Molecular Microbiology (1999) **32**(2), 437–445

Physical mapping of an origin of bidirectional replication at the centre of the *Borrelia burgdorferi* linear chromosome

Mathieu Picardeau,[1†] Jean R. Lobry[2] and B. Joseph Hinnebusch[1*]

[1]NIH, NIAID, Rocky Mountain Laboratories, Laboratory of Microbial Structure and Function, 903 S. 4th St., Hamilton, MT 59840, USA.

[2]Laboratoire de Biométrie, CNRS UMR 5558, Université Claude Bernard, Villeurbanne, France.

Summary

The *Borrelia burgdorferi* chromosome is linear, with telomeres characterized by terminal inverted repeats and covalently closed single-stranded hairpin loops. The replication mechanism of these unusual molecules is unknown. Previous analyses of bacterial chromosomes for which the complete sequence has been determined, including that of *B. burgdorferi*, revealed an abrupt switch in polarity of CG skew at known or putative origins of replication. We used nascent DNA strand analysis to physically map the *B. burgdorferi* origin to within a 2 kb region at the centre of the linear chromosome, and to show that replication proceeds bidirectionally from this origin. The results are consistent with replication models in which termination occurs at the telomeres after bidirectional, symmetrical elongation from the central origin. Sequences typical of origins of other bacterial chromosomes were not found at the origin of this spirochete. The most likely location of the replication origin of the linear chromosome is the 240 bp sequence between *dnaA* and *dnaN* where the switch in CG skew occurs.

Introduction

The Lyme disease spirochete *Borrelia burgdorferi* and other members of the genus have an unusual genome composed of a linear chromosome of approximately 1 Mb and several linear and circular plasmids. The linear chromosome and linear plasmids of *Borrelia* have covalently closed single-stranded hairpin loops at their ends and short inverted terminal repeats (Barbour and Garon, 1987; Hinnebusch and Barbour, 1991; Casjens *et al*., 1997). This genome structure is also found in a few eukaryotic viruses, the prophage plasmid of bacteriophage N15 of *Escherichia coli*, and some mitochondria (Hinnebusch and Tilly, 1993). The linear structure of these elements raises questions about their replication mechanism. Characterizing the replication origins of the linear chromosome and linear plasmids will be essential to understand this process. Although a genetic organization suggestive of a bacterial origin was found at the middle of the *B. burgdorferi* chromosome (Old *et al*., 1993), identification of replication origins in spirochetes has been hampered by a lack of genetic tools.

At present, the complete nucleotide sequences of more than 13 bacterial genomes have been reported. Analysis of these genomes has identified a characteristic strand compositional asymmetry (Lobry, 1996a; 1996b; Francino and Ochman, 1997; Freeman *et al*., 1998; Grigoriev, 1998; Mrazek and Karlin, 1998). In an unbiased situation, the base frequencies along the entire length of each chromosomal strand tend to be $[A] = [T]$ and $[C] = [G]$ (Lobry, 1995; Sueoka, 1995). However, significant deviations from this have been observed in all cases characterized by a $G > C$ base frequency in the leading strand and $C > G$ in the lagging strand. Thus, the CG skew of each chromosomal strand, calculated as $(C - G)/(C + G)$, switches polarity (from positive to negative or vice versa) at the origin and terminus of DNA replication. This global genomic feature can, therefore, be used to predict the location of the replication origin in bacteria for which it has not been experimentally determined. Analysis of the recently published chromosome sequence of *B. burgdorferi* showed a clear switch of the CG skew at the centrally located putative origin of replication (Fraser *et al*., 1997).

To provide direct experimental proof for a central location of the *B. burgdorferi* origin of replication, we used nascent DNA strand analysis (Vassilev and Johnson, 1989). This technique has become a standard method to accurately map eukaryotic replication origins (Vassilev and Johnson, 1989, 1990; Vassilev *et al*., 1990; Taira *et al*., 1994), but it has not previously been applied to prokaryotes. The method is very sensitive and does not require synchronized cells, but does require an a priori candidate origin

Received 16 December, 1998; accepted 29 January, 1999. †Present address: Laboratoire de Référence des Mycobactéries, Institut Pasteur, Paris, France. *For correspondence. E-mail joe_hinnebusch@nih.gov; Tel. (+1) 406 363 9260; Fax (+1) 406 363 9371.

in a region of known sequence, criteria furnished by CG skew and other sequence analysis (Old *et al*., 1993; Fraser *et al*., 1997). The method is, thus, well suited to the study of initiation of replication in *B. burgdorferi* and other bacteria. We first demonstrated the validity of this method by using the *E. coli* chromosome as a model. The physical mapping results for *B. burgdorferi*, together with CG skew analysis, demonstrate that the initiation site of the linear chromosome is centrally located within a region containing *dnaA* and *dnaN*, and that replication proceeds bidirectionally from that site.

Results

CG skew analysis of the E. coli *and* B. burgdorferi *chromosomes*

The 4.7 Mb *E. coli* chromosome shows a switch in polarity of CG skew from negative to positive at the terminus of replication, and from positive to negative at the origin of replication (Blattner *et al*., 1997; Mrazek and Karlin, 1998). When only the third base of each codon is considered, as in Fig. 1A, the CG skew is more pronounced. The rationale for this is that most base substitutions in this wobble position are synonymous and so are subject to a weaker counter-selective pressure (Blattner *et al*., 1997). Last codon positions, therefore, better record overall base composition biases, regardless of their source. In contrast to CG skew, AT skew was found to be weak (data not shown). A closer look at the *oriC* region showed that the CG skew polarity switch occurs a few hundred base pairs downstream of *oriC* (Fig. 1B). The polarity switch is gradual and was most obvious when 100 kb sequence windows were analysed. Examined in a slightly different way, when cumulative CG and AT skew for each gene along the entire

