Higher mutation rate helps to rescue genes from the elimination by selection

Malgorzata Dudkiewicz[a], Pawel Mackiewicz[a], Dorota Mackiewicz[a],
Maria Kowalczuk[a], Aleksandra Nowicka[a], Natalia Polak[a], Kamila Smolarczyk[a],
Joanna Banaszak[a], Miroslaw R. Dudek[b], Stanislaw Cebrat[a],*

[a] *Institute of Genetics and Microbiology, University of Wrocław, ul. Przybyszewskiego 63/77, 51-148 Wrocław, Poland*
[b] *Institute of Physics, University of Zielona Góra, ul. Szafrana 4A, 65-516 Zielona Góra, Poland*

Abstract

Directional mutation pressure associated with replication processes is the main cause of the asymmetry between the leading and lagging DNA strands in bacterial genomes. On the other hand, the asymmetry between sense and antisense strands of protein coding sequences is a result of both mutation and selection pressures. Thus, there are two different ways of superposition of the sense strand, on the leading or lagging strand. Besides many other implications of these two possible situations, one seems to be very important – because of the asymmetric replication-associated mutation pressure, the mutation rate of genes depends on their location. Using Monte Carlo methods, we have simulated, under experimentally determined directional mutation pressure, the divergence rate and the elimination rate of genes depending on their location in respect to the leading/lagging DNA strands in the asymmetric prokaryotic genome. We have found that the best survival strategy for the majority of genes is to sometimes switch between DNA strands. Paradoxically, this strategy results in higher substitution rates but remains in agreement with observations in bacterial genomes that such inversions are very frequent and divergence rate between homologs lying on different DNA strands is very high.

© 2004 Elsevier Ireland Ltd. All rights reserved.

Keywords: Mutation pressure; Selection pressure; DNA asymmetry; Leading strand; Gene surviving

1. Introduction

Evolution of genes requires two essential steps corresponding to mutation and selection pressures acting on the gene sequences and their products – proteins. The selection pressure should be understood as biological or biochemical requirement for gene prod-

* Corresponding author. Tel.: +48 71 3756 303;
fax: +48 71 3252 151.

E-mail addresses: mdudek@proton.if.uz.zgora.pl
(M.R. Dudek), cebrat@microb.uni.wroc.pl (S. Cebrat).
URL: http://smorfland.microb.uni.wroc.pl (S. Cebrat).

ucts, proteins, which have to fulfil different functions in the cell. Every gene is a double-stranded fragment of DNA consisting of sense (coding) and anti-sense (non-coding) strands. The sense strand codes for proteins while the other one is transcribed to mRNA which is next translated into the protein. Because of different roles of these strands, selection pressure on protein coding sequences generates the local bias in the nucleotide composition between these two strands. The bias is called DNA asymmetry and it is a consequence of selection for proper amino acid composition of the coded proteins and codon usage. The sense strand is usually purine-rich and particular codon positions have their own composition bias (Shepherd, 1981; Smithies et al., 1981; Wong and Cedergren, 1986; Karlin and Burge, 1995; Cebrat et al., 1997). On the other hand, the mutational pressure introduces nucleotide substitutions into DNA sequences and it is associated with the processes of replication and transcription. The mutation pressure associated with DNA replication generates in the majority of bacterial chromosomes a global asymmetry between two differently replicating DNA strands called the leading and the lagging strands (Lobry, 1996; Freeman et al., 1998; Grigoriev, 1998; McLean et al., 1998; Mrazek and Karlin, 1998; Mackiewicz et al., 1999a; Tillier and Collins, 2000a; Lobry and Sueoka, 2002). The leading strand is synthesised continuously and the lagging strand is synthesised from Okazaki fragments. The leading DNA strand usually is richer in guanine and thymine while the lagging DNA strand is richer in adenine and cytosine. Many other mechanisms may contribute to this asymmetry (see for review Francino and Ochman, 1997; Frank and Lobry, 1999; Kowalczuk et al., 2001a). However, the main force generating asymmetry seems to be the directional mutational pressure associated with replication, manifested by different nucleotide substitution patterns during the leading and lagging DNA strand synthesis. The most plausible explanation of the mutational pressure asymmetry seems to be the cytosine deamination theory proposed by Frank and Lobry (1999). The theory assumes that during replication process, stretches of the template for the newly synthesised lagging strand are temporarily single-stranded and are more exposed to damage and mutations, which leads in consequence to more frequent transitions of cytosines to thymines.

The asymmetry is so strong that it is reflected not only in the intergenic sequences but also in the codon usage, in amino acid composition of the coded proteins and even in the most conserved second codon positions (McInerney, 1998; Lafay et al., 1999; Mackiewicz et al., 1999b; Rocha et al., 1999). It indicates that selection pressure may accept some mutations in these positions which means that genes tolerate some codon substitutions (and in consequence some amino acid substitutions in their products). Thus, the elimination rate of genes should depend on the time they spend under the same directional mutational pressure, in fact it should grow in time.

To check this hypothesis we have analysed by computer simulations the evolution of genes in the asymmetric *Borrelia burgdorferi* genome subjected to the directional mutational pressure associated with replication. The *B. burgdorferi* genome seems to be very appropriate for such analyses because it shows the strongest asymmetry between the leading and lagging strands detected so far (McInerney, 1998; Mackiewicz et al., 1999c; Lafay et al., 1999).

2. Materials and methods

Simulations were performed with 564 leading strand genes from the *B. burgdorferi* genome (Fraser et al., 1997) whose sequence and annotations were downloaded from GenBank (ftp://www.ncbi.nlm.nih.gov). In one Monte Carlo Step (MCS) each nucleotide of the gene sequence was drawn with a probability $P_{mut} = 0.01$, then substituted by another nucleotide with the probability given in the nucleotide substitution matrix (Table 1). The matrix was constructed by the comparison of original genes with potential pseudogenes found in intergenic regions of the *B. burgdorferi* chromosome (Kowalczuk et al., 2001b). The data used in

Table 1
Nucleotide substitution matrix describing the mutational pressure on the leading strand of the *B. burgdorferi* genome

To				
From	A	T	G	C
A	–	0.103	0.067	0.023
T	0.065	–	0.035	0.035
G	0.164	0.116	–	0.015
C	0.070	0.261	0.047	–

All probabilities sum up to 1.

the construction of the matrix contained in the sum 3737 aligned sites. All numbers in the matrix sum up to 1 and correspond to the probabilities of the substitution of a given nucleotide by any other one. The matrix represents the most probable pure mutational pressure associated with replication acting on the leading strand. Because of strands' complementarity, the matrix acting on the lagging strand is a mirror reflection of the leading strand matrix. If the genes from the leading strand are under the mutational pressure characteristic for them it means that the sense strands of these genes are under the mutational pressure for the leading DNA strand. If such a gene is inverted, it means that its sense strand is under the mutational pressure characteristic for the lagging strand.

After each round of mutations, the nucleotide sequences were translated into amino acid sequences and compared to the original one. For each gene we calculated the selection parameter (T) which is the sum of absolute values of differences in fractions of each amino acid between the original sequence (f_0) and the sequence after mutations (f_t):

$$T = \sum_{i=1}^{20} |f_0^i - f_t^i|$$

It describes the deviation in the global amino acid composition of a protein coded by a given gene after

mutations, in comparison to its original sequence from the real genome.

If T was below an assumed threshold, the gene stayed mutated and went to the next MCS, if not, the gene was "killed" and replaced by its allele from the second genomic sequence, originally identical, simulated parallely.

As the value of the threshold T (equal to 0.3) we have assumed the average value counted for 442 pairs of orthologs belonging to two related genomes: *B. burgdorferi* and *Treponema pallidum*. These orthologs were extracted from COGs (clusters of orthologous groups of proteins) database downloaded from ftp://www.ncbi.nlm.nih.gov/pub/COG. Each COG contains protein sequences which are supposed to have evolved from one ancestral protein and fulfil the same function (Tatusov et al., 2001).

The number of accumulated substitutions and the number of replacements of genes were counted after each MCS. All simulations were performed for 1600 MCS, repeated 300 times and averaged.

3. Results

In Fig. 1 the number of accumulated amino acid substitutions introduced during the prolonged simulations
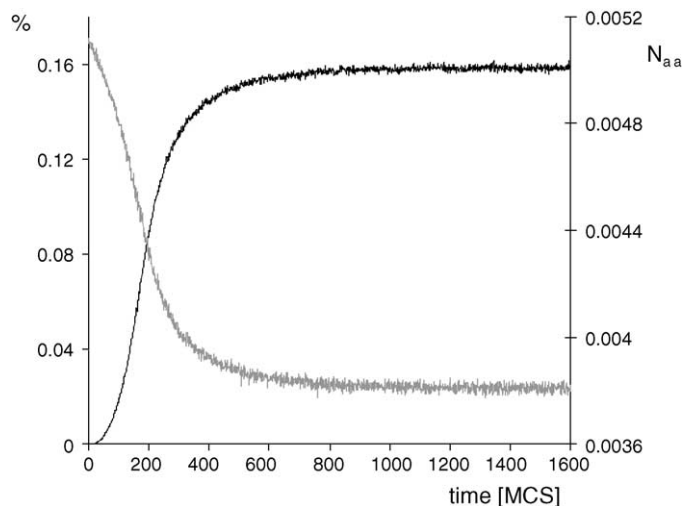

Fig. 1. The percentage of killed genes (the black line, the left *Y* axis) from the leading strand of the *B. burgdorferi* genome and the number of accepted amino acid substitutions per site (the grey line, the right *Y* axis) in the encoded proteins (N_{aa}) in the course of their evolution when the leading strand base substitution matrix was applied for the whole simulation.
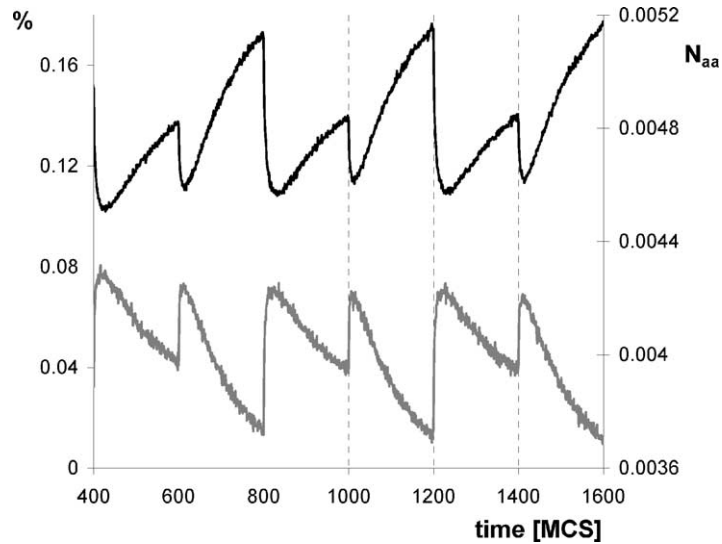
Fig. 2. The percentage of killed genes (the black line, the left Y axis) from the leading strand of the *B. burgdorferi* genome and the number of accepted amino acid substitutions per site (the grey line, the right Y axis) in the encoded proteins (N_{aa}) in the course of their evolution in the case when substitution matrix switched every 200 time steps from the one specific for the leading strand to the mirror one – from the lagging strand and vice versa.

and the percentage of "killed" genes are shown. During the whole period of the simulation, the genes positioned in the real genome on the leading DNA strand were under the mutation pressure characteristic for the leading strand. The number of substitutions introduced into sequences decreases slightly in time, as sequences adapt to the mutational pressure and their compositions reflect better the composition of the sequence being in the equilibrium with the mutational pressure. Such an effect of adaptation is manifested by compositional asymmetry in prokaryotic genomes. The percentage of killed genes (allelic replacements) increases in time and it asymptotically approximates a relatively high value 0.16%. The accumulation of substitutions measured as divergence rate between the original amino acid sequences and sequences after simulation decreases as the killing effect increases. In Fig. 2 the effect of the switching the DNA strand (and in consequence the replication-associated mutational pressure) by genes is shown. After 200 MCS the leading strand base substitution matrix was replaced by the "mirror" one – corresponding to the lagging strand's one, and after the next 200 steps back by the previous one, and so on. It mimics the inversion of genes.

In these simulations, the rate of genes elimination decreased after each "inversion". The number of ac-

cepted substitutions increased rapidly when the genes changed the strand. Moreover, there is statistically significant, strong negative correlation between the number of accumulated substitutions and the number of killed genes. We obtained similar results when the simulations were preformed on the lagging strand genes (data not shown).

4. Discussion

The results suggest that it is profitable for genes to be translocated from time to time to a differently replicating strand because such an inversion decreases the probability of their elimination and increases their survival. Paradoxically, accumulating more substitutions and higher divergence is accompanied by a lower killing effect. But this paradox could be easily explained by the higher probability of intragenic suppressor substitutions just after the switch. The translocation to the other strand results in the opposite directional mutational pressure acting on the translocated gene and may reverse its composition to a proper state. For example, if the level of hydrophobicity drops below the threshold because of the mutational trend introducing too much adenine into the second codon

positions, switching the mutation matrix replaces the directional mutation pressure by the one with a trend of introducing more thymine, which codes for hydrophobic amino acids if in the second codon position. The other simple explanation is intrinsic to the genetic code itself. Note, that all two-fold degenerate codons are discriminated by purine versus pyrimidine in the third codon positions. This corresponds to the compositional bias between the leading and the lagging DNA strands.

The introduced mechanism of selection by the crude amino acid composition parameter is a very degenerate one. However, application of tolerance for amino acid composition as selection parameter has its justification. The amino acid global composition of proteins is sufficiently unique and sensitive to be used for identification of particular proteins (Sibbald et al., 1991; Hobohm et al., 1994; Wilkins et al., 1996). An unknown protein may be identified by comparing its amino acid composition with compositions of proteins collected in the database with especially developed programmes (Wilkins et al., 1996, 1998; Appel et al., 1994; Hobohm and Sander, 1995; Grillasca et al., 2000). These methods can be applied even to cross-species protein identification because amino acid composition of functionally similar proteins is well conserved (Cordwell et al., 1995; Galat et al., 1996; Wilkins and Williams, 1997). Moreover, many authors have found that amino acid composition is unique for many protein families (Hobohm and Sander, 1995; Galat et al., 1996; Galat and Rioux, 1997) and compositional similarity search may be used in finding structural homologs even if sequence similarity falls below 25% (Hobohm and Sander, 1995). Nevertheless, simulations performed on sequences of individual genes with tolerance applied on the level of the hydrophobicity or isoelectric point of their products have given very similar results – lower rate of elimination of genes under temporarily switched mutational pressure, accompanied with higher divergence rate after switches. Although the results were obtained from analyses of simulations performed on one genome, they should be the same for many other bacterial genomes showing the DNA asymmetry which is caused by the asymmetric directional mutational pressure. This phenomenon of asymmetry seems to be universal in an overwhelming number of bacterial genomes. The results obtained in the computer simulations are in good agreement with analyses of genes evolution and rearrangements in prokaryotic

genomes. Many authors have observed very high divergence just for orthologs which are located on differently replicating strands in the compared genomes (Tillier and Collins, 2000b; Rocha and Danchin, 2001; Szczepanik et al., 2001; Mackiewicz et al., 2003a). Moreover, it was found that genes which have switched DNA strand accommodate to a new mutational pressure and, in respect to their composition, become similar to genes of the new strand (Lafay et al., 1999; Tillier and Collins, 2000b; Rocha and Danchin, 2001).

Our results show that many genes which change DNA strand have a much higher probability of surviving than the genes which stay on the same strand for a long time. It suggests that positions of genes should not be stable on chromosome for long evolutionary distances in compared genomes. Actually, many authors have shown that rearrangements are very common in bacterial chromosomes (Mushegian and Koonin, 1996; Kolsto, 1997; Watanabe et al., 1997; Bellgard et al., 1999; Itoh et al., 1999; Hughes, 2000) and the fraction of orthologs lying on the differently replicating strands increases very quickly with the phylogenetic distances between the compared genomes (Mackiewicz et al., 2003b). Accordingly, the fraction of orthologs staying in the analysed genomes on the same strand drops very quickly and reaches only a few percentage. These genes at conserved positions code for ribosomal proteins and stay mainly on the leading strand.

In conclusion, the directional, asymmetric mutational pressure would eventually generate in equilibrium two sequences differing in the nucleotide composition – one characteristic for the leading DNA strand and the other characteristic for the lagging DNA strand. Sequences of genes can oscillate between these two states due to inversions – relatively frequent recombination mechanism, avoiding the killing effect of the prolonged mutational pressure of the same substitution trends.

This strategy seems to be available only for prokaryotic genomes where the conservation of position of a particular gene in the genome is not very important – genes can move around the genome. This is not a case for eukaryotic, sexually reproducing organisms. In such species, the positions of genes are much more conserved because of the possibility of producing unbalanced gametes. Even if the replication-associated mutational pressure is also asymmetric (Gierlik et al., 2000; Niu et al., 2003) eukaryotic genes have a possi-

bility of switching between the two pressures because the strategy of the genome replication of eukaryotes is different. There are many autonomously replicated sequences (ARSs) initiating replication. Since there is no absolute synchronisation of the initiation of replication, the same region can be replicated from different ARS in the consecutive replication cycles. Thus, a gene sequence, "choosing" the proper position between ARSs can fit to the proper frequency of switches between the two asymmetric replication-associated mutation pressures. Such tuning could be executed simply by an insertion or a deletion of an intergenic sequence between the gene and ARS. It can explain why there are so many intergenic sequences in the eukaryotic genomes and why so many eukaryotic intergenic sequences are transposable while in prokaryotic genomes rather the coding sequences are movable.

Acknowledgment

References

Appel, R.D., Bairoch, A., Hochstrasser, D.F., 1994. A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. Trends Biochem. Sci. 19, 258–260.

Bellgard, M.I., Itoh, T., Watanabe, H., Imanishi, T., Gojobori, T., 1999. Dynamic evolution of genomes and the concept of genome space. Ann. N.Y. Acad. Sci. 870, 293–300.

Cebrat, S., Dudek, M.R., Mackiewicz, P., Kowalczuk, M., Fita, M., 1997. Asymmetry of coding versus non-coding strands in coding sequences of different genomes. Microb. Comp. Genomics 2, 259–268.

Cordwell, S.J., Wilkins, M.R., Cerpa-Poljak, A., Gooley, A.A., Duncan, M., Williams, K.L., Humphery-Smith, I., 1995. Cross-species identification of proteins separated by two-dimensional gel electrophoresis using matrix-assisted laser desorption ionisation/time-of-flight mass spectrometry and amino acid composition. Electrophoresis 16, 438–443.

Francino, M.P., Ochman, H., 1997. Strand asymmetries in DNA evolution. Trends Genet. 13, 240–245.

Frank, A.C., Lobry, J.R., 1999. Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. Gene 238, 65–77.

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., 1997. Genomic sequence of a Lyme disease spirochaete *Borrelia burgdorferi*. Nature 390, 580–586.

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

Galat, A., Rioux, V., 1997. Convergence of amino acid compositions of certain groups of protein aids in their identification on two-dimensional electrophoresis gels. Electrophoresis 18, 443–451.

Galat, A., Gerbod, M.C., Bouet, F., Riviere, S., 1996. Proteins and their amino acid compositions: uniqueness, variability, and applications. Arch. Biochem. Biophys. 330, 229–237.

Gierlik, A., Kowalczuk, M., Mackiewicz, P., Dudek, M.R., Cebrat, S., 2000. Is there replication-associated mutational pressure in the *Saccharomyces cerevisiae* genome? J. Theor. Biol. 202, 305–314.

Grigoriev, A., 1998. Analysing genomes with cumulative skew diagrams. Nucleic Acids Res. 26, 2286–2290.

Grillasca, J.P., Planells, R., Aubert, D., Antonini, F., 2000. SPAC: identification of polypeptides using their amino-acid composition. Comput. Chem. 24, 235–238.

Hobohm, U., Sander, C., 1995. A sequence property approach to searching protein databases. J. Mol. Biol. 251, 390–399.

Hobohm, U., Houthaeve, T., Sander, C., 1994. Amino acid analysis and protein database compositional search as a rapid and inexpensive method to identify proteins. Anal. Biochem. 222, 202–209.

Hughes, D., 2000. Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. Genome Biol. 1:REVIEWS0006.

Itoh, T., Takemoto, K., Mori, H., Gojobori, T., 1999. Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. Mol. Biol. Evol. 16, 332–346.

Karlin, S., Burge, C., 1995. Dinucleotide relative abundance extremes: a genomic signature. Trends Genet. 11, 283–290.

Kolsto, A.B., 1997. Dynamic bacterial genome organization. Mol. Microbiol. 24, 241–248.

Kowalczuk, M., Mackiewicz, P., Mackiewicz, D., Nowicka, A., Dudkiewicz, M., Dudek, M.R., Cebrat, S., 2001a. DNA asymmetry and the replicational mutational pressure. J. Appl. Genet. 42, 553–577.

Kowalczuk, M., Mackiewicz, P., Mackiewicz, D., Nowicka, A., Dudkiewicz, M., Dudek, M.R., Cebrat, S., 2001b. High correlation between the turnover of nucleotides under mutational pressure and the DNA composition. BMC Evol. Biol. 1, 13.

Lafay, B., Lloyd, A.T., McLean, M.J., Devine, K.M., Sharp, P.M., Wolfe, K.H., 1999. Proteome composition and codon usage in spirochaetes: species-specific and DNA strand-specific mutational biases. Nucleic Acids Res. 27, 1642–1649.

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

Lobry, J.R., Sueoka, N., 2002. Asymmetric directional mutation pressures in bacteria. Genome Biol. 3:RESEARCH0058.

Mackiewicz, P., Gierlik, A., Kowalczuk, M., Dudek, M.R., Cebrat, S., 1999a. Asymmetry of nucleotide composition of prokaryotic chromosomes. J. Appl. Genet. 40, 1–14.

Mackiewicz, P., Gierlik, A., Kowalczuk, M., Dudek, M.R., Cebrat, S., 1999b. How does replication-associated mutational pressure influence amino acid composition of proteins? Genome Res. 9, 409–416.

Mackiewicz, P., Gierlik, A., Kowalczuk, M., Szczepanik, D., Dudek, M.R., Cebrat, S., 1999c. Mechanisms generating long-range cor-

relation in nucleotide composition of the *Borrelia burgdorferi* genome. Physica A 273, 103–115.

Mackiewicz, D., Mackiewicz, P., Kowalczuk, M., Dudkiewicz, M., Dudek, M.R., Cebrat, S., 2003a. Rearrangements between differently replicating DNA strands in asymmetric bacterial genomes. Acta Microb. Pol. 52, 245–261.

Mackiewicz, P., Mackiewicz, D., Kowalczuk, M., Dudkiewicz, M., Dudek, M.R., Cebrat, S., 2003b. High divergence rate of sequences located on different DNA strands in closely related bacterial genomes. J. Appl. Genet. 44, 561–584.

McInerney, J.O., 1998. Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi*. Proc. Natl. Acad. Sci. U.S.A. 95, 10698–10703.

McLean, M.J., Wolfe, K.H., Devine, K.M., 1998. Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. J. Mol. Evol. 47, 691–696.

Mrazek, J., Karlin, S., 1998. Strand compositional asymmetry in bacterial and large viral genomes. Proc. Natl. Acad. Sci. U.S.A. 95, 3720–3725.

Mushegian, A.R., Koonin, E.V., 1996. Gene order is not conserved in bacterial evolution. Trends Genet. 12, 289–290.

Niu, D.K., Lin, K., Zhang, D.-Y., 2003. Strand compositional asymmetries of nuclear DNA in eukaryotes. J. Mol. Evol. 57, 325–333.

Rocha, E.P., Danchin, A., 2001. Ongoing evolution of strand composition in bacterial genomes. Mol. Biol. Evol. 18, 1789–1799.

Rocha, E.P., Danchin, A., Viari, A., 1999. Universal replication biases in bacteria. Mol. Microbiol. 32, 11–16.

Shepherd, J.C., 1981. Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary justification. Proc. Natl. Acad. Sci. U.S.A. 78, 1596–1600.

Sibbald, P.R., Sommerfeldt, H., