ELSEVIER

Physica A 273 (1999) 103–115

PHYSICA A

www.elsevier.com/locate/physa

Mechanisms generating long-range correlation in nucleotide composition of the *Borrelia burgdorferi* genome

P. Mackiewicz[a], A. Gierlik[a], M. Kowalczuk[a], D. Szczepanik[a], M.R. Dudek[b], S. Cebrat[a,*]

[a] *Department of Genetics, Institute of Microbiology, Wrocław University, ul. Przybyszewskiego 63/77, 51-148 Wrocław, Poland*

[b] *Institute of Theoretical Physics, Wrocław University, pl. Maxa Borna 9, 50-204 Wrocław, Poland*

Received 15 July 1999

Abstract

We have analysed protein coding and intergenic sequences in the *Borrelia burgdorferi* (the Lyme disease bacterium) genome using different kinds of DNA walks. Genes occupying the leading strand of DNA have significantly different nucleotide composition from genes occupying the lagging strand. Nucleotide compositional bias of the two DNA strands reflects the aminoacid composition of proteins. 96% of genes coding for ribosomal proteins lie on the leading DNA strand, which suggests that the positions of these as well as other genes are non-random. In the *B. burgdorferi* genome, the asymmetry in intergenic DNA sequences is lower than the asymmetry in the third positions in codons. All these characters of the *B. burgdorferi* genome suggest that both replication-associated mutational pressure and recombination mechanisms have established the specific structure of the genome and now any recombination leading to inversion of a gene in respect to the direction of replication is forbidden. This property of the genome allows us to assume that it is in a steady state, which enables us to fix some parameters for simulations of DNA evolution.

PACS: 87.14.G; 05.10-a

Keywords: Long-range correlation; Replication; Transcription

* Corresponding author. Fax: +48-71-3252151.

E-mail addresses: cebrat@angband.microb.uni.wroc.pl (S. Cebrat), mdudek@ift.uni.wroc.pl (M.R. Dudek)

0378-4371/99/$ - see front matter
PII: S0378-4371(99)00345-3

1. Introduction

1.1. Replication-associated mutational pressure

The bacterial chromosome is a double-stranded DNA molecule. Each DNA molecule is built of two antiparallel (Watson and Crick) strands (see Fig. 1a). Each of these strands is a directional structure with one end termed as 5′ and the other as 3′. Eubacterial chromosomes usually are circular and have only one origin of replication (where replication starts in both directions) and one termination region (where the two replication forks meet and replication ends) (Fig. 1a). The *B. burgdorferi* genome is rather exceptional – it is linear, and replication forks move until they reach the ends of the chromosome. DNA synthesis is semi-conservative, which means that after replication in the new DNA molecule one strand is the old one (matrix) and the other is the newly synthesised one. Synthesis of a new strand is possible only in one direction: from the 5′ end to the 3′ end. Since matrix strands are antiparallel to newly synthesised strands, the mechanisms of replication of the two strands have to be different. In fact one strand is synthesised continuously – it is called leading (Fig. 1a) and the other is synthesised

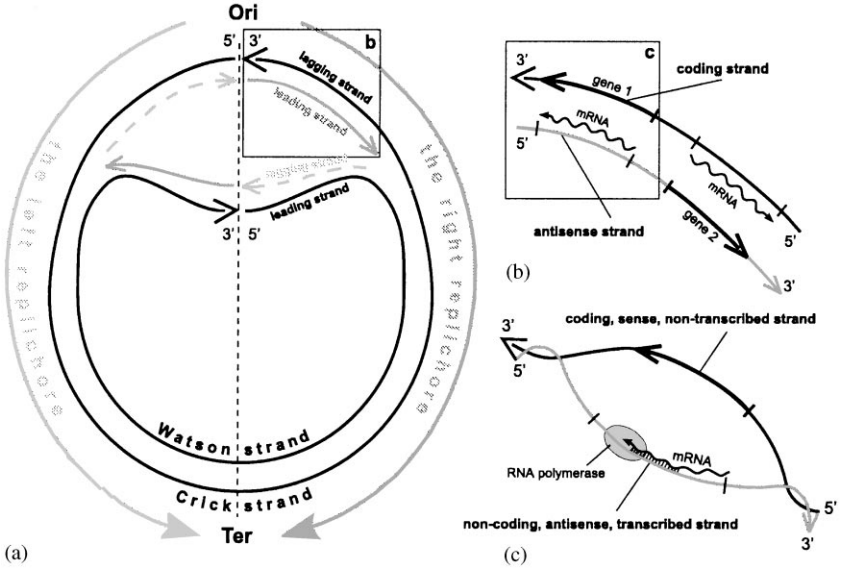

Fig. 1. The topology of replication and transcription of bacterial circular chromosome. Synthesis of a new strand is possible only in one direction: from the 5′ end to the 3′ end. One strand is synthesised continuously – it is called leading (a – solid line) and the other is synthesised discontinuously by joining fragments and it is called lagging (dotted line) The role of the leading/lagging strand is switched at the origin of replication (Ori) and at the terminus of replication (Ter). The two parts of the chromosome from the origin to the terminus of replication are called replichores, usually of approximately the same length. Gene 2 in (b) is located on the leading strand, because its coding strand is replicated as the leading strand and the direction of transcription (also from 5′ end to the 3′ end of the gene) is the same as the direction of replication fork movement.

discontinuously by joining fragments (dotted line in Fig. 1a) and it is called lagging [1,2]. In eubacterial circular genomes the role of the leading/lagging strand is switched at the terminus of replication. At this point the leading strand becomes lagging and vice versa. The same kind of topological switch is at the origin of replication. The part of the chromosome from the origin to the terminus of replication is called a replichore [3]. Usually the two replichores in a eubacterial genome are approximately of the same length. Different mechanisms of replication of the two DNA strands implicate different mutational pressures which introduce specific nucleotide substitutions into newly synthesised DNA strands ([4–12]; for a review see [13–15]). That is why a specific bias in DNA nucleotide composition between leading and lagging strands is observed in eubacterial chromosomes [3,14,16–29].

