and GC disparity curves [$(x_n+y_n)/2$ and $(x_n+y_n)/2$, respectively], can be used to predict replication origins (Zhang et al. 2003).

Results

The RY disparity curve for the genome of *Methanocaldococcus jannaschii*

For the replication origins (*oriC*s) that have been known in Archaea, the RY or MK disparity curves show a peak at the locations of the corresponding *oriC*s (Zhang and Zhang 2002, 2003). The RY or MK disparity curves display the distributions of purine/pyrimidine and amino/keto, respectively, along genomes. Therefore, for the known *oriC*s, the purine/pyrimidine or amino/keto are asymmetrically distributed around the *oriC*s. For the genome of *Methanocaldococcus jannaschii*, the RY disparity curve displays a single global minimum at the position about 695 kb, which means that at about 695 kb, the genome undergoes a change from a CT-rich region to an AG-rich one (Fig. 1). The MK disparity curve shows globally four peaks that include a peak at the same position with that of the RY disparity curve. The analysis of MK disparity curve will be discussed in another section. Based on the behaviors of the *Z* curves for the Archaea in which the *oriC*s have been identified, the location around the global minimum of the RY disparity curve of *M. jannaschii* is a high-interest candidate region of a potential *oriC*.

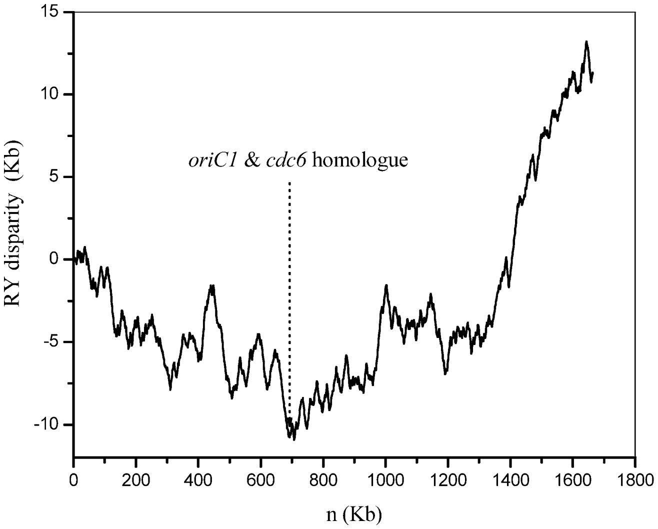


Fig. 1 The RY disparity curve for *Methanocaldococcus jannaschii*. The RY disparity curve displays a single global minimum at the poison of about 695 kb, which means that at this site, the genome undergoes a change from a CT-rich region to an AG-rich one. A *cdc6* homologue was found at the peak of the RY disparity curve, suggesting that the region around the global minimum contains a replication origin (*oriC*), based on the behaviors of the *Z* curves for the known *oriC*s

Existence of a Cdc6 homologue at the peak of the RY disparity curve

Usually, a *cdc6* gene is found close to the *oriC*. This is very similar to the case of bacteria, in which a gene coding for DnaA is frequently close to the *oriC*. We scanned the region around the peak for a potential *cdc6* gene, although it was previous thought that no Cdc6 homologue exists in *M. jannaschii* (Bernander 1998, 2000). Surprisingly, we found that an ORF, MJ0774, is highly similar to the *cdc6* gene (Fig. 2). The ORF MJ0774 encodes a 409-amino acid long polypeptide and is annotated as a hypothetical protein. We searched the amino acid sequence against the National Center for Biotechnology Information Conserved Domain Database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), and a Cdc6 protein was assigned to MJ0774 segment from amino acids 13 to 404. The alignment of the MJ0774 (13–404) with the consensus sequence of Cdc6 proteins (12–355) showed that MJ0774 is a homologue of the Cdc6 protein (Fig. 2). In addition, a helix-turn-helix domain was found at the region from residue 327–403, and this domain is believed to be involved in the DNA binding (Liu et al. 2000).

