Minireview
Where does DNA replication start in archaea?
Amit Vas and Janet Leatherwood

Address: Department of Molecular Genetics and Microbiology, State University of New York, Stony Brook, NY 11794-5222, USA.
E-mail: janet.leatherwood@sunysb.edu

Abstract

Genome-wide measures of DNA strand composition have been used to find archaeal DNA replication origins. Archaea seem to replicate using a single origin (as do eubacteria) even though archaeal replication factors are more like those of eukaryotes.

Use of genome sequences is a powerful art that goes beyond finding protein homologs: it has changed how we can approach basic biological questions. This is particularly apparent for the enigmatic archaebacteria. Here, more than for other organisms, available genome data far exceed traditional biological study. A recent striking example of the insights that can be gained from archaeal genomics is provided by a report in *Science* from Myllykallio *et al.* [1] showing the use of DNA strand compositional bias, or GC skew, to find the likely replication origin in three *Pyrococcus* species. One reason for widespread interest in archaeal replication origins is the similarity between the factors involved in DNA replication in archaea and eukaryotes. Archaeal homologs of eukaryotic replication factors and DNA polymerase suggest that archaebacteria could become an important model to aid understanding of eukaryotic DNA replication.
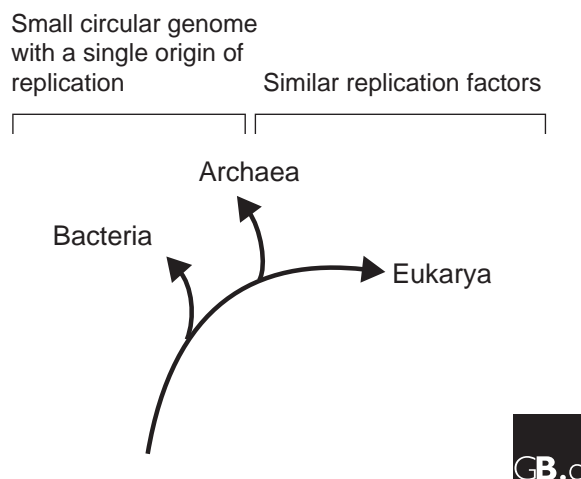
Where do archaea fit?

Are archaea like humans or bacteria? This was the issue raised when archaeal genome sequencing revealed some areas of surprising similarity between these prokaryotes and eukaryotes. Though archaeal metabolism and operon gene organization is certainly most similar to prokaryotic eubacteria, the archaeal factors for transcription, translation and DNA replication seem more akin to those found in eukaryotes. Thus, the third kingdom, archaebacteria, might serve

as a simple model for mechanisms of eukaryotic cell function. And we are left wondering just how much these prokaryotes resemble ourselves. (For more extensive reviews of this issue see [2-6]).

Archaea (as exemplified by *Pyrococcus* sp.) replicate their circular genome from a single DNA replication origin as do bacteria, even though they may use eukaryotic-like proteins to do so (Figure 1; [1]). This single-origin replication is definitely un-human, as our DNA replication depends on initiation at thousands of different origins. The multiple sites of initiation are essential for timely replication of large eukaryotic genomes. The archaebacterial *Pyrococcus* genomes by contrast are smaller even than that of *Escherichia coli*, so perhaps we should not be surprised that archaea can replicate like *E. coli* using a single origin.

Identifying a replication origin in archaea may be more important than finding whether they use one origin or many. In fact, *E. coli* lacking RNaseH start replication at many different sites, yet these multiple replication origins are not at all eukaryotic-like [7]. The observation that a single origin is used fails to enlighten us as to the mechanism of initiation, but identification of an origin does provide one of the most powerful tools for future studies of DNA replication initiation. Myllykallio *et al.* [1] reported genomic analyses that strongly suggest a well conserved 600 base-pair sequence is the replication origin in three related archaea [1,8].
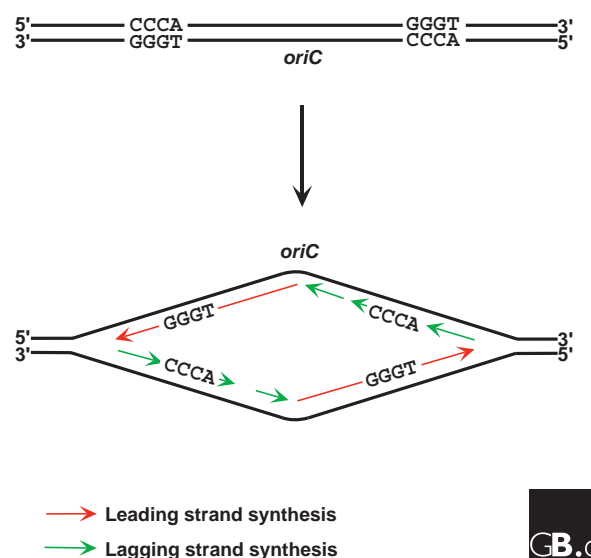
Figure 1
Evolutionary relationships between bacteria, archaea and eukaryotes takes into account the similarities between archaea replication, transcription and translation factors with eukaryotic factors. The report by Myllykallio *et al.* [1] shows that the archaea share chromosome organization and replication pattern with prokaryotes although they use many eukaryotic-like factors to duplicate their chromosomes.

Figure 2
A bi-directional replication fork: DNA replicated by lagging-strand synthesis on one side of the origin will be replicated by leading-strand synthesis on the other side. In bacteria, there is a switch in the strand bias of guanine content at the origin. Myllykallio *et al.* [1] measured the strand bias resulting from GGGT as illustrated.

The importance of GC skew

How can nucleotide sequence be used to find replication origins? In most eubacteria, there is a statistical overrepresentation of guanine (G) in DNA of the leading strand and more cytosine (C) in DNA of the lagging strand. This GC skew changes sign at the replication origin and terminus (Figure 2), though this change is most notable at the origin as termination may occur in a wider region. What creates GC skew is poorly understood but could include differences in errors, damage, and/or repair for lagging versus leading strand synthesis. Skewed distribution of short sequences may be even more predictive of bacterial origins than GC skew alone [9]. Myllykallio *et al.* [1] measured strand distribution of the tetramer GGGT across several archaea genomes to look for a singularity where GGGT skew changed sign. This occured at the same place in genomes of three related *Pyrococcus* species analyzed, suggesting this region may be the replication origin for these organisms.

Several archaeal genomes lack obvious GC skew that would indicate a single DNA replication origin [10,11]. This has fueled speculation that archaea use multiple origins for DNA replication and could provide clues to the selection and use of many replication origins in eukaryotes. But GC skew is not clear for all eubacteria and may be obscured by biological constraints of factor binding sites and gene coding sequences (and see McLean *et al.* [11] for a discussion of these issues and a clear, thoughtful skew analysis of 12 prokaryotes including three archaea). By practicing on

bacteria with known origins, scientists and mathematicians are finding countless ways to count Gs and Cs and have predicted single origins for some archaea [9-13]. So what is special about the report by Myllykallio *et al.*?

Myllykallio *et al.* did three important things. First, they obtained a signal revealing skewed strand distribution of nucleotides and the skew changed sign abruptly at one position in the genome (the putative origin). Use of a tetramer (GGGT) and mathematics to smooth out local variations were required to see a signal at all. Since it is not understood exactly what causes nucleotide skew and is even less clear what causes skewed distribution of short sequences, finding a signal was only the first step.

Second, they exploited the awesome power of comparative genomics. They compared three fully sequenced *Pyrococcus* species. Nucleotide skew as well as other properties predict origin location in the same place in all three genomes. For instance, a prediction for bacteria-like replication is that genes encoding replication factors will be clustered around the replication origin. Most notably, the gene for the bacterial replication initiator *dnaA* is so consistently linked to the origin that it is predictive of origin location. Several of the *Pyrococcus* replication factors cluster around the predicted origin, including Orc1/Cdc6, which resembles the putative eukaryotic initiator origin recognition complex and, therefore, is analogous to DnaA. Finally, the bacterial replication terminus is a hot spot for

rearrangement [14,15] and comparative genomics reveals this to be true for the three *Pyrococcus* species studied by Myllykallio *et al.* [1].

Third and most importantly, they tested the hypothesis derived by computer or *in silico* experiments using old-fashioned laboratory experimentation. They grew these third kingdom creatures — keeping them warm at 95°C — labeled newly synthesized DNA *in vivo*, and then determined which genome regions replicate first and last. As with any good story, all the pieces fit. Tetramer skew analysis predicted origins in the same place for all three species: the region identified has a highly conserved intergenic sequence that might bind replication factors [8], and, finally, DNA replication was shown to begin in this putative origin region.

