Genomics 96 (2010) 173–180

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

The probability of nonsense mutation caused by replication-associated mutational pressure is much higher for bacterial genes from lagging than from leading strands

Vladislav Victorovich Khrustalev *, Eugene Victorovich Barkovsky

Department of General Chemistry, Belarussian State Medical University, Minsk, Dzerzinskogo, 83, Belarus

ARTICLE INFO

Article history:
Received 9 February 2010
Accepted 12 June 2010
Available online 18 June 2010

Keywords:
Replichore
Chirochore
Replication-associated mutational pressure
Preterminal codon usage
8-oxo-G

ABSTRACT

We studied nucleotide usage biases in 4-fold degenerated sites of all the genes from leading and lagging strands of 30 bacterial genomes. The level of guanine in 4-fold degenerated sites (G4f) is significantly lower in genes from lagging strands than in genes from leading strands, probably because of the faster rates of guanine oxidation in single-stranded DNA leading to G to T transversions. The rates of cytosine deamination causing C to T transitions are also higher in lagging strands. We showed that the level of codons able to form stop-codons by the way of G to T transversions and C to T transitions is always higher than the level of codons able to form stop-codons by the way of C to A transversions and G to A transitions. This circumstance can be an explanation of the lower percent of ORFs in lagging strands of bacterial replichores than in leading strands.

Introduction

There are two facts known about bacterial replichores (chirochores) since their discovery [1]. The first fact is the following. Nucleotide content of the leading strands of bacterial DNA is different from the nucleotide content of the lagging strands [1–3]. There is a single origin of replication (OriC) in bacterial "chromosome" [4] and a single region of its termination (ter) [5]. In fact, the same DNA strand is leading downstream the OriC and it is lagging upstream the OriC. So, bacterial genomic DNA is separated into two parts of an opposite "chirality" (in terms of nucleotide content) by "OriC" and "ter" regions [1,2].

In this work we confirmed that there is a difference in nucleotide content of leading and lagging strands [1–3,6,7] and reported that some general features of the nucleotide content distribution between leading and lagging strands of bacteria do exist. We concentrated our attention on the differences in nucleotide usage in 4-fold degenerated sites, while classical GC-skews analysis is based on calculation of total nucleotide content [1–3,7]. All the possible nucleotide mutations are synonymous in 4-fold degenerated sites. So, the level of cytosine in 4-fold degenerated sites (C4f) and the level of guanine in them (G4f) are the most sensitive indicators of replication-associated mutational pressure [6].

The second fact known about bacterial replichores is that the density of open reading frames is higher for the leading strand than for the lagging strand of each replichore [2,6–8]. In other words, for every "replichore" the percent of coding regions situated on leading strand is higher than the percent of coding regions situated on lagging strand. In our opinion, the same mutational processes should be responsible for nucleotide usage bias observed between leading and lagging strands and for the difference in the density of coding regions between leading and lagging strands.

How can the open reading frame disappear? It can disappear by the way of nonsense mutation. What is the substrate for nonsense mutations? There are so-called preterminal codons which can become terminal by the way of a single nucleotide substitution [9,10]. The level of preterminal codon usage shows the inverse linear correlation with G+C of bacterial genes [10]. There are codons which can become terminal by the way of C to T transition and those which can become terminal by the way of G to A transition. C to T transitions (caused by cytosine deamination) should be more frequent in lagging strands [11]. It means that G to A transitions should be more frequent in leading strands.

In this work we have shown that the usage of codons which can become terminal by the way of C to T transition (PCU C to T) is always higher than the usage of codons which can become terminal by the way of G to A transition (PCU G to A). So, the probability of nonsense transition is higher for ORFs from lagging strands, than for ORFs from leading strands.

Similar situation has been discovered by us for nonsense transversions. The usage of codons which can become terminal by the way of G to T transversion (PCU G to T) is always higher than the usage of codons which can become terminal by the way of C to A transversion (PCU C to A). G to T transversions caused by the oxidation of guanine should be more frequent in ORFs from lagging strands than in ORFs from leading strands [12], unlike C to A transversions.

We came to the conclusion that the nature of genetic code and the predecessor's effect are responsible for the higher amount of the

* Corresponding author. Minsk, 220029, Communisticheskaya 7-24, Belarus.
E-mail address: vvkhrustalev@mail.ru (V.V. Khrustalev).

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.06.002

substrate for nonsense C to T and G to T mutations (in comparison with the substrate for nonsense G to A and C to A mutations) in all coding districts. Biochemical causes of replication-associated mutational pressure (RAMP) are responsible for the higher frequencies of C to T and G to T mutations in genes from lagging strands.

Results

Nucleotide usage biases in 4-fold degenerated sites of genes from lagging and leading strands

We studied 30 completely sequenced bacterial genomes with different levels of total GC-content. As one can see in Fig. 1, bacterial species from different phylums have been included in this work (Proteobacteria, Cyanobacteria, Firmicutes, Fusobacteria, Aquificae, Thermotogae, Spirochaetes, Deinococcus-Thermus).