1.2. Transcription-associated mutational pressure

To understand the relations between gene location and the structure of chromosome, description of another process – gene transcription – should be superimposed on the replicating chromosome. A protein coding gene is a fragment of double-stranded DNA which codes for a single polypeptide chain. It is arbitrarily accepted that one strand of a gene is called the coding strand (sense strand or non-transcribed strand) and the other the non-coding, anti-sense or transcribed strand (Fig. 1c). A gene is defined as located on the leading strand if its coding strand is replicated as leading and the direction of transcription (also from 5′ end to the 3′ end of the gene) is the same as the direction of replication fork movement (gene 2 in Fig. 1b). Replication-associated mutational pressure is not the only mechanism introducing compositional bias into the DNA molecule. Asymmetric nucleotide substitutions could also occur during transcription of genes [13,22,30,31]. Since transcription-associated mutational pressure preferentially affects the non-transcribed strand of genes, its local effect observed on one DNA strand (i.e. leading) depends on the location of the gene on the chromosome, while replication-associated mutational pressure affects uniformly the whole replichore from the origin to the terminus of replication. It is relatively easy to separate the effect of replication from the effect of transcription by subtracting or adding the values of asymmetry of Watson (W) strand and the asymmetry of Crick (C) strand [26–28,32].

1.3. Asymmetry introduced by protein coding functions

Both replication- and transcription-associated mutational pressures introduce substitutions into DNA strands “not seeing” positions in codons. Nevertheless, protein coding sequences have their own asymmetry connected with coding functions. This asymmetry is forced by selection for specific amino acid composition of coded proteins, and has a very specific character – each position in codons shows significantly different nucleotide composition [14,33–36] and differently responds to mutational pressure introducing nucleotide substitutions. All these mechanisms are responsible for establishing specific “coding asymmetry” of bacterial genomes [28,37–40]. Let us assume

that replication-associated mutational pressure results in purine-rich leading strand and that the coding strands of genes are also rich in purines. Thus, if a gene is located on the leading strand (according to the convention accepted above), both trends – that connected with coding functions and that implicated by mutational pressure – cumulate. But when a gene with its purine-rich coding strand is located in inverted position, then its transcribed pyrimidine-rich strand is located on the leading, purine-rich strand. It is obvious that under the same replication-associated mutational pressure, the frequencies of mutations in these two genes will be different. Since some mutations are accepted and accumulate, one can expect different nucleotide composition of genes located on leading and lagging DNA strands. That should be true only in the case of relatively rare recombination events, which means that the coding sequence stays relatively long in the same position, cumulating non-random substitutions. This phenomenon is very clearly seen in the *B. burgdorferi* genome.

2. Materials and methods

All calculations were done on the *B. burgdorferi* genomic sequence [41] downloaded from www.ncbi.nlm.nih.gov.

2.1. DNA walks

To show correlations in nucleotide composition of the *B. burgdorferi* genome, we have performed different kinds of DNA walks. These methods were described in detail previously [32] and many technical issues are presented at our www site: http://smORFland.microb.uni.wroc.pl. We have analysed distribution of individual nucleotides along the DNA strand by performing detrended DNA walks. In this kind of walks the walker moved one unit up in two-dimensional space when it visited the analysed nucleotide, and down when it visited a different nucleotide. The value of its movement down was corrected in such a way that the walk ended at $y = 0$. Values on the x-axis represent position on chromosome. This kind of detrended walks was performed also when other compositional biases of the genome were analysed. For the analysis of protein coding sequences we have used a specific modification of two-dimensional DNA walks of Berthelsen et al. [42]. We have analysed separately the composition of first, second and third positions in codons and we have presented the results as three separate DNA walks on one plot, called a spider. Analyses were done separately for each Open Reading Frame (ORF) or for all ORFs found in the genome spliced together. In the latter case the resulting plot was called a genomic spider. We have also performed DNA walks separately for spliced ORFs lying on leading or lagging strands, or on Watson or Crick strands. The last DNA walks enable showing the differences in nucleotide composition of codon positions between ORFs situated on leading and lagging strands. DNA walks performed for spliced ORFs show average trends in nucleotide composition of particular positions in codons, while DNA walks

