

Differential Gene Survival under Asymmetric Directional Mutational Pressure

Paweł Mackiewicz¹, Małgorzata Dudkiewicz¹, Maria Kowalczyk¹,
Dorota Mackiewicz¹, Joanna Banaszak¹, Natalia Polak¹,
Kamila Smolarczyk¹, Aleksandra Nowicka¹, Mirosław R. Dudek², and
Stanisław Cebrat^{1*}

¹ Department of Genetics, Institute of Microbiology, University of Wrocław,
ul. Przybyszewskiego 63/77, PL-54148 Wrocław, Poland

{malgosia, pamac, nowicka, kowal, dorota, polak, smolar,
cebrat}@microb.uni.wroc.pl

<http://smORFland.microb.uni.wroc.pl>

² Institute of Physics, University of Zielona Góra, ul. A. Szafrana 4a,
PL-65516 Zielona Góra, Poland

mdudek@proton.if.uz.zgora.pl

Abstract. We have simulated, using Monte Carlo methods, the survival of prokaryotic genes under directional mutational pressure. We have found that the whole pool of genes located on the leading DNA strand differs from that located on the lagging DNA strand and from the subclass of genes coding for ribosomal proteins. The best strategy for most of the non-ribosomal genes is to change the direction of the mutational pressure from time to time or to stay at their recent position. Genes coding for ribosomal proteins do not profit to such an extent from switching the directional pressure which seems to explain their extremely conserved positions on the prokaryotic chromosomes.

1 Introduction

Most of the natural DNA sequences are asymmetric. There are two main mechanisms introducing DNA asymmetry: the replication-associated directional mutational pressure and the selection for protein coding sequences (see for review: [1], [2]). The replication-associated mutational pressure generates some kind of a global asymmetry between the two strands called the leading and the lagging DNA strands. On the other hand, the selection for coding sequences generates a local asymmetry between sense (coding) and anti-sense (complementary to the sense) strands of genes. This asymmetry results from the coding function requirement of genes. Thus, as in the case of two chiral molecules, the two possible ways of superposition of a coding sequence on the asymmetric bacterial chromosome are not equivalent. For example, if the sense strand of a gene located on the leading strand has more G than C, and C is more often substituted by other

* To whom all correspondence should be sent.

nucleotides than G on the leading strand, then inversion of this sequence, which transfers the C-rich anti-sense strand of the gene to the leading strand, would increase the mutation rate of the gene. Thus, a gene sequence remaining for a long time on one DNA strand tends to acquire some asymmetry characteristic for the mutational pressure while sequences occasionally inverted oscillate between the two compositional stages and their composition depends on the time which they spend on each strand and on how frequent they are translocated. In this paper we have simulated the effect of changing the mutational pressure on the gene survival.

2 Methods

Simulations have been performed on 564 leading strand genes and 286 lagging strand genes from the *Borrelia burgdorferi* genome [3], whose sequence and annotations were downloaded from GenBank. The replication-associated mutational pressure (RAMP) describing the nucleotide substitution frequencies has been parameterized as described by Kowalczyk et al. [4]. The matrix describing RAMP of the lagging strand is the mirror reflection of the RAMP for the leading DNA strand. In one Monte Carlo Step (MCS) each nucleotide of the gene sequence (its sense strand) was drawn with a probability $p_{mut} = 0.01$, then substituted by another nucleotide with the probability described by the corresponding parameter in the substitution matrix. After each round of mutations, we translated the nucleotide sequences into the amino acid sequences and compared the resulting composition of the proteins with the original. For each gene we calculated the selection parameter (T) for the amino acid composition which is the sum of absolute values of differences between fractions of amino acids as follows:

$$T = \sum_{i=1}^{20} |f_i(0) - f_i(t)|, \quad (1)$$

where: $f_i(0)$ is a fraction of a given amino-acid in the original sequence (before mutations) and $f_i(t)$ is a fraction of a given amino acid in the sequence after mutations in t MCS. If T was below the assumed threshold, a gene stayed mutated and went to the next round of mutations (the next MC step). If T trespassed the threshold - the gene was "killed" and replaced by its allele from the second genomic sequence, originally identical, simulated parallelly. As a value of the threshold we have assumed the average value T between pairs of orthologs belonging to two related genomes: *B. burgdorferi* and *Treponema pallidum* which equals 0.3. All simulations were performed for 1000 Monte Carlo steps, repeated 100 times and averaged. For comparison, the numbers of killed genes from different sets were normalized by the number of genes in the given set. In the simulations we have applied both stable and changing replication associated mutational pressure (RAMP). Stable RAMP means that during the whole simulation genes were subjected only to one pressure characteristic for the leading or the lagging strand. In the simulations with changing RAMP genes were

alternately under the RAMP characteristic for the leading or the lagging DNA strand, changing with different frequencies. These simulations were carried out in different conditions described by the two parameters: F - the fraction of MC steps during the whole simulation in which the genes were subjected to mutational pressure characteristic for the strand on which they are normally located in the genome, N - Number of switches of the RAMP from leading to lagging one or *vice versa*. In sum, we have analyzed 87 different conditions of RAMP changing (different combinations of values F and N).