Identification of a replication origin

For most *oriC*s in Archaea, the locations are often found to be in an intergenic sequence adjacent to the *cdc6* genes. The intergenic sequence that contains the minimal *oriC* is often highly rich in A and T bases. The high AT composition is thought to be necessary to facilitate the melting of the DNA double helix by the proteins containing helicase activity. In addition, most identified *oriC*s contain various forms of repeat elements, which are thought to be required for DNA binding (Kelman 2000; Tye 2000).

We then investigated the regions around the *cdc6* homologue. Indeed, there is a roughly 700-bp region close to the *cdc6* homologue that shows many characteristics of *oriC*s. This intergenic region is between the ORF MJ0773 and MJ0774, from 694,540 – 695,226 bp of the genome. The region contains 687 bp and is highly rich in AT content (80%). In addition, we have also found six copies of the direct repeat element tttgattcat. There are also two copies of the consecutive repeat tattgtttgattcatgagatttttaat, which are flanked by two AT stretches. There are two copies of the consecutive repeat gtagataatta, which are followed by an AT stretch (Fig. 3). A consensus-repeat sequence was defined based on the *oriC*s of *Pyrococcus* and *Methanobacterium* (Lopez et al. 1999). However, the repeat sequences in the *oriC* of *M. jannaschii* show no similarity with the above-mentioned consensus-repeat sequence as well as the sequence close to the initiation starting point of *Pyrococcus abyssi* (Matsunaga et al. 2003).

The identified region has the characteristics that it is rich in AT composition and contains multiple copies of repeats as well as AT stretches. In addition, the region is at a global minimum of the RY disparity curve and adjacent to a *cdc6* homologue. This region has almost all the characteristics of known *oriC*s; therefore, it strongly suggests that the region is an *oriC*, which is designated as *oriC1*.

Features of the MK disparity curve for *M. jannaschii*

Recently, a technique called marker frequency analysis has been successfully applied in vivo to identify the location of an *oriC* of the archaeon *Archaeoglobus fulgidus. M. jannaschii,* however, displayed marker frequency distributions with multiple peaks and valleys (Maisnier-Patin et al. 2002). Nevertheless, an intriguing alternative explanation proposed by the authors is that the complex pattern of the marker frequency distributions for *M. jannaschii* is due to the presence of multiple *oriC*s (Maisnier-Patin et al. 2002). Indeed, the features of the MK disparity curve for *M. jannaschii* are consistent with this hypothesis.

The MK disparity curve for *M. jannaschii* shows four peaks that include the one associated with the identified *oriC1* (Fig. 4). The locations of these peaks are 127, 695 (*oriC1*), 986, and 1,388 kb, respectively. The MK disparity curve for the genome of *Halobacterium* sp. NRC-1 also shows four peaks, and based on this, two *oriC*s and

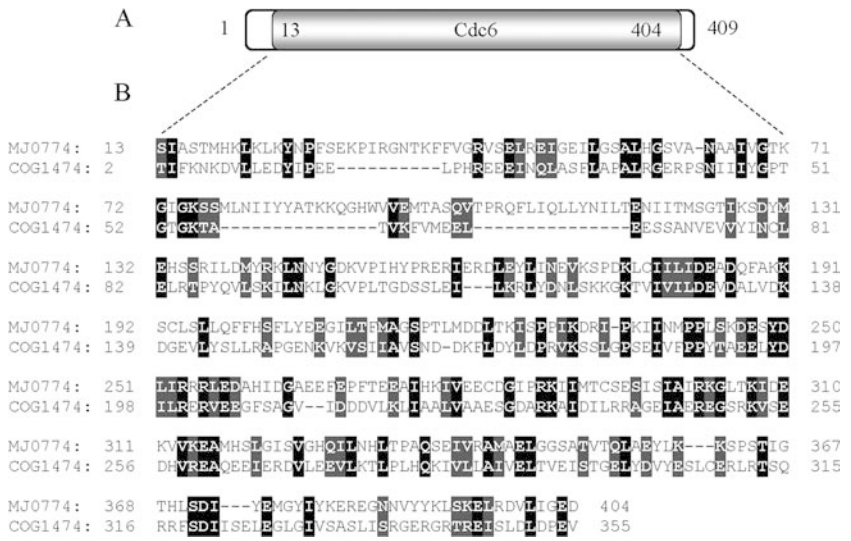
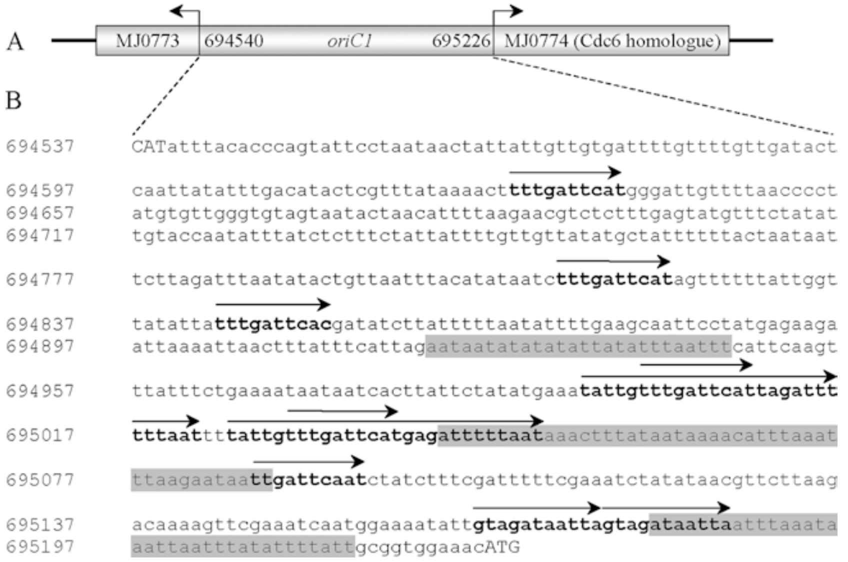