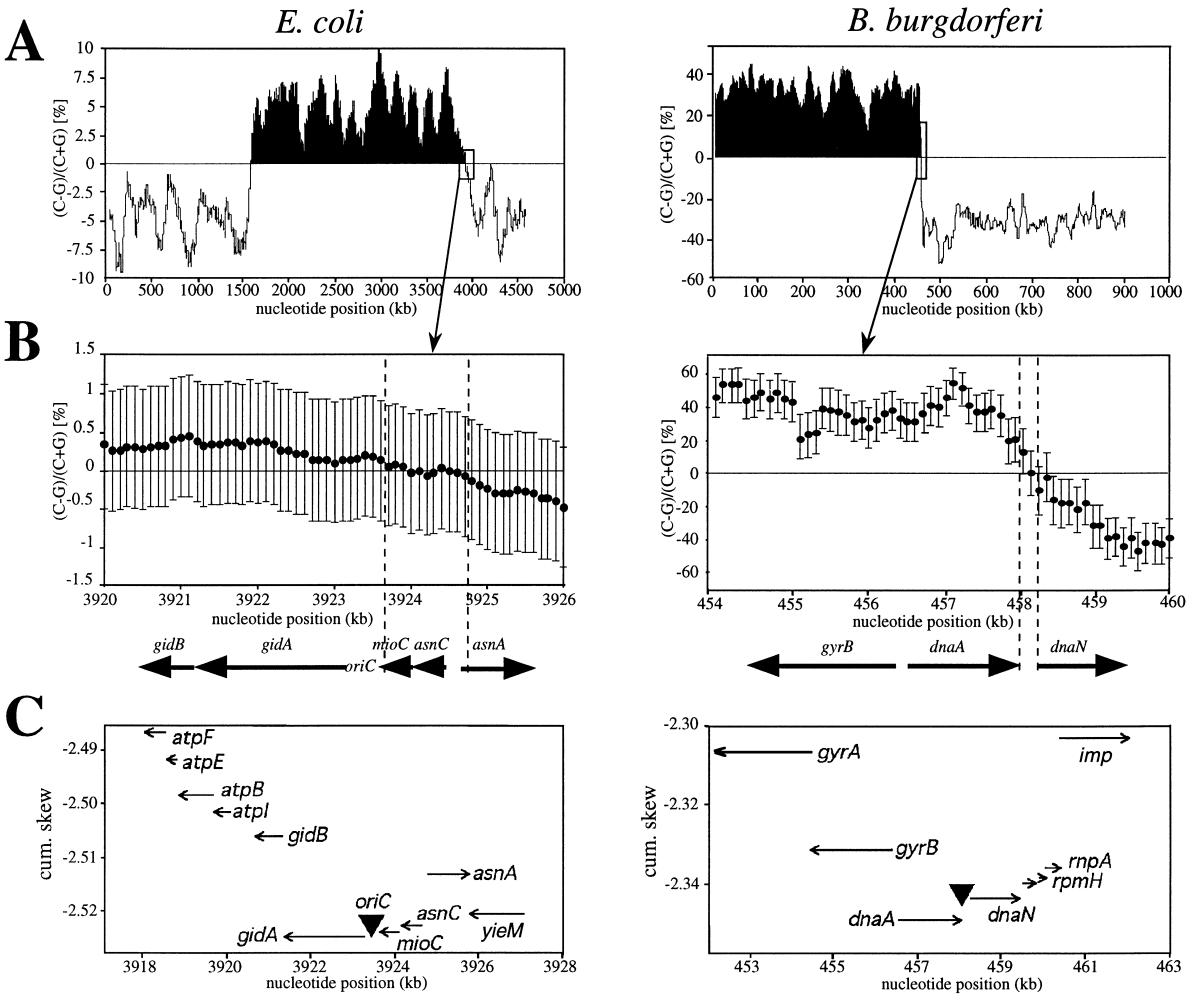

Fig. 1. Base composition skew diagrams calculated using the third position of each codon for the *E. coli* and the *B. burgdorferi* chromosomes.
A. CG skew analysis of entire chromosome. *E. coli*, 100 kb window size, 1 kb step; *B. burgdorferi*, 10 kb window size, 0.1 kb step.
B. CG skew analysis of 6 kb region containing the polarity switch, which is demarcated by dashed lines. *E. coli*, 100 kb window size, 0.1 kb step; *B. burgdorferi*, 1 kb window size, 0.1 kb step. Standard deviations were calculated according to Lobry (1996a).
C. Cumulative skew values of individual genes in the region of the CG skew polarity switch.

chromosome was calculated, the minimum value occurred in genes that flank *oriC* (Fig. 1C).

A plot of the CG skew of the 911 kb chromosome of *B. burgdorferi* revealed a single clear switch in polarity from positive to negative at the middle of the chromosome (Fraser *et al*., 1997). Again, the CG skew plot fluctuated least and the polarity switch was greatest when only the third position of the codons was examined (Fig. 1A). A finer analysis of this central region showed that the CG switch occurs between the genes *dnaA* and *dnaN* at position 458 kb, almost exactly at the midpoint of the chromosome (Fig. 1B). A switch from positive to negative values in AT skew occurs in the same region; this switch was also more pronounced when only the third position of codons was used in the calculation (data not shown). The CG skew polarity switch is both more acute and of greater amplitude for the *B. burgdorferi* chromosome, from a mean value of + 30% to − 30%, than for *E. coli*, from + 5% to − 5% (Fig. 1A). The cumulative skew diagram of the complete series of *B. burgdorferi* gene sequences showed a minimum between *dnaA* and *dnaN*, analogous to the position of *oriC* in the identical analysis of *E. coli* (Fig. 1C).

Mapping origins of replication by nascent DNA strand length

The PCR-based method (Vassilev and Johnson, 1989) for mapping chromosomal origins of replication is summarized in Fig. 2. In unsynchronized exponential-phase bacterial cultures, cells that have just initiated DNA replication will have very short nascent strands, whereas those nearing termination will have nearly chromosomal-length nascent strands. Therefore, assuming bidirectional replication from a unique origin, nascent strands in the total population will have a pyramid-shaped size distribution, with the apex centred on the replication origin (Fig. 2). Brief incubation with the nucleotide BrdU adds a short segment of labelled DNA to the ends of the strands, and strands that initiate during the pulse will be completely labelled. Single-stranded, BrdU-labelled nascent DNA is then size fractionated by alkaline gel electrophoresis of undigested genomic DNA. Small nascent DNA strands of different size classes are gel-purified from separate fractions and subjected to two rounds of immunoprecipitation with anti-BrdU antibody to further reduce non-BrdU-incorporated DNA (contaminating parental DNA) to insignificant levels (Vassilev and Johnson, 1990; Vassilev *et al*., 1990). DNA isolated from the gel-fraction that contains the shortest nascent strands should, thus, be greatly enriched for origin sequences, and each successive size fraction should contain nascent strands incorporating the origin and incrementally greater amounts of flanking DNA. PCR using primers for an origin-proximal sequence will therefore yield a product from all size fractions, whereas primers for origin-distal sequences will yield products only from fractions containing longer nascent strands. To visualize this, replicate dot blots of products amplified from each fraction in multiplex PCRs containing all primer sets are hybridized with individual origin-proximal and -distal probes.

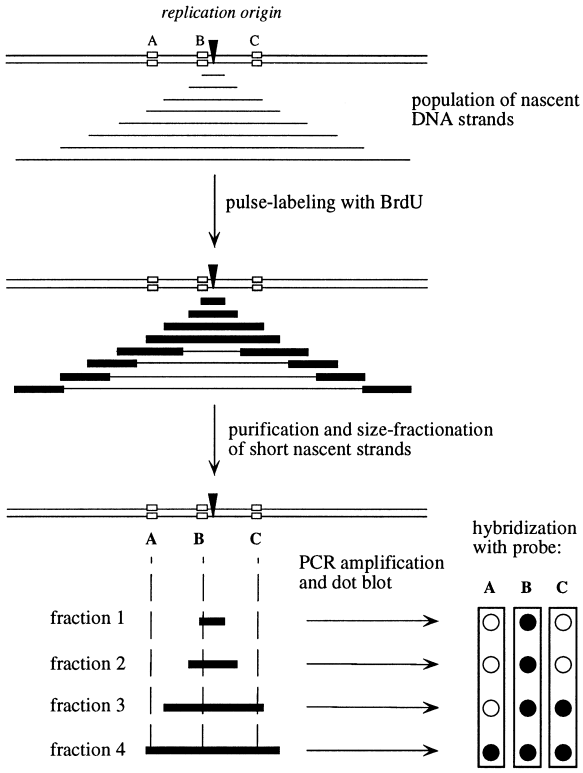

Fig. 2. Principle of origin mapping by nascent DNA strand length analysis. After pulse-labelling exponential phase cells with BrdU, short nascent strands are size fractionated by alkaline gel electrophoresis and purified by immunoprecipitation with anti-BrdU antibody. Multiplex PCR using primer sets for A–C in a single reaction is performed on each fraction, and amplified products are dot blotted to replicate membranes for individual hybridization with probes A–C. From the hybridization pattern, it can be deduced that segment B is closest to the replication origin because it could be PCR amplified from even the shortest nascent strands.

Verification of the method using E. coli oriC

To test the validity of this method and to standardize the conditions, we used the *E. coli* chromosome as a model system. Three PCR amplification segments spanning about 7 kb of the *oriC* locus were selected (probes A–C; Fig. 3A), and two other segments were selected in a region more than 800 kb from *oriC* (probes X and Z; Fig. 3A). The primers used to amplify these segments were the same length and of similar GC content, and each pair generated a single amplified product of 337–363 bp (Table 1). Each of the five products hybridized to a single fragment in

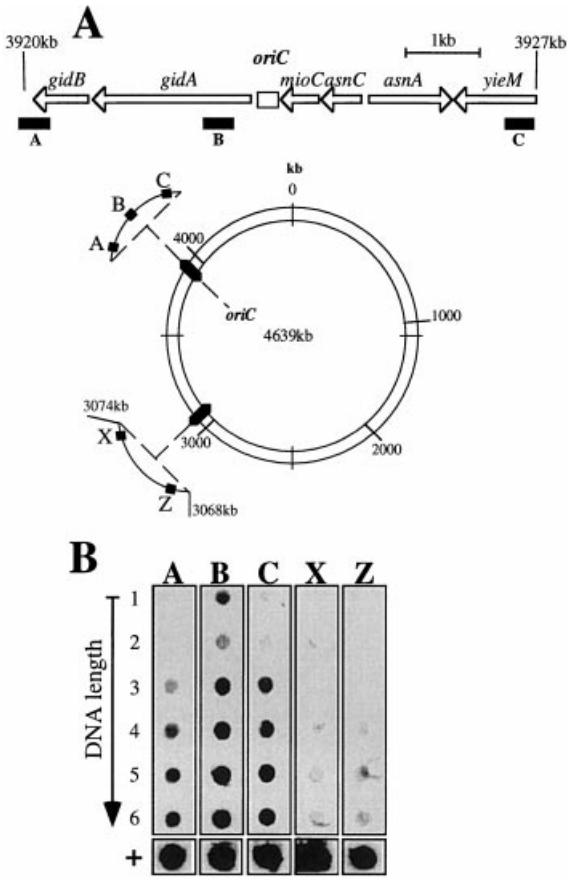

Fig. 3. Mapping of the replication origin of the *E. coli* chromosome by nascent strand analysis.
A. *oriC* region and chromosome map showing location of probes (solid boxes). Arrows represent individual genes in the *oriC* region and the open box indicates the 245 bp *oriC*.
B. Replicate dot blots of PCR products from six nascent DNA fractions hybridized with probes A–C, X and Z. Lower panel (+), control blots of PCR products from *E. coli* genomic DNA.

Southern blots of digested genomic DNA. Specificity was also checked in cross-hybridization experiments, which showed that each probe hybridized only with its homologue (data not shown). To avoid effects of sample-to-sample variation during PCR, amplification of each nascent DNA fraction was performed by addition of all primer pairs in a single tube. This multiplex PCR was first evaluated to verify that each primer set amplified its target sequence with approximately equal efficiency (see *Experimental procedures*).

Size-fractionated nascent DNA strands (0.3–10 kb) were used as templates to amplify the segments A–C, X and Z. The pattern obtained from dot blots of these PCRs individually hybridized with probes A–C, X and Z (Fig. 3B) showed that only fractions 1 and 2, corresponding to the shortest nascent DNA strands, contained sequences of segment B, closest to *oriC*. Only longer strands in fractions 3–6 incorporated sequences of segments A and C, located 2–3 kb on either side of *oriC*. Segments X and Z, located more than 800 kb from *oriC*, were not amplified from any of the fractions. A control PCR using total genomic DNA as template was analysed identically. As expected, all five probe segments were amplified from this control sample (Fig. 3B). These results were confirmed in two separate experiments in which nascent DNA fractions were analysed using primer sets for either segments A–C or segments X and Z (data not shown). In conclusion, this method proved capable of correctly mapping the known replication origin of *E. coli*.

Mapping of B. burgdorferi *origin of replication*

To map the replication origin of the *B. burgdorferi* chromosome, three segments in the *dnaA* region where the switch of the CG skew occurs were selected (probes A–C; Fig. 4A). Two additional segments 60 kb and 205 kb from the candidate origin of replication served as controls (probes X and Z; Fig. 4A). Preliminary testing of primer and probe

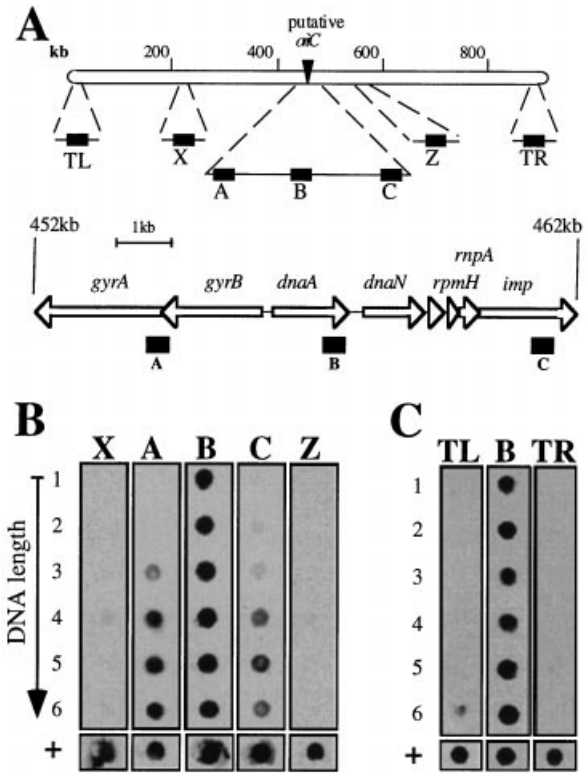

Fig. 4. Mapping of the replication origin of the *B. burgdorferi* chromosome.
A. Linear chromosome map and close-up of candidate origin region showing location of the probes used for mapping. Genes and their direction of transcription are indicated by arrows.
B. Replicate dot blots of PCR products from six nascent DNA fractions hybridized with probes X, A–C and Z.
C. Results using telomeric probes. Lower panels (+), control blots of PCR products from genomic DNA of *B. burgdorferi*.

Table 1. Primers used in this study.

Probes	Length(bp)	Primers	Location (bp)*
For *E. coli*			
Probe X	337	5′-AGGCAATTTTGCTTTTGGCG-3′	3 068 117–3 068 136
		5′-CTGCTCAGGAAATCAAAGAC-3′	3 068 434–3 068 453
Probe Z	340	5′-TTGACCATTCCGCTAAAACG-3′	3 072 758–3 072 777
		5′-GCTGTTTGTCGATAACGATG-3′	3 073 077–3 073 096
Probe A	357	5′-AATGATGGTTCTTAGCGCCG-3′	3 920 485–3 920 504
		5′-TGAGCAAGGCCGTTTCTACG-3′	3 920 822–3 920 841
Probe B	350	5′-GTCCCGGTTTTCAGACGACC-3′	3 922 660–3 922 679
		5′-AACGCAAGCAAAGGACCGGC-3′	3 922 990–3 923 009
Probe C	363	5′-TGAGCCGGAAGTATCCACAC-3′	3 926 239–3 926 258
		5′-GGAACTGAAACGCCTGGCAG-3′	3 926 583–3 926 602
For *B. burgdorferi*			
Probe TL	328	5′-TACTTCTCGATGGGAATTCC-3′	179–198
		5′-TAGTTCTCAGAGTTAAATCG-3′	488–507
Probe X	362	5′-TCTTGAGAAAAATGCACCCC-3′	248 784–248 803
		5′-GTATAAGAACTCCCGAAAGC-3′	249 126–249 145
Probe A	348	5′-AACGAAGTCCCATCTCATAC-3′	454 292–454 311
		5′-ACTATGGATCCTGCTAGAAG-3′	454 620–454 639
Probe B	336	5′-AACGTAAGAGACCTTGAAGC-3′	457 602–457 621
		5′-TGTTGAATGGGTTTTCCCTC-3′	457 918–457 937
Probe C	331	5′-CCTGTAATTTTGCAGCTTCC-3′	461 637–461 656
		5′-ATATTAAAAGCCCTGATGGC-3′	461 948–461 967
Probe Z	344	5′-TTTGGAGAGATTAAGGCAGG-3′	520 219–520 238
		5′-CTCAAGTCCATCATATTGAC-3′	520 544–520 563
Probe TR	369	5′-AGTTTAGGATAGTGCCCAAG-3′	909 711–909 730
		5′-ATATAGGGGCTTTATACGCC-3′	910 061–910 080

*Databases used in primer location were *http://web.pasteur.fr/Bio/Colibri.html/* for the *E. coli* chromosome and *http://www.tigr.org/* for the *B. burgdorferi* chromosome.

specificity was performed as described for *E. coli*. Application of the method in a single experiment utilizing primer sets for X, A–C and Z together resulted in the pattern shown in Fig. 4B. Probe B hybridized with the dot-blotted PCR products from all fractions, including those derived from the shortest nascent strands in fractions 1 and 2. In contrast, probes A and C, each ≈3 kb from the candidate origin locus, hybridized only with amplified products derived from longer nascent strands in fractions 3–6 (Fig. 4B). Probes X and Z did not hybridize with any of the fractions. A separate experiment using only primer sets A–C yielded the same pattern (data not shown).

Because initiation at telomeric sites has been proposed for viral and mitochondrial genomes that have a structure like the *Borrelia* linear chromosome (Baroudy *et al*., 1982; Nosek *et al*., 1998), probes TL and TR, corresponding to the two ends of the chromosome, were used in a third experiment. Telomeric sequences were not amplified from any of the nascent strand fractions, indicating that initiation does not occur at the telomeres (Fig. 4C). In summary, the results indicate that the origin of replication of the *B. burgdorferi* chromosome must be within the 6 kb region between segments A and C, close to segment B (Fig. 4). Because the nascent strands in the first fraction were < 1 kb, the origin should be within ± 1 kb from segment B. The hybridization patterns of Fig. 4 also indicate that replication is bidirectional, like that of *E. coli*, because unidirectional replication would result in an asymmetric pattern.

Sequence analysis of B. burgdorferi *origin region*

We analysed the *B. burgdorferi* nucleotide sequence in the region around segment B (Fig. 4A) for sequence patterns that have been described for known bacterial origins (Smith *et al*., 1991). Because the bacterial chromosomal origins that have been genetically delineated occur in intergenic regions, we gave special attention to the 240 bp region between *dnaA* and *dnaN*, where the abrupt polarity switch in CG skew occurs, and the 176 bp region between *gyrB* and *dnaA* (Fig. 4A). Neither intergenic region contained the 9 bp canonical binding site of the initiation protein DnaA, 5′-TTAT(C/A)CA(C/A)A-3′ (Fuller *et al*., 1984). A search using a more relaxed consensus sequence, 5′-T(T/C)(A/T)T(A/C)CA(C/A)A-3′ (Fuller *et al*., 1984; Moriya *et al*., 1988), allowing one mismatch revealed one match within the *gyrB–dnaA* intergenic sequence and two matches within the *dnaA–dnaN* intergenic sequence (Fig. 5). The *dnaA–dnaN* intergenic sequence also contained a series of four other 9 bp imperfect direct repeats, 5′-(A/T)A(A/C)(A/C)TACAA-3′, three 11–17 bp runs of A and T (Fig. 5A), and a 19.6% GC content compared with 28.6% for the entire *B. burgdorferi* chromosome (Fraser *et al*.,

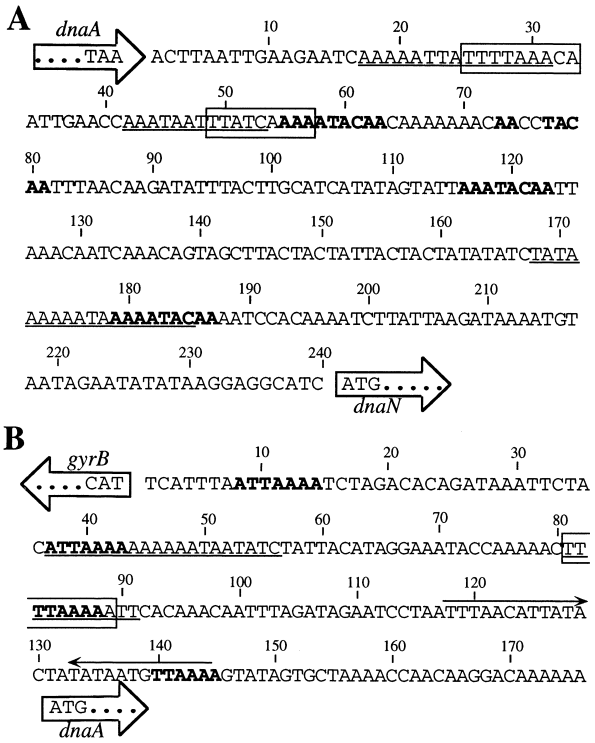

Fig. 5. Nucleotide sequence of intergenic sequences near the origin of the *B. burgdorferi* chromosome.
A. Region of 240 bp between *dnaA* and *dnaN*.
B. Region of 176 bp between *gyrB* and *dnaA*. Open arrows depict *dnaA*, *dnaN* and *gyrB* sequences. Bold letters indicate direct repeats. AT-rich sites are underlined and possible DnaA-binding sites are boxed. Thin arrows indicate inverted repeats.

1997). The 176 bp sequence between *gyrB* and *dnaA* was also AT rich (19.9% GC) and included runs of consecutive A and T but only short direct repeats and a 12 bp inverted repeat, which could form a stem loop structure (Fig. 5B). The 492 bp sequence between *dnaN* and *rpmH*, 300 bp of which consists of a putative gene of unknown function (Fig. 4A), did not contain repeats or recognizable DnaA boxes. These results indicate that the origin sequence of *Borrelia* does not resemble previously characterized origins of Gram-negative and Gram-positive bacteria.