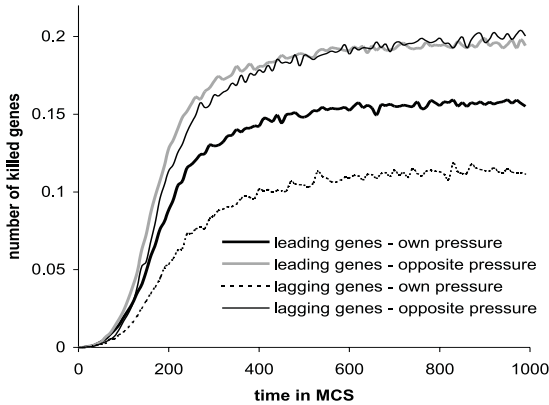


Fig. 1. The normalized number of killed genes from the leading and lagging strands of the *B. burgdorferi* genome. The genes were subjected to mutational pressure characteristic for them (their own pressure) and the mutational pressure characteristic for the complementary DNA strand (the opposite pressure)

3 Results and Discussion

After simulations of genes subjected to stable mutational pressure we found that (Fig. 1): (i) The effect of killing grew in time and approximated to a relatively high level. (ii) The killing effect for the genes staying under their own pressure is higher for the leading strand genes than for the lagging strand genes. (iii) Both sets of genes are better adapted to the mutational pressure characteristic for their recent positions in the genome than to the pressure from the opposite strand. Furthermore, the killing effect under the opposite RAMP is equally deleterious for both sets of genes.

In the earlier studies we have found that frequent changes of RAMP could be the best general strategy for gene survival [5]. In the present studies we are showing the relationship between the frequency of gene transpositions (inversions)

between differently replicating DNA strands and their survival. The diagram in Fig. 2 shows which percent of a given set of genes has the highest survival chance under one of the 87 combinations of tested parameters (F and N) after 1000 MCS of simulation. Generally, genes prefer to stay longer under the RAMP to which they are actually subjected, but there are no preferred positions for the ribosomal genes located in the *B. burgdorferi* genome on the leading strand.

In Fig. 3 we have presented how the number of killed genes depends on N for different F values. These analyses show that too frequent switching the direction of mutational pressure does not enhance significantly the gene survival. Usually switching every several hundreds of steps is close to the optimal gene survival. Relationship between the number of killed genes and F has a distinct minimum (Fig. 4). Ribosomal genes do not profit as much from switching their positions (data not shown).