A
1
13
Cdc6
404
409
B
MJ0774: 13 SIASTMHKLKLKYNPFSEKPIRGNTKFFVGRVSELREIGEILGSALHGSVA-NAAIVGTK 71
COG1474: 2 TIFKNKDVLLEDYIPEE-----------LPHREEEINQLASFLAPALRGERPSNIIIYGPT 51
MJ0774: 72 GIGKSSMLNIIYYATKKQGHWVVEMTASQVTPRQFLIQLLYNILTENIITMSGTIKSDYM 131
COG1474: 52 GTGKTA---------------TVKFVMEEL---------------EESSANVEVVYINCL 81
MJ0774: 132 EHSSRILDMYRKLNNYGDKVPIHYPRERIERDLEYLINEVKSPDKLCIILIDEADQFAKK 191
COG1474: 82 ELRTPYQVLSKILNKLGKVPLTGDSSLEI---LKRLYDNLSKKGKTVIVILDEVDALVDK 138
MJ0774: 192 SCLSLLQFFHSFLYEEGILTFMAGSPTLMDDLTKISPPIKDRI-PKIINMPPLSKDESYD 250
COG1474: 139 DGEVLYSLLRAPGENKVKVSIIAVSND-DKFLDYLDPRVKSSLGPSEIVFPPYTAEELYD 197
MJ0774: 251 LIRRRLEDAHIDGAEEFEPFTEEAIHKIVEECDGIPRKIIMTCSESISIAIRKGLTKIDE 310
COG1474: 198 ILRERVEEGFSAGV--IDDDVLKLIAALVAAESGDARKAIDILRRAGEIAEREGSRKVSE 255
MJ0774: 311 KVVKEAMHSLGISVGHQILNHLTPAQSEIVRAMAELGGSATVTQLAEYLK---KSPSTIG 367
COG1474: 256 DHVREAQEEIERDVLEEVLKTLPLHQKIVLLAIVELTVEISTGELYDVYESLCERLRTSQ 315
MJ0774: 368 THLSDI---YEMGYIYKEREGNNVYYKLSKELRDVLIGED 404
COG1474: 316 RRFSDIISELEGLGIVSASLISRGERGRTREISLDLDPEV 355

Fig. 2 **A** Schematic diagram of MJ0774. A homologous search was performed in the National Center for Biotechnology Information Conserved Domain Database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), and a Cdc6 protein was assigned to MJ0774 as a homologous counterpart with the highest score. In addition, the segment from residues 327–403 contains a helix-turn-helix domain (not shown), which may mediate DNA binding. (Figure not drawn to scale.) **B** Alignment of MJ0774 with the Cdc6 consensus sequence with the expect value (*E*) = 4e-13. Those residues that are *shaded in black* are identical, and those *shaded in gray* are interchangeable, based on substitution matrices



```
694537   CATatttacacccagtattcctaataactattattgttgtgattttgttttgttgatact
694597   caattatatttgacatactcgtttataaaactttgattcatgggattgttttaacccct
694657   atgtgttgggtgtagtaatactaacattttaagaacgtctctttgagtatgtttctatat
694717   tgtaccaatatttatctctttctattattttgttgttatatgctatttttttactaataat
694777   tcttagatttaatatactgttaatttacatataatctttgattcatagtttttttattggt
694837   tatattatttgattcacgatatcttattttttaatattttgaagcaattcctatgagaaga
694897   attaaaattaactttatttcattagaataatatatatattatatttaatttcattcaagt
694957   ttatttctgaaaataataatcacttattctatatgaaatattgtttgattcattagattt
695017   tttaatttttattgtttgattcatgagatttttaataaactttataataaaacatttaaat
695077   ttaagaataattgattcaatctatctttcgattttcgaaatctatataacgttcttaag
695137   acaaaagttcgaaatcaatggaaaatattgtagataattagtagataattaatttaaata
695197   aattaatttatattttattgcggtggaaacATG
```

Fig. 3 **A** Schematic diagram of the *oriC* of *M. jannaschii*. The *oriC* is within the intergenic region of MJ0773 and MJ0774, which is a Cdc6 homologue. *Arrows* indicate the directions of transcription. (Figure is not drawn to scale.) **B** The sequence of the *oriC*. The sequence is highly rich in A and T residues, with an AT content of 80%. The start codons of the ORF MJ0773 and MJ0774, respectively, are *capitalized*. There are six copies of the repeat tttgattcat; two consecutive repeats of tattgtttgattcatgagatttttaat, which are flanked by two AT stretches; and two consecutive repeats of gtagataatta, which are followed by an AT stretch. The repeat elements are shown in *boldface*, whereas the AT stretches are *shaded*. *Arrows* indicate directions of the repeats

replication termini (*terC*s) were predicted (Zhang and Zhang 2003). The overall shape of the MK disparity curve for *M. jannaschii* is similar to that of *Halobacterium* sp. NRC-1. By studying the positions of the four peaks, it is possible that the peak at 1,388 kb is associated with another *oriC*, whereas the peaks at 127 and 986 kb correspond to *terC*s. We noticed that the lengths between the position of the peak at 1,388 kb and the two *terC*s are exactly the same, i.e., 402 kb, and this is consistent with the characteristics of most identified *oriC*s, i.e., in the genomes with a single *oriC*, the *oriC* and *terC* basically divide the whole genomes into two parts of similar lengths. However, we also noticed that the lengths between the *oriC1* and two predicted *terCs* are not the same. It has already been known that some horizontally transferred elements are present in the genome of *M. jannaschii* (Bult et al. 1996). Although the exact amount of horizontally transferred DNA is not clear, these horizontal transfer events could be a reason that the two replichores have different sizes, i.e., the horizontally transferred DNA increased the length of one of the replichores. In addition, a gene coding for replication factor C (MJ1422) is situated at the position of the peak of putative *oriC2*. However, now there is no evidence to suggest that the gene coding for replication factor C is close to *oriC*s. Nevertheless, some archaeal origins are indeed situated in the regions close to some replication factors, such as DNA polymerases and helicases (Salzberg et al. 1998).

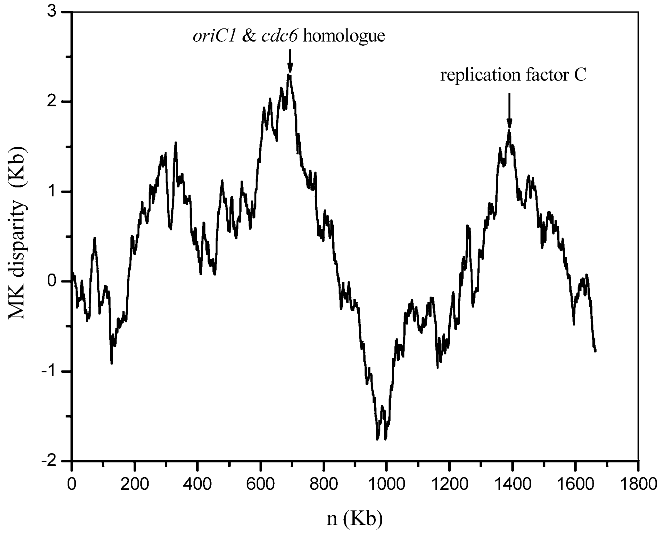


Fig. 4 The MK disparity curve for *M. jannaschii,* which shows four peaks, including the one that is associated with the identified *oriC*, *oriC1*. The locations of these peaks are 127, 695 (*oriC1*), 986, and 1,388 kb, respectively. It is possible that the peak at 1,388 kb is associated with another *oriC*, whereas the peaks at 127 and 986 kb are associated with replication termini (*terC*s). Note that the lengths between the position of the peak at 1,388 kb and two *terC*s are exactly the same, i.e., 402 kb. However, the lengths between the *oriC1* and two predicted *terCs* are not the same. One possible explanation is that the horizontal transferred DNA could increase the length of one of the replichores. A gene coding for replication factor C (MJ1422) is situated at the position of the peak at 1,388 kb. Refer to the text for details

Discussion

It is previously thought that *Methanocaldococcus jannaschii* is the only archaeon that lacks a clear Cdc6 homologue (Bernander 1998, 2000). However, the ORF MJ0774 indeed aligns well with the Cdc6 consensus sequence. This fact suggests that for all the Archaea whose genome sequences are available, there are either one or multiple copies of Cdc6 homologues. Therefore, it appears that the *cdc6* gene is an indispensable gene required for DNA replication in the archaeal domain.

During the past several years, great advancement has been made in the understanding of the replication process of the Archaea (Bohlke et al. 2002; Cann and Ishino 1999; Edgell and Doolittle 1997; Kelman and Kelman 2003). The replication of genomes starts with the recognition of the minimal *oriC* by some initiation factors, which in turn recruit the proteins having helicase activity to unwind the DNA double helix. This is followed by the further recruitment of other replication machinery to continue the replication process. It is still not clear in Archaea which protein functions as the replication initiation factor. It has been recently demonstrated that Cdc6 binds specifically to *oriC* in vivo, providing the first in vivo evidence that Cdc6 may function as a replication initiation factor, i.e., an origin binding protein (Matsunaga et al. 2001). This is very similar to the case in bacteria, in which DnaA is the replication initiation factor, which is also found to be close to the *oriC* (Kornberg and Baker 1992). Indeed, both Cdc6 and DnaA belong to the AAA$^+$ superfamily of ATPases (Davey et al. 2002). The close proximity of the *oriC1* and Cdc6 homologue in *M. jannaschii* adds another line of evidence to support the role of Cdc6 as an initiator. The reason that *cdc6* is close to the *oriC* is not fully understood, although it was hypothesized that the juxtaposition could facilitate the formation of the replication complex at *oriC*s (Lopez et al. 1999).

The underlying mechanism for the asymmetry of the RY and MK disparity curves (or AT and GC disparity curves) around the *oriC*s and *terC*s should be due to the asymmetrical replication pattern of DNA replication. For instance, on the leading strand the DNA is synthesized continuously, whereas on the lagging strand the DNA is synthesized discontinuously, i.e., short Okazaki fragments are synthesized and joined in this process. The asymmetrical process, such as the enzymes involved for different strands and structural difference in replication forks, may cause profound impact on the error and repair rate of the two strands, which eventually leads to the asymmetry of the nucleotide composition around *oriC*s and *terC*s (Francino and Ochman 1997; Frank and Lobry 1999). However, it is not clear why for some *oriC*s, both RY and MK disparity curves have the same peaks, whereas for others, only the RY or MK disparity curve shows the peak corresponding to *oriC*s and *terC*s. Nevertheless, no matter which components, the asymmetrical pattern of the *Z* curve clearly suggests the potential regions of *oriC*s and *terC*s.