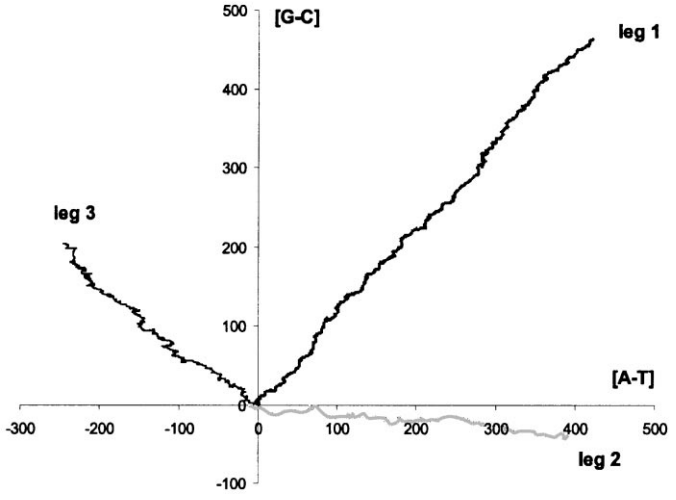

Fig. 2. DNA walks performed in two-dimensional space separately for each position in codons for gene BB0512, 6500 bp long, with co-ordinates on chromosome: 172 165–174 331 bp, situated on the leading strand. Three DNA walks were done independently for each nucleotide position in codons. The first walker starts from the first nucleotide position of the first codon and jumps every third nucleotide until the end of the examined sequence has been reached. Similarly, the second and the third walkers start from the second and the third nucleotide positions of the first codon, respectively. The three walks together have been called a spider and a single walk has been called a spider leg. Every jump of a walker is associated with a unit shift in the two-dimensional space depending on the type of nucleotide visited. The shifts are: (0,1) for G, (1,0) for A, $(0,-1)$ for C and $(-1,0)$ for T.

done separately for each ORF enable parameterisation of individual coding sequences. We have used parameters describing individual ORFs (i.e. their spiders). They were: arcus tangent [(G–C)/(A–T)] for the first, second, and third positions in codons. Using pairs of these parameters it was possible to prepare distributions of ORFs on the finite surface of torus projection.

3. Results and discussion

3.1. Correlations in nucleotide composition of coding sequences

Examples of DNA walks performed in two-dimensional space separately for each position in codons are presented in Fig. 2. Gene BB0512 described by these walks, 6500 bp long, is situated on the leading strand, and its co-ordinates are 172 165–174 331 bp. Its function has not been found yet. Each walk starts at the start translation codon of the analysed gene. The walk representing the first positions in codons is called leg 1. Leg 2 and leg 3 represent DNA walks performed for the second and the third positions in codons, respectively. Trends in nucleotide composition of each position are significantly different. One could argue that correlations in nucleotide composition of

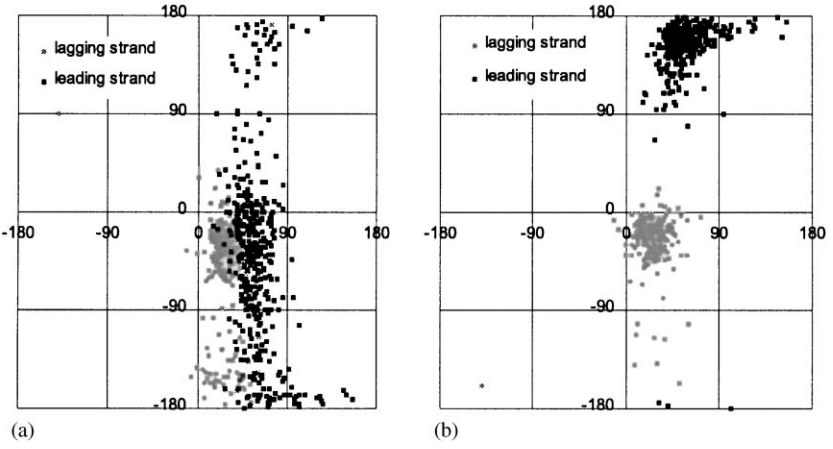

Fig. 3. Distribution of ORFs longer than 100 codons on the torus projection. Each ORF is represented by a point with co-ordinates: (a) arcus tangent [(G–C)/(A–T)] for the first position against arcus tangent [(G–C)/(A–T)] for the second position; (b) arcus tangent [(G–C)/(A–T)] for the first position against arcus tangent [(G–C)/(A–T)] for the third position.

codon positions have to be very specific for individual genes. Thus, we have performed this analysis for every ORF longer than 100 codons in the *B. burgdorferi* genome. In case of overlapping ORFs the longer one was chosen. To describe the obtained set of ORFs, we have plotted arcus tangent [(G–C)/(A–T)] for the first positions against arcus tangent [(G–C)/(A–T)] for the second positions, and against arcus tangent [(G–C)/(A–T)] for the third positions. Note that arcus tangent [(G–C)/(A–T)] corresponds to the angle between the x-axis and the leg describing the analysed position. The resulting plots are distributions of points representing individual ORFs on torus projections (Fig. 3a and b). We have two distinct sets of ORFs forming two compact groups of points on the plot, especially when the first and the third positions in codons are taken into consideration.