Discussion

The origin-mapping results for the *B. burgdorferi* linear chromosome provide direct experimental support for the use of strand compositional asymmetry analysis to predict the location of the origin of replication of bacterial genomes. Based on CG skew, each strand of the chromosome of *E. coli* and *B. burgdorferi* can be divided into two halves, from the middle of the chromosome to the terminus (Fig. 1A). The same bipolar asymmetry is observed in most of the bacterial genomes analysed to date (Lobry, 1996a, 1996b; Freeman *et al*., 1998; Grigoriev, 1998; Mrazek and Karlin, 1998). Thus, for circular chromosomes such as *E. coli*, two polarity switches of CG skew are observed, one at the origin and the other at the terminus of replication. A single switch is evident at the centre of the *B. burgdorferi* chromosome, which our results confirm is the origin. The CG skew asymmetry in *B. burgdorferi* is more pronounced than in *E. coli* and, in fact, is the major source of variation in codon use between genes on the leading strand and those on the lagging strand (McInerney, 1998). Previous analysis of the *B. burgdorferi* chromosome also identified certain oligomers, such as the octamer TTGTTTTT, that are distributed differently on the leading strand compared with the lagging strand (Fraser *et al*., 1997; Salzberg *et al*., 1998).

Long-range strand compositional asymmetry may reflect differences with respect to replication and transcription of the leading and lagging strands (Francino and Ochman, 1997; Mrazek and Karlin, 1998). For example, during replication the leading strand is continuously synthesized whereas the lagging strand is synthesized discontinuously in short Okazaki fragments. In addition, most bacterial genes (66% in the case of *B. burgdorferi*) are encoded on the leading strand (Francino and Ochman, 1997; Fraser *et al*., 1997), making it the primary transcription template strand. The lagging strand is more often non-duplexed during both these processes and more susceptible to DNA-damaging events (Francino and Ochman, 1997).

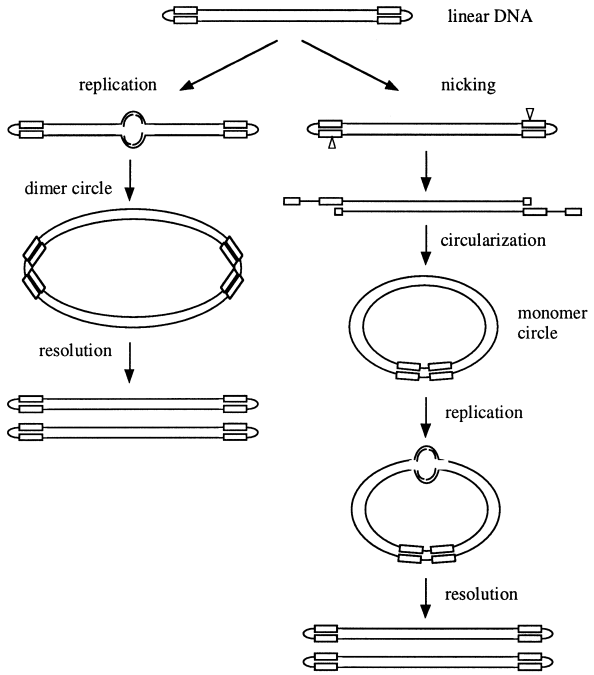

Fig. 6. Replication models for the *Borrelia* linear chromosome with bidirectional replication from a central origin. Boxes indicate the terminal inverted repeats; hairpin telomeres are represented at the ends of the chromosome.

Nascent DNA strand analysis mapped the replication initiation site of the *B. burgdorferi* chromosome to an ≈2 kb region containing *dnaA*, *dnaN* and *gyrB*, which are also adjacent to the origin in other prokaryotes (Smith *et al*., 1991). Additional observations taken together suggest that the *B. burgdorferi* replication origin is most probably in the 240 bp sequence between *dnaA* and *dnaN*. First, the abrupt CG skew switch occurs precisely there. Second, the bacterial chromosomal origins that have been functionally delineated to date are intergenic sequences and the smallest of these is the 245 bp minimal *oriC* of *E. coli* (Oka *et al*., 1980; Smith *et al*., 1991). Of the intergenic regions near the *B. burgdorferi* origin, the 240 bp *dnaA–dnaN* intergenic region most closely approximates to this size. It should be noted, however, that spirochetes are phylogenetically distant from those bacteria whose chromosomal origins have been defined. In fact, we did not find sequence patterns that typify the chromosomal origins of Gram-positive and Gram-negative bacteria (Smith *et al*., 1991) at the *B. burgdorferi* origin. Notwithstanding this, the prototypical bacterial *oriC* must accommodate two fundamental steps of initiation: binding of the initiation protein DnaA to origin-specific sites (DnaA boxes), followed by local unwinding of adjacent AT-rich sites to allow entry of the proteins that form the replication complex (Kornberg and Baker, 1992). For example, the *E. coli oriC* contains five 9 bp DnaA boxes, and DnaA boxes have been found at other bacterial origins (Smith *et al*., 1991). We and others (Old *et al*., 1993) detected no perfect matches to consensus DnaA boxes in the origin region, suggesting that the sequence of the *B. burgdorferi* DnaA-binding site is atypical. Indeed, the *B. burgdorferi* DnaA homologue differs in the normally highly conserved region that may be involved in DNA binding (Old *et al*., 1993). We did not detect DnaA boxes in the putative replication origin of another pathogenic spirochete *Treponema pallidum*, for which the complete sequence of the genome has been determined (Fraser *et al*., 1998). Although the *B. burgdorferi* origin locus did not contain an obvious eubacterial *oriC* sequence pattern, the *dnaA–dnaN* intergenic region did contain a complex pattern. This included a series of four 9 bp imperfect direct repeats, two sequences with single bp mismatches to DnaA-binding sites, and runs of consecutive AT that could facilitate strand separation.

Our data address the various models that have been proposed for the replication of linear elements with covalently closed hairpin ends. The physical mapping results indicate that replication of the *B. burgdorferi* chromosome initiates from an internal origin that is equidistant from the telomeres, and that replication from this origin is bidirectional and symmetric. Bidirectional replication can also be implied from the CG skew profile: viral and mitochondrial genomes that replicate other than by a θ-type bidirectional mechanism do not show strand composition bipolar asymmetry (Grigoriev, 1998). Our results rule out replication models involving telomeric origins, such as the self-priming and *de novo* start models proposed for the structurally similar genomes of poxviruses and certain mitochondria (Baroudy *et al*., 1982; Nosek *et al*., 1998). Rather, the mechanism of replication of the *Borrelia* linear elements probably involves circular intermediates (Fig. 6). This has been suggested by a previous study in which a stably replicating circular isoform of a *Borrelia hermsii* linear plasmid was detected (Ferdows *et al*., 1996). Because all other spirochetes analysed so far have circular chromosomes, it can be speculated that the linear structure of the *Borrelia* chromosome evolved from a circular progenitor and that a typical mechanism for initiation of replication has been conserved. The mechanisms for termination of replication and resolution of the telomeres remain important unanswered questions.

Future biochemical and genetic analyses should further define what constitutes a minimal origin of replication for the *B. burgdorferi* linear chromosome and its initiation mechanism. Nascent DNA strand analysis could also be used to map the replication origins of *Borrelia* plasmids. These studies should clarify whether the linear plasmids are actually minichromosomes, as has been proposed (Barbour, 1993), or if their replication origins are each distinct as is the case for compatible plasmids in other bacteria. These analyses should also aid in the development of new genetic tools such as shuttle vectors for gene delivery.

Experimental procedures

Labelling and isolation of short nascent DNA strands

Exponential-phase cultures of *E. coli* DH5α grown at 37°C in M63 medium and *B. burgdorferi* B31 grown at 34°C in BSK medium were pulse-labelled with 5-bromo-2′-deoxyuridine (BrdU) (Sigma) at a final concentration of 20 μM for 1 min and 5 min respectively. Immediately after pulse-labelling, replication was stopped by centrifugation at 4°C, brief washing and cell lysis. Genomic DNA was isolated by a previously described method (Meier *et al*., 1985). Size fractionation of DNA was performed under denaturing conditions by alkaline gel electrophoresis (Staib and Grummt, 1997). Approximately 2 μg of genomic DNA per well was electrophoresed in 1.2% agarose in 50 mM NaOH/1 mM EDTA at 0.6 V cm^{-1} for 32–48 h. The gel was then neutralized and stained in 0.5 M Tris, pH 7.5, containing 0.5 μg ml^{-1} ethidium bromide. For each lane, six agarose slices that contained nascent DNA of different size classes in the range of 0.3–10 kb were removed. DNA in the gel slices was recovered by electroelution and was ethanol precipitated in the presence of 100 μg of herring sperm DNA. DNA from each fraction was subjected to two rounds of immunoprecipitation with mouse anti-BrdU monoclonal antibody (Calbiochem) and rabbit anti-mouse IgG (Sigma) as previously described (Staib and Grummt, 1997). BrdU-labelled DNA was recovered after phenol/chloroform treatment by ethanol precipitation in the presence of 20 μg of yeast tRNA.

PCR and hybridization assays

Multiplex PCRs contained 0.5 μM of each primer pair (Table 1) and individual nascent DNA fractions. Hot start PCR (Chou *et al*., 1992) was performed in a Perkin Elmer thermal cycler: 94°C for 2 min (one cycle); 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s (35 cycles); and a final 10 min at 72°C. Amplified products were purified by using Sephacryl S-300 columns (Pharmacia), denatured in 0.25 N NaOH/0.5 M NaCl, and dot blotted onto replicate nylon membranes (GeneScreen, NEN Research Products).

Probes were generated by PCR of total genomic DNA using primers for a single target segment and labelled with horseradish peroxidase by using the ECL labelling kit (Amersham). Blots were hybridized overnight at 42°C and treated as recommended by the manufacturer to detect hybridization.

For both *E. coli* and *B. burgdorferi*, the efficiency of primer sets A–C (Table 1) in the multiplex PCR was evaluated using a control template encompassing the entire origin region. Serial dilutions of the multiplex PCR were dot blotted and individually hybridized with probes corresponding to each target sequence. Hybridization signals were compared to assess the relative yield of each product in the multiplex PCR.

Sequence analysis

CG skew plots were generated using a previously described program (Lobry, 1996a) that compiled base frequencies within a moving window along the published strand of the *B. burgdorferi* genome (accession number AE000783) and the *E. coli* genome (accession number U00096) to calculate the skew index $(C-G)/(C+G)$, where C and G denote the corresponding base counts within the window. To increase the signal to noise ratio, only bases in the third position of the codons were taken into account. AT skew $(A-T)/(A+T)$ was calculated by the same algorithm to check the consistency of results. Combined CG and AT skew was calculated and cumulatively summed (Freeman *et al*., 1998; Grigoriev, 1998; Lobry, 1996b) for each gene along the entire chromosome instead of a series of fixed windows. Again, only third-codon positions were considered. Values for individual genes in the region of CG skew polarity switch, where the minimum cumulative value occurred, were graphed (Fig. 1C).

The *B. burgdorferi* replication origin locus sequence was analysed using the GCG software package (Genetics Computer Group).

Acknowledgements

We thank Patti Rosa, Isabelle Saint Girons, Tom Schwan and Kit Tilly for critical review of the manuscript. M.P. is a recipient of a fellowship from Fondation Roux (Institut Pasteur) and from Fondation Philippe.