Differences in nucleotide usage in 4-fold degenerated sites between genes from leading and genes from lagging strands for each replichore have been calculated (see Supplementary material, Tables 1–4). Although the difference in each nucleotide usage is not constant among all the replichores, general conclusions have been made after the appliance of paired differences test to the data obtained.

Level of G4f is significantly higher ($P<0.001$) in genes from leading strands than in genes from lagging strands; average difference is equal to $2.91 \pm 0.36\%$. Level of C4f is significantly lower ($P<0.001$) in genes from leading strands than in genes from lagging strands; average difference is equal to $3.54 \pm 0.37\%$.

The level of T4f is significantly higher ($P<0.05$) in genes from leading strands than in genes from lagging strands; however, an average difference is rather low; it is equal to $1.00 \pm 0.50\%$. There is no significant difference between the level of A4f in genes from leading and lagging strands.

The usage of codons able to form stop-codons by the way of a single nucleotide mutation occurred under the influence of replication-associated mutational pressure

The data described in previous subsection is consistent with experimental observations: the rates of cytosine and adenine deamination, as well as the rates of guanine and thymine oxidation

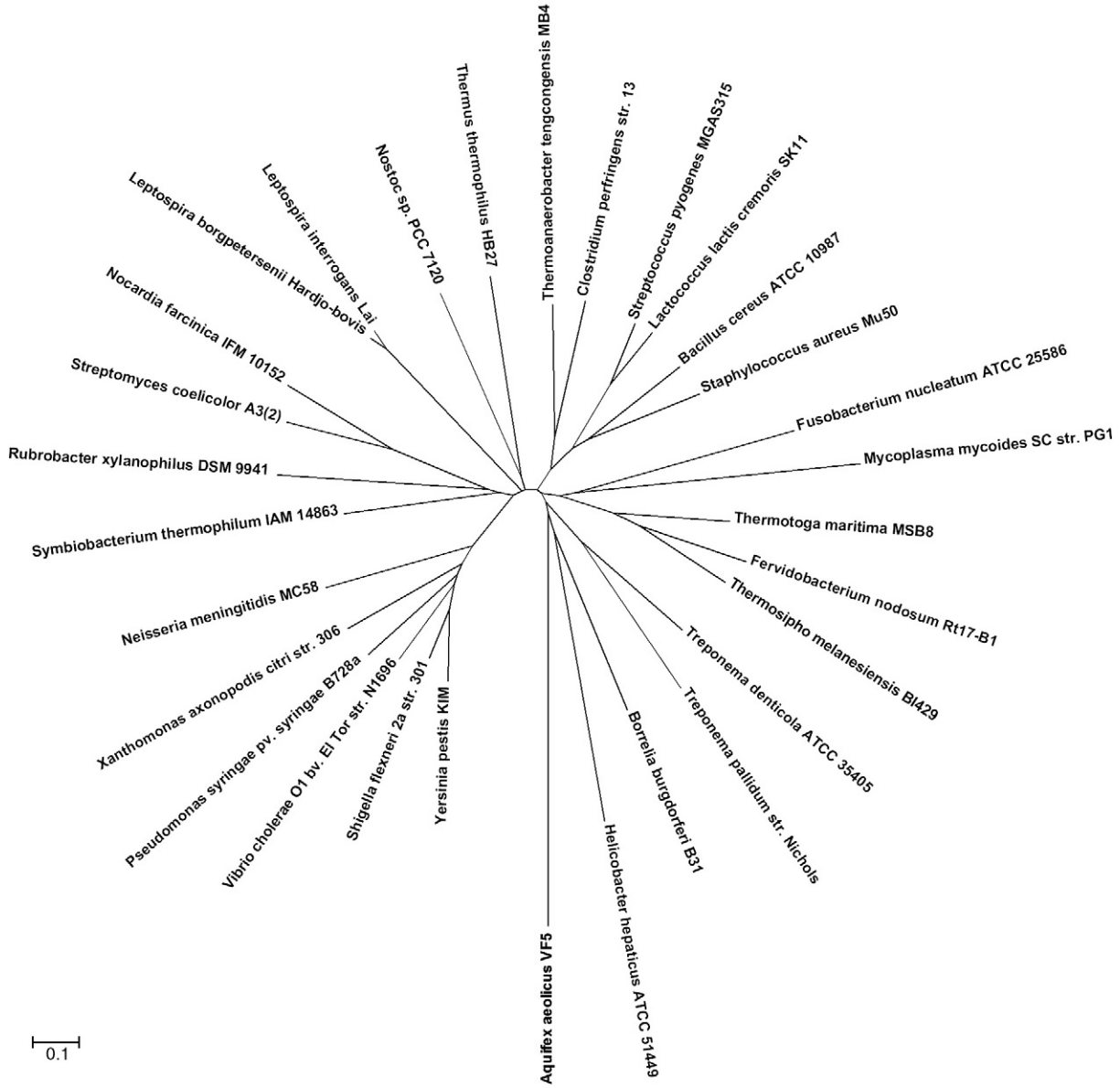

Fig. 1. NJ-tree built for amino acid sequences of DNA-polymerase I from bacterial species included in this study. PAM-matrix has been used for the alignment; evolutionary distances have been calculated by PAM-matrix (Dayhoff).