3.1.1. The effect of gene position on its nucleotide composition

The two sets of ORFs observed in plot 3b correspond to sets of ORFs situated on the leading DNA strand (upper right part of the plot) and lagging strand (close to the centre). This means that codon composition of *B. burgdorferi* genes depends strongly on the direction of their transcription in relation to the direction of movement of replication fork. The most sensitive to location on chromosome are the third positions in codons (y-axis in Fig. 3b). It is very clearly seen in DNA walks performed for ORFs spliced separately from leading and lagging strands (Fig. 4). In two sets of ORFs very strong correlations are seen in the first and the third positions. However, while in the first positions purines prevail in both sets, the third positions of ORFs situated on the lagging strand are relatively rich in adenine and cytosine and the third positions of ORFs on the leading strand are rich in guanine and thymine. There is

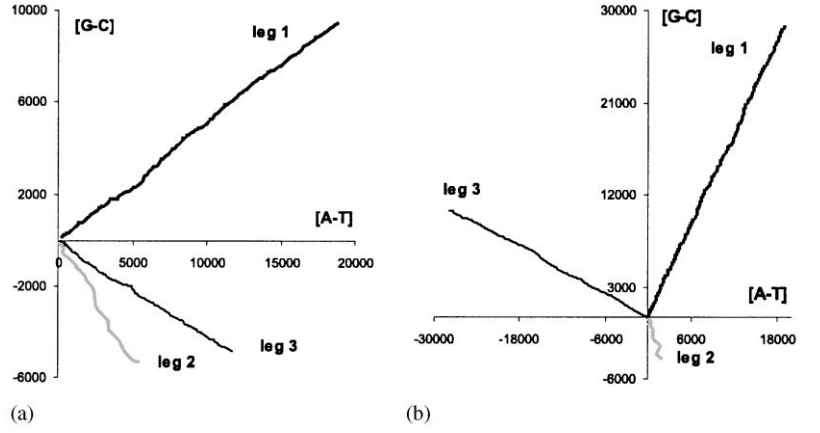

Fig. 4. Spiders representing spliced ORFs longer than 100 codons; (a) ORFs located on the lagging strand, (b) ORFs located on the leading strand.

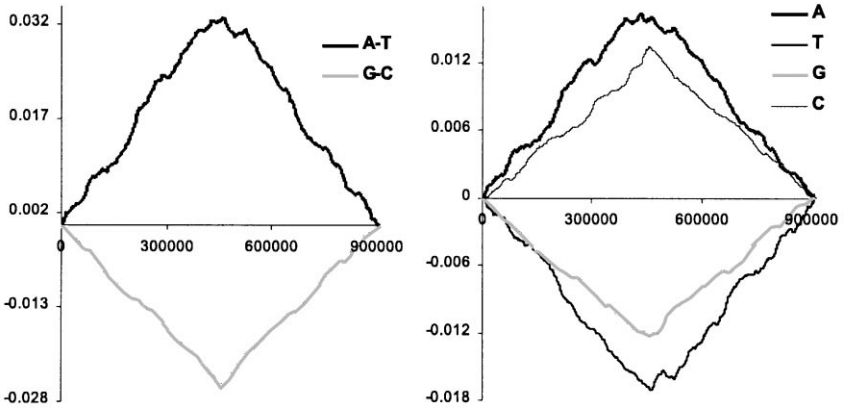

Fig. 5. Detrended DNA walks on whole *B. burgdorferi* chromosome, strand W; (a) walker moved up when the visited nucleotide was a purine and down when it was a pyrimidine, analyses were done separately for the pairs A–T and G–C. (b) Analysis of relative abundance of each of four nucleotides along the chromosome.

a very weak correlation in nucleotide composition of the second position in codons of ORFs situated on the leading strand. That is why the set of these ORFs forms a belt in Fig. 3a. Much better, comparative results were obtained in detrended DNA walks performed to analyse relative abundance of particular nucleotides along the whole DNA strand (Fig. 5). The plots are normalised in such a way that the values on y-axes, representing relative abundance of nucleotide, can be compared. Values on x-axes represent positions on chromosome. The *B. burgdorferi* genome is linear and the origin of replication is situated in the middle of the chromosome. Thus, the extrema of the plots represent the point where the leading/lagging role of DNA strand switches, in fact it is the origin of replication. Some authors argue that the correlation seen in codon

composition of ORFs is a result of substitutions introduced during transcription. If it is true, the substitutions introduced during transcription should introduce the same trends in DNA composition independent of the strand (leading versus lagging). Thus, when the asymmetry introduced to ORFs of the leading strand is added to the asymmetry of ORFs of the lagging strand, it should cumulate. If replication-associated processes are responsible for the compositional asymmetry of ORFs, it should be of different sign for ORFs on leading and lagging strands. This asymmetry will cumulate when the asymmetry of the leading strand is subtracted from the asymmetry of the lagging strand. We have performed additions and subtractions of DNA walks on spliced ORFs situated on W strand and C strand. The results of such DNA walk transformations are presented in Fig. 6. It is obvious that replication-associated mechanisms are responsible for the asymmetry of coding sequences in the *B. burgdorferi* genome. The asymmetry introduced by transcription-associated mechanisms is invisible in this genome.

3.2. DNA asymmetry introduced by replication-associated mutational pressure into coding and intergenic sequences

For further analysis of the *B. burgdorferi* genome, we subtracted DNA walks performed separately for each position in codons. The results were compared with the asymmetry in intergenic sequences (Fig. 7). The results indicate that replication-associated mutational pressure influences composition of intergenic sequences as well as each position of codons. However, the results of this influence are different for different positions in codons. Since replication does not "see" positions in codons, the differences between positions in accumulation of substitutions must be a result of selection. If we assume that the differences observed between the ORFs of leading and lagging strands are the result of replication-associated mutational pressure, we have to assume that there should be enough time to accumulate many mutations in coding sequences, which means that positions of genes in the genome are conserved and recombination with inversion is forbidden in this genome or it is an extremely rare event. There is a question implicated by such structure of the genome. Were the positions of genes on the chromosome pre-established in the past randomly or not? We have analysed the positions of genes coding for ribosomal proteins and found that 96% of these genes are coded by the leading strand. Thus, we suppose that positions of other genes are also non-random and that they were established non-randomly. From the biological point of view, one should expect that genes are positioned on chromosome in such a way that the probability of mutation is the lowest. Then the killing effect of mutational pressure is the lowest. Since the probability of substitution under a given replication-associated mutational pressure for the same sequence depends on the direction of this sequence in relation to the replication fork movement, positions of genes are important. Thus it should be possible to find some resemblance between composition of coding strands of genes located on the leading strand and anti-sense strands of genes located on the lagging strand. In fact we have found in the *B. burgdorferi* genome two sets of genes located on the two DNA strands with very similar sequences on the leading

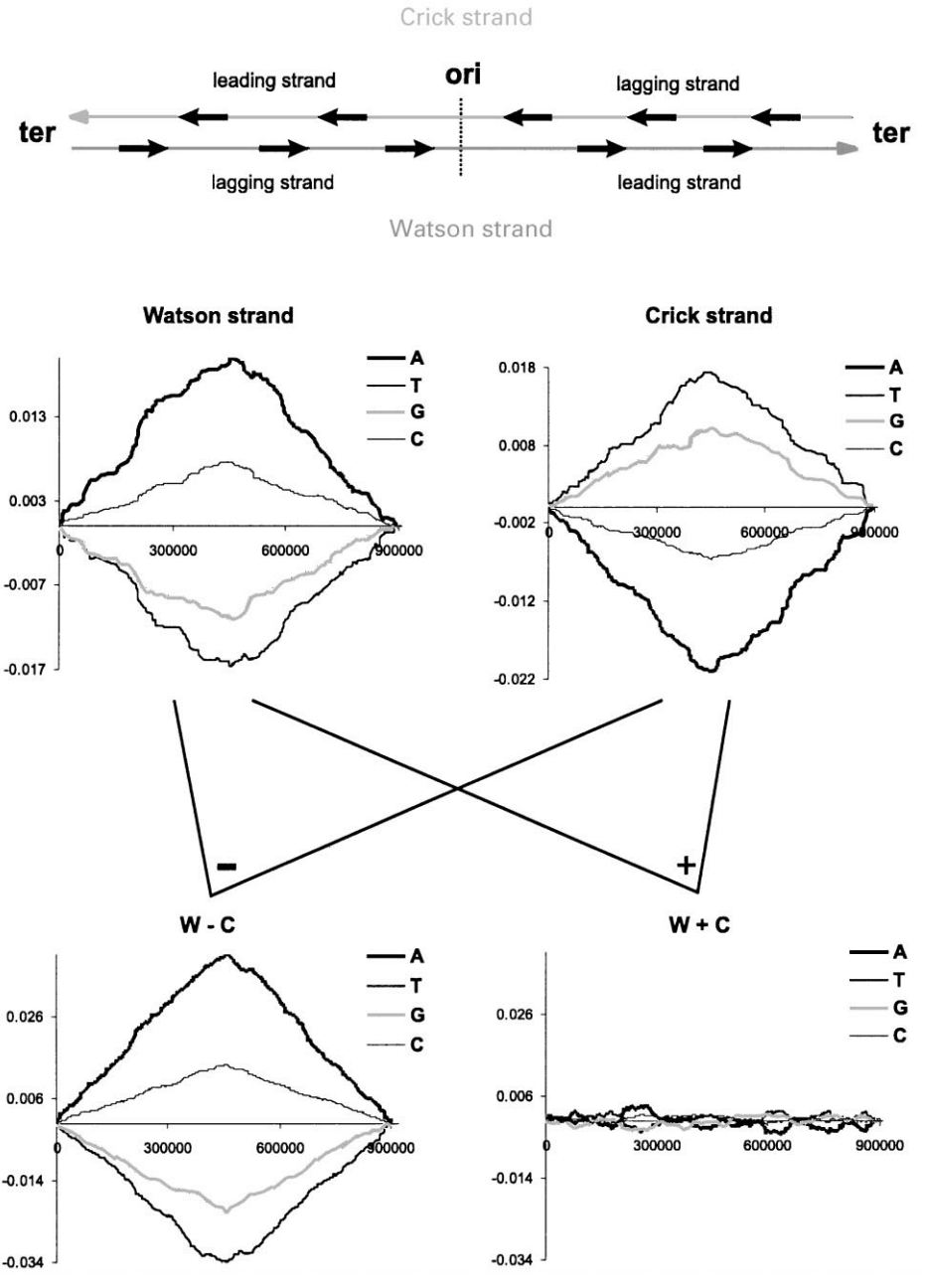

Fig. 6. Addition and subtraction of detrended DNA walks. In the upper part, the scheme of chromosome partition for leading/lagging parts and the location of ORFs are depicted. The upper line in the scheme represents C strand. It is analysed from the left end, thus, its first part is leading and at Ori its role switches to lagging. The upper two plots present detrended DNA walks describing nucleotide composition of ORFs lying on Watson strand and Crick strand, respectively. Lower left plots are the results of subtraction of walks on C strand from the corresponding walks on W strand. Lower right plots are the results of addition of walks on C strand to the corresponding walks on W strand.