Towards a mechanism

The species studied by Myllykallio *et al.* [1] performs the incredible feat of replicating its genome at 95°C — a temperature hot enough to melt DNA duplexes. It is amazing that replication under extreme conditions, using many eukaryotic-like factors that differ considerably from bacterial replication machinery, could result in the bacterial GC skew. Perhaps the same skew will be detectable in eukaryotes once we better understand how to filter out biological noise and focus on chromosomal regions that are replicated most often by a fork passing in a single direction.

Myllykallio *et al.* [1] used their information to calculate replication fork movement to be at a rate of 20 kilobases per minute. This is slower than DNA replication in *E. coli*, but is still ten times faster than fork movement in eukaryotes [7]. Archaea have a DNA polymerase resembling eukaryotic DNA polymerases, but they also have their own unique DNA polymerase [16]; perhaps this unique archael polymerase is required for the faster movement of the replication fork. Nevertheless, many factors for initiation and replication fork function in archaebacteria have clear homologs in eukaryotes.

Many scientists trying to decipher eukaryotic DNA replication are leaping at the chance to study something a bit simpler. Archaeal replication proteins look far more like eukaryotic replication proteins than those of eubacteria and there are fewer of them. For example, eukaryotic DNA replication requires six related mini-chromosome maintenance (MCM) proteins named MCM2-7, but the archaeal bacterium *Methanobacter thermoautotrophicum* has only one MCM homolog [17]; eukaryotic DNA replication requires the replication factor C complex, comprised of five proteins, while *M. thermoautotrophicum* has only two subunits for replication factor C [18]; eukaryotic DNA replication initiation requires the origin recognition complex (ORC) and Cdc6 (three ORC subunits and Cdc6 share sequence similarity [19]), whereas the *M. thermoautotrophicum* has only two ORC/Cdc6-like subunits [20].

Archaeal replication is thought to be the evolutionary precursor of eukaryotic replication, so archaea may use homo-oligomers evolved from an ancestral factor, where eukaryotes use hetero-oligomeric complexes of related proteins evolved from this same ancestral factor. Biochemical studies are already yielding valuable data from studying the simpler archaeal systems. For example, archaebacterial MCM has helicase function *in vitro* [21-23], and this lends strong support to the hypothesis that MCMs in eukaryotes function as a replicative helicase [24], an idea that is reviewed elsewhere [17,25]. In another example, the archaeal homolog of DNA polymerase alpha subunit p50 has primase activity *in vitro*, strongly supporting the long-held, but never proven, hypothesis that p50 is the catalytic subunit of eukaryotic DNA polymerase alpha primase [26]. Information on archaeal replication origins will be critical in reconstituting an *in vitro* archaebacterial replication system.

Myllykallio *et al.* [1] provide the best evidence to date that archaebacteria replicate DNA from a single origin. If this is indeed so, these organisms have no need to coordinate replication initiations at various sites. But they must still couple replication with growth and division; and how they do so is an interesting puzzle. *E. coli* achieves such regulation via SeqA, a negative regulator of replication initiation [27-29]. Eukaryotes do so via the 'master' cell-cycle regulators, the cyclin-dependent kinases (Cdks) and cyclins, and eukaryotic replication is also dependent on the replication specific kinase Cdc7 [25,30]. Archaea lack recognizable homologs of SeqA or cyclin dependent kinases, cyclins, or the kinase Cdc7. Perhaps archaebacteria have their own ways to couple replication, growth, and division, which may be achieved by some of the proteins encoded by the 50% of archaeal coding sequences specific to archaea.