In summary, a homologue of the *cdc6* gene was found in the genome of *M. jannaschii*, and the position of the gene is at a global minimum of the RY disparity curve and at one of the peaks of the MK disparity curve. In addition, an intergenic region that is between the *cdc6* gene and an adjacent ORF shows many characteristics of *oriC*s, i.e., highly rich in AT content (80%), and contains multiple copies of repeat elements and AT stretches. Therefore, it is strongly suggests that an *oriC* in the genome of *M. jannaschii* has been identified. Moreover, based on the behaviors of the MK disparity curve, it suggests that another *oriC* is located around 1,388 kb.

Acknowledgements The authors thank the anonymous referees for their constructive comments. The present study was supported in part by the 973 Project of China (grant 1999075606).

References

Bernander R (1998) Archaea and the cell cycle. Mol Microbiol 29:955–961

Bernander R (2000) Chromosome replication, nucleoid segregation and cell division in archaea. Trends Microbiol 8:278–283

Berquist BR, DasSarma S (2003) An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, *Halobacterium* sp. strain NRC-1. J Bacteriol 185:5959–66

Bohlke K, Pisani FM, Rossi M, Antranikian G (2002) Archaeal DNA replication: spotlight on a rapidly moving field. Extremophiles 6:1–14

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

Cann IK, Ishino Y (1999) Archaeal DNA replication: identifying the pieces to solve a puzzle. Genetics 152:1249–1267

Davey MJ, Jeruzalmi D, Kuriyan J, O'Donnell M (2002) Motors and switches: AAA+ machines within the replisome. Nat Rev Mol Cell Biol 3:826–835

Edgell DR, Doolittle WF (1997) Archaea and the origin(s) of DNA replication proteins. Cell 89:995–998

Francino MP, Ochman H (1997) Strand asymmetries in DNA evolution. Trends Genet 13:240–245

Frank AC, Lobry JR (1999) Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. Gene 238:65–77

Jones WJ, Leigh JA, Mayer F, Woese CR, Wolfe RS (1983) *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. Arch Microbiol 136:254–261

Kelman LM, Kelman Z (2003) Archaea: an archetype for replication initiation studies? Mol Microbiol 48:605–615

Kelman Z (2000) The replication origin of archaea is finally revealed. Trends Biochem Sci 25:521–523

Kornberg A, Baker TA (1992) DNA replication. Freeman, New York

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

Maisnier-Patin S, Malandrin L, Birkeland NK, Bernander R (2002) Chromosome replication patterns in the hyperthermophilic euryarchaea *Archaeoglobus fulgidus* and *Methanocaldococcus (Methanococcus) jannaschii*. Mol Microbiol 45:1443–1450

Matsunaga F, Forterre P, Ishino Y, Myllykallio H (2001) In vivo interactions of archaeal Cdc6/*orc1* and minichromosome maintenance proteins with the replication origin. Proc Natl Acad Sci USA 98:11152–11157

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

Morell V (1996) Life's last domain. Science 273:1043–1045

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

Rocha EP, Danchin A, Viari A (1999) Universal replication biases in bacteria. Mol Microbiol 32:11–16

Salzberg SL, Salzberg AJ, Kerlavage AR, Tomb JF (1998) Skewed oligomers and origins of replication. Gene 217:57–67

Tye BK (2000) Insights into DNA replication from the third domain of life. Proc Natl Acad Sci USA 97:2399–2401

Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA 74:5088–5090

Zhang CT, Zhang R (1991) Analysis of distribution of bases in the coding sequences by a diagrammatic technique. Nucleic Acids Res 19:6313–6317

Zhang CT, Zhang R, Ou HY (2003) The Z curve database: a graphic representation of genome sequences. Bioinformatics 19:593–599

Zhang R, Zhang CT (1994) Z curves, an intuitive tool for visualizing and analyzing the DNA sequences. J Biomol Struct Dyn 11:767–782

Zhang R, Zhang CT (2002) Single replication origin of the archaeon *Methanosarcina mazei* revealed by the Z curve method. Biochem Biophys Res Commun 297:396–400

Zhang R, Zhang CT (2003) Multiple replication origins of the archaeon *Halobacterium* species NRC-1. Biochem Biophys Res Commun 302:728–734