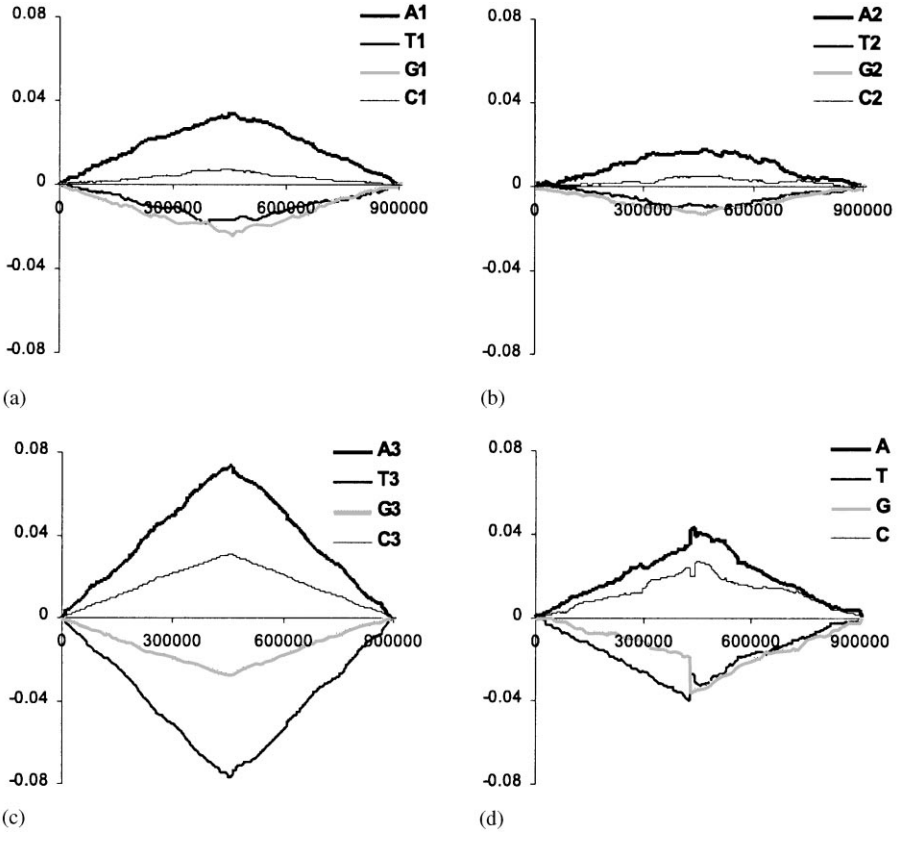

Fig. 7. Results of subtraction of detrended DNA walks on the spliced ORFs of strand C of *B. burgdorferi* from detrended DNA walks on spliced ORFs lying on strand W. Analyses were done separately for the first, the second and the third positions in codons (a, b, and c, respectively) and compared with the walks on spliced intergenic sequences (d).

strand, thus, with similar mutational pressure acting on them. These genes code for trans-membrane proteins. Their nucleotide compositions are depicted by spiders shown in Fig. 8. Spiders of genes lying on the leading strand are very similar to the spiders done for the anti-sense of genes lying on the lagging strand and vice versa. In this figure relative positions of these genes on chromosome are also shown. These very genes have to be analysed with the assumption that they are in the phase relation shown in the lower part of Fig. 8. This phase relation is not accidental. In many genomes over $\frac{2}{3}$ of all overlapping ORFs overlap in this phase relation [43,44]. We have proven that it is even possible to generate a coding sequence by another coding sequence in the anti-sense in this phase relation. Note that spiders performed for the coding strands of these genes differ significantly only in the third positions. These positions overlap in sense and anti-sense strands and their nucleotide compositions correspond to the composition of the third positions of other genes situated on the leading or the