References

1. Myllykallio H, Lopez P, Lopez-Garcia P, Heilig R, Saurin W, Zivanovic Y, Philippe H, Forterre P: **Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon.** *Science* 2000, **288**:2212-2215.
2. Dennis PP: **Ancient ciphers: translation in Archaea.** *Cell* 1997, **89**:1007-1010.
3. Edgell DR, Doolittle WF: **Archaea and the origin(s) of DNA replication proteins.** *Cell* 1997, **89**:995-998.
4. Belfort M, Weiner A: **Another bridge between kingdoms: tRNA splicing in archaea and eukaryotes.** *Cell* 1997, **89**:1003-1006.
5. Reeve JN, Sandman K, Daniels CJ: **Archaeal histones, nucleosomes, and transcription initiation.** *Cell* 1997, **89**:999-1002.
6. Olsen GJ, Woese CR: **Archaeal genomics: an overview.** *Cell* 1997, **89**:991-994.
7. Kornberg A, Baker TA: *DNA Replication.* New York: W. H. Freeman and Company, 1992.
8. **Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon - Supplementary material** [www.sciencemag.org/feature/data/1048232.shl]
9. Salzberg SL, Salzberg AJ, Kerlavage AR, Tomb JF: **Skewed oligomers and origins of replication.** *Gene* 1998, **217**:57-67.
10. Mrazek J, Karlin S: **Strand compositional asymmetry in bacterial and large viral genomes.** *Proc Natl Acad Sci U S A* 1998, **95**: 3720-3725.

11. McLean MJ, Wolfe KH, Devine KM: **Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes.** *J Mol Evol* 1998, **47**:691-696.

12. Karlin S: **Bacterial DNA strand compositional asymmetry.** *Trends Microbiol* 1999, **7**:305-308.

13. Lobry JR: **Asymmetric substitution patterns in the two DNA strands of bacteria.** *Mol Biol Evol* 1996, **13**:660-665.

14. Bierne H, Michel B: **When replication forks stop.** *Mol Microbiol* 1994, **13**:17-23.

15. Louarn JM, Louarn J, Francois V, Patte J: **Analysis and possible role of hyperrecombination in the termination region of the *Escherichia coli* chromosome.** *J Bacteriol* 1991, **173**:5097-5104.

16. Cann IK, Ishino Y: **Archaeal DNA replication: identifying the pieces to solve a puzzle.** *Genetics* 1999, **152**:1249-1267.

17. Tye BK: **Insights into DNA replication from the third domain of life.** *Proc Natl Acad Sci U S A* 2000, **97**:2399-2401.

18. Kelman Z, Hurwitz J: **A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon *Methanobacterium thermoautotrophicum* deltaH.** *J Biol Chem* 2000, **275**:7327-7336.

19. Quintana DG, Dutta A: **The metazoan origin recognition complex.** *Front Biosci* 1999, **4**:D805-D815.

20. Smith DR, Doucette-Stamm LA, Deloughery C, Lee H, Dubois J, Aldredge T, Bashirzadeh R, Blakely D, Cook R, Gilbert K, *et al.*: **Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics.** *J Bacteriol* 1997, **179**:7135-7155.

21. Shechter DF, Ying CY, Gautier J: **The intrinsic DNA helicase activity of *Methanobacterium thermoautotrophicum* delta H minichromosome maintenance protein.** *J Biol Chem* 2000, **275**:15049-15059.

22. Kelman Z, Lee JK, Hurwitz J: **The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* DeltaH contains DNA helicase activity.** *Proc Natl Acad Sci U S A* 1999, **96**:14783-14788.

23. Chong JP, Hayashi MK, Simon MN, Xu RM, Stillman B: **A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase.** *Proc Natl Acad Sci U S A* 2000, **97**:1530-1535.

24. Ishimi Y: **A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex.** *J Biol Chem* 1997, **272**:24508-24513.

25. Leatherwood J: **Emerging mechanisms of eukaryotic DNA replication initiation.** *Curr Opin Cell Biol* 1998, **10**:742-748.

26. Desogus G, Onesti S, Brick P, Rossi M, Pisani FM: **Identification and characterization of a DNA primase from the hyperthermophilic archaeon *Methanococcus jannaschii*.** *Nucleic Acids Res* 1999, **27**:4444-4450.

27. Lu M, Campbell JL, Boye E, Kleckner N: **SeqA: a negative modulator of replication initiation in *E. coli*.** *Cell* 1994, **77**:413-426.

28. Torheim NK, Skarstad K: ***Escherichia coli* SeqA protein affects DNA topology and inhibits open complex formation at oriC.** *EMBO J* 1999, **18**:4882-4888.

29. Wold S, Boye E, Slater S, Kleckner N, Skarstad K: **Effects of purified SeqA protein on oriC-dependent DNA replication in vitro.** *EMBO J* 1998, **17**:4158-4165.

30. Sclafani RA: **Cdc7p-Dbf4p becomes famous in the cell cycle.** *J Cell Sci* 2000, **113**:2111-2117.