References

Barbour, A.G. (1993) Linear DNA of *Borrelia* species and antigenic variation. *Trends Microbiol* **6:** 236–329.

Barbour, A.G., and Garon, C.F. (1987) Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends. *Science* **237:** 409–411.

Baroudy, B.M., Venkatesan, S., and Moss, B. (1982) Incompletely base-paired flip-flop terminal loops link the two DNA strands of the vaccinia virus genome into one uninterrupted polynucleotide chain. *Cell* **28:** 315–324.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., *et al*. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277:** 1453–1462.

Casjens, S., Murphy, M., DeLange, M., Sampson, L., van Vugt, R., and Huang, W.M. (1997) Telomeres of the linear chromosomes of Lyme disease spirochaetes: nucleotide sequence and possible exchange with linear plasmid telomeres. *Mol Microbiol* **26:** 581–596.

Chou, Q., Russel, M., Birch, D.E., Raymond, J., and Bloch, W. (1992) Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Res* **20:** 1717–1723.

Ferdows, M.S., Serwer, P., Griess, G.A., Norris, S.J., and Barbour, A.G. (1996) Conversion of a linear to a circular plasmid in the relapsing fever agent *Borrelia hermsii*. *J Bacteriol* **178:** 793–800.

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

Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., *et al*. (1997) Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390:** 580–586.

Fraser, C.M., Norris, S.J., Weinstock, G.M., White, O., Sutton, G.G., Dodson, R., *et al*. (1998) Complete genome sequence of *Treponema pallidum*, the syphillis spirochete. *Science* **281:** 375–388.

Freeman, J.M., Plasterer, T.N., Smith, T.F., and Mohr, S.C. (1998) Patterns of genome organization in bacteria. *Science* **279:** 1827.

Fuller, R.S., Funnel, B.E., and Kornberg, A. (1984) The DnaA protein complex with the *E. coli* chromosomal origin (*oriC*) and other sites. *Cell* **38:** 889–900.

Grigoriev, A. (1998) Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res* **26:** 2286–2290.

Hinnebusch, J., and Barbour, A.G. (1991) Linear plasmids of *Borrelia burgdorferi* have a telomeric structure and sequence similar to those of eucaryotic virus. *J Bacteriol* **173:** 7233–7239.

Hinnebusch, J., and Tilly, K. (1993) Linear plasmids and chromosomes in bacteria. *Mol Microbiol* **10:** 917–922.

Kornberg, A., and Baker, T.A. (1992) *DNA Replication*. New York: W.H. Freeman.

Lobry, J.R. (1995) Properties of a general model of DNA evolution under no-strand-bias conditions. *J Mol Evol* **40:** 326–330.

Lobry, J.R. (1996a) Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol* **13:** 660–665.

Lobry, J.R. (1996b) A simple vectorial presentation of DNA sequences for the detection of replication origins in bacteria. *Biochimie* **78:** 323–326.

McInerney, J.O. (1998) Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi*. *Proc Natl Acad Sci USA* **95:** 10698–10703.

Meier, J.T., Simon, M.I., and Barbour, A.G. (1985) Antigenic variation is associated with DNA rearrangements in a relapsing fever *Borrelia*. *Cell* **41:** 403–409.

Moriya, S., Fukuoka, T., Ogasawara, N., and Yoshikawa, H. (1988) Regulation of initiation of the chromosomal replication by DnaA-boxes in the origin of the *Bacillus subtilis* chromosome. *EMBO J* **7:** 2911–2917.

Mrazek, J., and Karlin, S. (1998) Strand compositional asymmetry in bacterial and large viral genomes. *Proc Natl Acad Sci USA* **95:** 3720–3725.

Nosek, J., Tomaska, L.U., Fukuhara, H., and Kovac, L. (1998) Linear mitochondrial genomes: 30 years down the line. *Trends Genet* **14:** 184–188.

Oka, A., Sugimoto, K., Takanami, M., and Hirota, Y. (1980) Replication origin of the *Escherichia coli* K-12 chromosome: the size and the structure of the minimum DNA segment carrying the information for autonomous replication. *Mol Gen Genet* **178:** 9–20.

Old, I.G., Margarita, D., and Saint Girons, I. (1993) Unique genetic arrangement in the *dnaA* region of the *Borrelia burgdorferi* linear chromosome: nucleotide sequence of the *dnaA* gene. *FEMS Microbiol Lett* **111:** 109–114.

Salzberg, S.L., Salzberg, A.J., Kerlavage, A.R., and Tomb, J.F. (1998) Skewed oligomers and origins of replication. *Gene* **217:** 57–67.

Smith, D.W., Yee, T.W., Baird, C., and Krihnapillai, V. (1991) Pseudomonad replication origins: a paradigm for bacterial origins? *Mol Microbiol* **5:** 2581–2587.

Staib, C., and Grummt, F. (1997) Mapping replication origins by nascent DNA strands length. *Methods Companion Methods Enzymol* **13:** 293–300.

Sueoka, N. (1995) Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. *J Mol Evol* **37:** 137–153.

Taira, T., Iguchi-Ariga, S.M.M., and Ariga, H. (1994) A novel DNA replication origin in the human heat shock protein 70 gene promoter. *Mol Cell Biol* **14:** 6386–6397.

Vassilev, L., and Johnson, E.M. (1989) Mapping initiation sites of DNA replication *in vivo* using polymerase chain reaction amplification of nascent strand segments. *Nucleic Acids Res* **19:** 7693–7705.

Vassilev, L., and Johnson, E.M. (1990) An initiation zone of chromosomal DNA replication located upstream of the c-*myc* gene in proliferating HeLa cells. *Mol Cell Biol* **10:** 4899–4904.

Vassilev, L.T., Burhans, W.C., and DePamphilis, M.L. (1990) Mapping an origin of DNA replicating at a single-copy locus in exponentially proliferating mammalian cells. *Mol Cell Biol* **10:** 4685–4689.