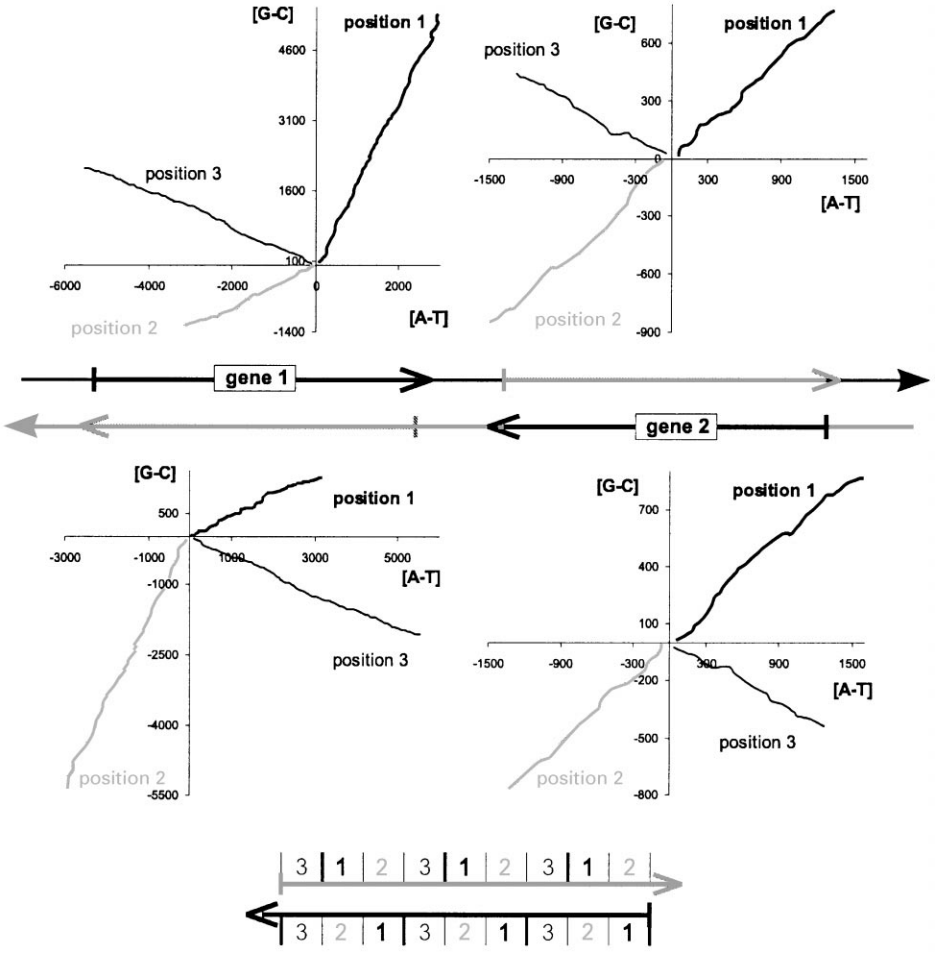

Fig. 8. Spiders performed for genes coding for trans-membrane proteins, lying on the leading strand (upper left) and the lagging strand (lower right). Spiders done for complementary (anti-sense) sequences of these genes are presented in the lower left and upper right parts of the figure, respectively. The phase relations between the coding sequences and their anti-sense sequences are shown under the plots.

lagging strands. Furthermore, the asymmetry in third positions resembles that of intergenic sequences. In this phase relation the first position in the sense strand corresponds to the second position in the anti-sense strand and the second position in the sense strand corresponds to the first position in the anti-sense strand. In the second position of these very genes thymine and cytosine prevail, which means that in the anti-sense of the first positions adenine and guanine prevail. This is a general feature of all coding sequences. Since the composition of the sense strand of trans-membrane genes lying on the leading strand resembles the composition of the anti-sense strand of genes lying on the lagging strand, the mutational pressure acting on these two sets of

genes is the same. The significant difference between these two classes of genes is in their third positions in codons. These positions are degenerated. In half of codons a substitution does not change the sense. In 60 codons (of 64 all codons) any transition in the third position does not change the sense.

4. Conclusions

In the *B. burgdorferi* genome genes are non-randomly positioned on chromosome – on the leading or lagging strand – and they have stayed at their position for a long time. It seems that translocation of genes with inversion in regard to the replication fork movement in this genome is not allowed. As a result, differences between genes lying on leading and lagging strands are observed. These differences are a result of accumulation of nucleotide substitutions introduced by replication-associated mutational pressure. Mutational pressure does not recognise codon structure of protein coding sequences, therefore, it introduces mutations independently of the role of a nucleotide in coding for the amino acid. It is selection which is responsible for removing deleterious mutations according to different rules for different positions in codons. The two effects, that of mutational pressure and that of selection, are responsible for establishing very specific differences in asymmetry in each position in codons and between DNA strands. If we assume that evolutionary mechanisms tend to optimise the effect of mutation pressure leading to the nucleotide composition of a sequence which is the least prone to mutate under the given mutational pressure, we can try to find the mutational pressure establishing DNA composition of intergenic sequences which is the least deleterious for coding sequences lying on the two DNA strands. The next paper presents the results of our first attempts to simulate mutational pressure under selection conditions depending on positions in codons.

Acknowledgements

This work was supported by the State Committee for Scientific Research, Grant No. 6PO4A 030 14.

References

[1] R. Okazaki, T. Okazaki, K. Sakabe, K. Sugimoto, A. Sugino, Proc. Natl. Acad. Sci. USA 59 (2) (1968) 598–605.
[2] A. Kornberg, T.A. Baker, DNA Replication, Freeman, New York, 1992.
[3] F.R. Blattner, G. Plunkett III, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, Ch.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, Y. Shao, Science 277 (1997) 1453–1462.
[4] A.R. Fersht, J.W. Knill-Jones, Proc. Natl. Acad. Sci. USA 78 (1981) 4251–4255.
[5] H. Echols, M.F. Goodman, Ann. Rev. Biochem. 60 (1991) 477–511.
[6] K.J. Marians, Ann. Rev. Biochem. 61 (1992) 673–719.