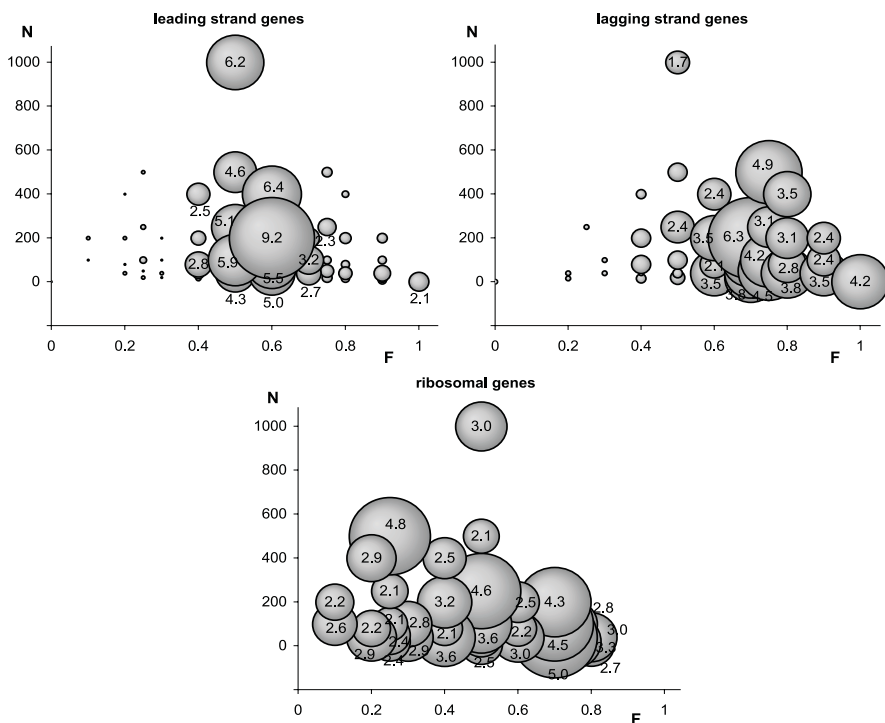


Fig. 2. Diagram presenting the best survival strategy for three sets of genes. This diagram shows which percent of a given set of genes has the highest survival chance under one of the 87 combinations of tested parameters (F and N) of changing mutational pressure after 1000 MCS of simulation

As it can be seen in Fig. 5 the number of accepted amino acid substitutions in coded proteins per site (substitutions which did not eliminate the gene function)

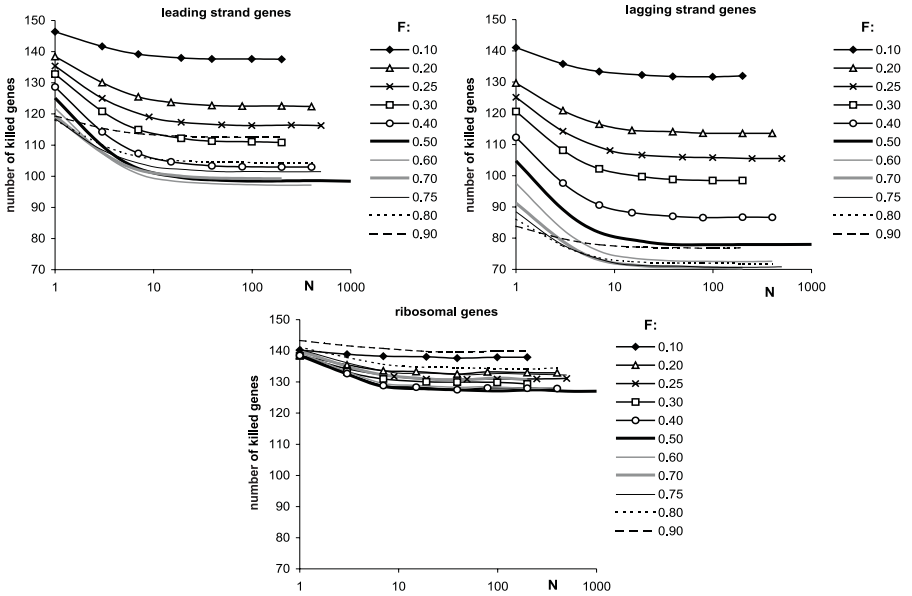


Fig. 3. Relationship between the number of killed genes and N for different F values for three sets of genes after 1000 MCS of simulation

is also higher. That means that the observed divergence of genes which recently changed their positions on chromosome should be higher, which was actually observed in numerous genomic analyses ([6] – [8]). In Fig. 5 it is also clear that the number of accepted substitutions is the lowest for the ribosomal proteins which are actually extremely conserved.

The last observations, these from simulations as well as from genome analyses lead to the conclusion that switching the direction of the mutational pressure does not diminish the total frequency of mutations but rather introduces intragenic suppression mutations which complement the former mutations in the same gene. Such intragenic suppression should be much more effective for longer genes (see accompanying paper). The ribosomal genes, in all the genomes analyzed thus far are usually located on the leading strand [9]. Our simulations have shown that they do not profit very much from transpositions (switching the mutational pressure) and the deleterious effect of the prolonged opposite mutational pressure is the same for the leading and lagging DNA strands. Since these genes are very intensively transcribed it is important for them to concert the direction of replication fork movement and the direction of transcription. This eliminates the possible deleterious effect of head on collisions of replication and transcription complexes ([10], [11]). The location of sense strands of these genes on the leading strand eliminates this effect.

