GATC avoidance in bacteria with DpnI/DpnII complementary R-M systems

A.S. Ershova,[2,3,4] I.S. Rusinov,[1] A.S. Karyagina,[2,3,4] S.A. Spirin,[1,2,5] A.V. Alexeevski[1,2,5]

[1]Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics, Moscow 119992, Russia;[2]Lomonosov Moscow State University, Belozersky Institute of Physical and Chemical Biology, Moscow 119992, Russia; [3]Gamaleya Institute of Epidemiology and Microbiology, Moscow 123098, Russia; [4]Institute of Agricultural Biotechnology, the Russian Academy of Sciences, Moscow 127550, Russia; [5]Scientific Research Institute for System Studies, the Russian Academy of Science (NIISI RAS), Moscow 117281, Russia.

asershova@gmail.com

Comparison of the observed and statistically expected numbers of oligonucleotide “words” in genomes allows to estimate selection pressure on the words. It was shown that recognition sites of Type II restriction-modification (R-M) systems are under negative pressure and are often underrepresented in prokaryotic genomes [1]. R-M systems are able to discriminate between host and foreign DNA.

Classical R-M system is able to methylate its recognition site as well as cleave DNA if their recognition site is unmethylated. Host DNA is methylated and as a result is protected from cleavage. Nevertheless, due to occasional methylation mistakes, classical Type II R-M systems can be toxic for the host. It leads to the negative selection against their sites in a host genome.

Methyl-directed R-M systems are able to cleave modified (e.g. methylated) DNA and are not toxic for a host if the host does not encode DNA-methylatransferase with the similar recognition site.

In this work we studied occurrences of sites of the methyl-directed R-M systems and compared them with classical Type II ones in 2141 complete prokaryotic genomes.

To estimate the expected number of short sequences in a genome we used S. Karlin's method [3]. A site was considered to be underrepresented if Karlin's contrast value was less than 0.78.

For all the known sites of methyl-directed R-M systems (17 Type IIM sites and 3 Type IV ones) Karlin's contrast value was analyzed in genomes that encode R-M systems with the corresponding specificity (54 and 13, correspondingly).

There are no underrepresented sites among all 13 pairs {site, genome} corresponding to Type IV R-M systems. The only underrepresented site of IIM R-M systems is GATC. This site is underrepresented in 38 of 42 cases. This almost total avoidance is surprising, since methyl-directed R-M systems are not dangerous for a bacterial genome.

GATC is known to be a recognition site of methyl-directed IIM R-M systems as well as classical Type II ones.

These systems are found in different strains of the same bacterial species. For example, in *Streptococcus pneumoniae* there are strains encoding DpnII-like R-M systems, which cleave unmethylated GATC and methylate their own DNA, as well as strains encoding Type IIM DpnI-like R-M systems, which cleave methylated GATC [4]. Such complementary DpnI/DpnII-like systems can defend mixed bacterial population from phage attacks more effectively [5].

However, such systems make difficult DNA exchange between strains with different methylation status of GATC [6]. *S. pneumoniae* with Type II DpnII-like R-M system possesses an additional methyltransferase which methylates GATC in single-strand DNA and thus prevents degradation of acquired unmethylated DNA [6].

We found that in most cases GATC avoidance is associated with the presence of DpnI-like Type IIM and DpnII-like Type II R-M system genes in different strains of the same species (S*treptococcus pneumoniae, Neisseria meningitidis, Eubacterium rectale*) or genus (*Moraxella catarrhalis, Sulfurospirillum deleyianum*).

We hypothesize that GATC avoidance facilitates DNA exchange between strains with different methylation status of GATC. In this case GATC avoidance shows the importance of

natural gene transfer for *Streptococcus, Neisseria, Eubacterium, Moraxella* and *Sulfurospirillum*.

GATC is underrepresented in 17% of all 2141 complete prokaryotic genomes. This underrepresentation could be explained by the importance of GATC palindrome for different cellular processes.

GATC plays a significant role in different DNA-based activities in dam+ bacteria encoding orphan DNA methyltransferase Dam, which methylates GATC. For example, in *E. coli* GATC is involved in gene transcription, DNA mismatch repair, initiation of chromosome replication and maintaining nucleoid structure [7]. In addition, unmethylated GATC can be hydrolyzed by mutH components of mismatch repair systems [8]. However, dam+ bacteria avoid GATC only in 3.5% cases.

Probably avoidance of GATC is resulted in the combined influence of DpnI/DpnII-like R-M systems and is a marker of DNA exchange in heterogeneous population.

The work is partially supported by RFBR grant 14-04-91350 (AE, AA) and Russian Science Foundation 14-50-00029 (IR, SS).

1. Rocha EP, Danchin A, Viari A. Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. Genome Res. 2001;11:946-58.
2. Schbath S, Prum B, de Turckheim E. Exceptional motifs in different Markov chain models for a statistical analysis of DNA sequences. J Comput Biol. 1995;2(3):417-37.
3. Karlin S, Cardon LR. Computational DNA sequence analysis. Annu Rev Microbiol. 1994;48:619-54.
4. Lacks SA, Mannarelli BM, Springhorn SS, Greenberg B. Genetic basis of the complementary DpnI and DpnII restriction systems of S. pneumoniae: an intercellular cassette mechanism. Cell. 1986;46(7):993-1000.
5. Bernheimer HP. Lysogenic pneumococci and their bacteriophages. J Bacteriol. 1979;138(2):618-24.

6. Johnston C, Martin B, Granadel C, Polard P, Claverys JP. Programmed protection of foreign DNA from restriction allows pathogenicity island exchange during pneumococcal transformation. PLoS Pathog. 2013;9(2):e1003178.
7. Leonard AC, Grimwade JE. INITIATION OF DNA REPLICATION. Ecosal Plus. 2010;2010. doi: 10.1128/ecosalplus.4.4.1.
8. Deschavanne P, Radman M. Counterselection of GATC sequences in enterobacteriophages by the components of the methyl-directed mismatch repair system. J Mol Evol. 1991;33:125-32.