[7] T.A. Kunkel, Bioessays 14 (1992) 303–308.
[8] X. Veaute, R.P. Fuchs, Science 261 (5121) (1993) 598–600.
[9] S. Waga, B. Stillman, Nature 369 (1994) 207–212.
[10] W.A. Rosche, T.Q. Trinh, R.R. Sinden, J. Bacteriol. 177 (1995) 4385–4391.
[11] T. Iwaki, A. Kawamura, Y. Ishino, K. Kohno, Y. Kano, N. Goshima, M. Yara, M. Furusawa, H. Doi, F. Imamoto, Mol. Gen. Genet. 251 (1996) 657–664.
[12] I.J. Fijalkowska, P. Jonczyk, M. Maliszewska-Tkaczyk, M. Bialoskorska, R.M. Schaaper, Proc. Natl. Acad. Sci. USA 95 (1998) 10 020–10 025.
[13] M.P. Francino, H. Ochman, Trends Genet. 13 (6) (1997) 240–245.
[14] J. Mrazek, S. Karlin, Proc. Natl. Acad. Sci. USA 95 (1998) 3720–3725.
[15] A.C. Frank, J.R. Lobry, Gene 238 (1999) 65–77.
[16] J.R. Lobry, Biochimie 78 (1996) 323–326.
[17] J.R. Lobry, Mol. Biol. Evol. 13 (5) (1996) 660–665.
[18] F. Kunst, N. Ogasawara, I. Moszer, A.M. Albertini, G. Alloni, V. Azevedo, M.G. Bertero, P. Bessieres, A. Bolotin, S. Borchert et al., Nature 390 (1997) 249–256.
[19] C.M. Fraser, S. Casjens, W.M. Huang, G.G. Sutton, R. Clayton, R. Lathigra, O. White, K.A. Ketchum, R. Dodson, E.K. Hickey et al., Nature 390 (1997) 580–586.
[20] C.M. Fraser, S.J. Norris, G.M. Weinstock, O. White, G.G. Sutton, R. Dodson, M. Gwinn, E.K. Hickey, R. Clayton, K.A. Ketchum et al., Science 281 (1998) 375–388.
[21] S.G. Andersson, A. Zomorodipour, J.O. Andersson, T. Sicheritz-Ponten, U.C. Alsmark, R.M. Podowski, A.K. Naslund, A.S. Eriksson, H.H. Winkler, C.G. Kurland, Nature 396 (6707) (1998) 133–140.
[22] J.M. Freeman, T.N. Plasterer, T.F. Smith, S.C. Mohr, Science 279 (1998) 1827.
[23] A. Grigoriev, Nucleic Acids Res. 26 (10) (1998) 2286–2290.
[24] M.J. McLean, K.H. Wolfe, K.M. Devine, J. Mol. Evol. 47 (6) (1998) 691–696.
[25] S.L. Salzberg, A.J. Salzberg, A.R. Kerlavage, J.F. Tomb, Gene 217 (1998) 57–67.
[26] S. Cebrat, M.R. Dudek, A. Gierlik, M. Kowalczuk, P. Mackiewicz, Physica A 265 (1–2) (1999) 78–94.
[27] P. Mackiewicz, A. Gierlik, M. Kowalczuk, M.R. Dudek, S. Cebrat, J. Appl. Genet. 40 (1) (1999) 1–14.
[28] P. Mackiewicz, A. Gierlik, M. Kowalczuk, M.R. Dudek, S. Cebrat, Genome Res. 9 (5) (1999) 409–416.
[29] M. Picardeau, J.R. Lobry, B.J. Hinnebush, Mol. Microbiol. 32 (1999) 437–445.
[30] A. Beletskii, A.S. Bhagwat, Proc. Natl. Acad. Sci. USA 93 (1996) 13 919–13 924.
[31] M.P. Francino, L. Chao, M.A. Riley, H. Ochman, Science 272 (1996) 107–109.
[32] S. Cebrat, M.R. Dudek, Eur. Phys. J. B 3 (1998) 271–276.
[33] C.T. Zhang, R. Zhang, Nucleic Acids Res. 19 (22) (1991) 6313–6317.
[34] G. Gutierrez, L. Marquez, A. Marin, Nucleic Acids Res. 24 (13) (1996) 2525–2528.
[35] S. Cebrat, M.R. Dudek, P. Mackiewicz, M. Kowalczuk, M. Fita, Microb. Comp. Genomics 2 (4) (1997) 259–268.
[36] J. Wang, J. Biomol. Struct. Dyn. 16 (1998) 51–57.
[37] G. Perriere, J.R. Lobry, J. Thioulouse, Comput. Appl. Biosci. 12 (1996) 519–524.
[38] J.O. McInerney, Proc. Natl. Acad. Sci. USA 95 (18) (1998) 106 698–106 703.
[39] B. Lafay, A.T. Lloyd, M.J. McLean, K.M. Devine, P.M. Sharp, K.H. Wolfe, Nucleic Acids Res. 27 (1999) 1642–1649.
[40] E.P. Rocha, A. Danchin, A. Viari, Mol. Microbiol. 32 (1) (1999) 11–16.
[41] C.M. Fraser, S. Casjens, W.M. Huang, G.G. Sutton, R. Clayton, R. Lathigra, O. White, K.A. Ketchum, R. Dodson, E.K. Hickey et al., Nature 390 (1997) 580–586.
[42] Ch.L. Berthelsen, J.A. Glazier, M.H. Skolnick, Phys. Rev. A 45 (1992) 8902–8913.
[43] S. Cebrat, M.R. Dudek, Trends Genet. 12 (1996) 12.
[44] S. Cebrat, P. Mackiewicz, M.R. Dudek, Biosystems 42 (2) (1998) 165–176.