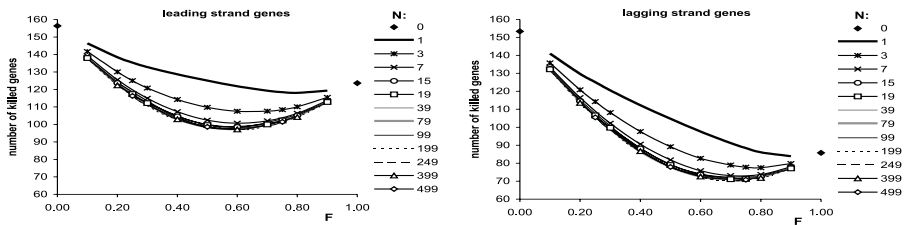


Fig. 4. Relationship between the number of killed genes and F for different N values for two sets of genes after 1000 MCS of simulation

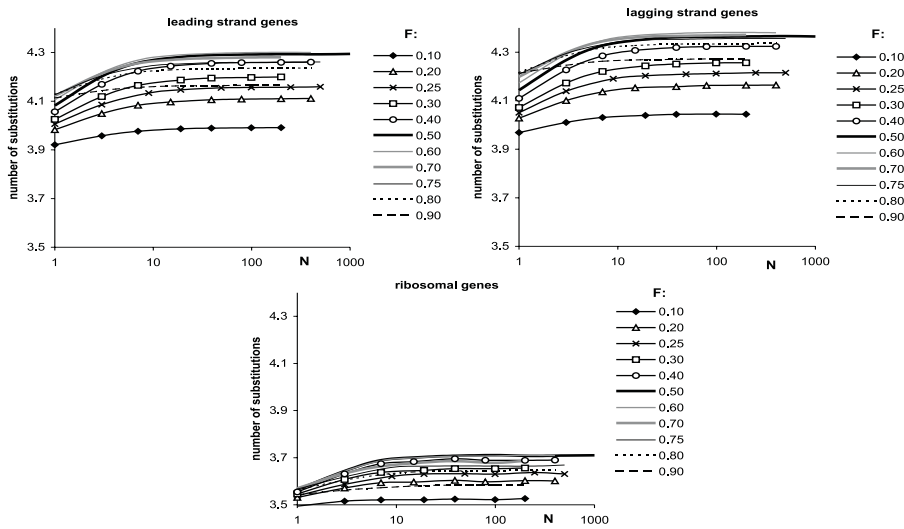


Fig. 5. Relationship between the number of accepted amino acid substitutions in coded proteins per site and N for different F values for three sets of genes after 1000 MCS of simulation

Acknowledgements. The work was supported by the grant number 1016/S/IMi/03 and is done in the frame of COST Action P10 program. M.K. was supported by the Foundation for Polish Science.

References

1. Frank, A.C., Lobry, J.R.: Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene* **238** (1999) 65–77
2. Kowalczyk, M., Mackiewicz, P., Mackiewicz, D., Nowicka, A., Dudkiewicz, A., Dudek, M.R., Cebart, S.: DNA asymmetry and the replicational mutational pressure. *J. Appl. Genet.* **42** (2001) 553–577

3. Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K. et al.: Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390** (1997) 580–586
4. Kowalczyk, M., Mackiewicz, P., Mackiewicz, D., Nowicka, A., Dudkiewicz, M., Dudek, M.R., Cebrat, S.: High correlation between the turnover of nucleotides under mutational pressure and the DNA composition. *BMC Evol. Biol.* **1** (2001) (1):13
5. Dudkiewicz, M., Mackiewicz, P., Nowicka, A., Kowalczyk, M., Mackiewicz, D., Polak, N., Smolarczyk, K., Dudek, M.R., Cebrat, S.: Properties of Genetic Code under Directional, Asymmetric Mutational Pressure. *Lect. Notes Comput. Sc.* **2657** (2003) 343–350
6. Tillier, E.R., Collins, R.A.: Replication orientation affects the rate and direction of bacterial gene evolution. *J. Mol. Evol.* **51** (2000) 459–463
7. Szczepanik, D., Mackiewicz, P., Kowalczyk, M., Gierlik, A., Nowicka, A., Dudek, M.R., Cebrat, S.: Evolution rates of genes on leading and lagging DNA strands. *J. Mol. Evol.* **52** (2001) 426–433
8. Mackiewicz, P., Mackiewicz, D., Kowalczyk, M., Dudkiewicz, M., Dudek, M.R., Cebrat, S.: High divergence rate of sequences located on different DNA strands in closely related bacterial genomes. *J. Appl. Genet.* **44** (2003) 561–584
9. McLean, M.J., Wolfe, K.H., Devine, K.M.: Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *J. Mol. Evol.* **47** (1998) 691–696
10. Brewer, B.J.: When polymerases collide: replication and the transcriptional organization of the *E. coli* chromosome. *Cell* **53** (1988) 679–686
11. French, S.: Consequences of replication fork movement through transcription units *in vivo*. *Science* **258** (1992) 1362–1365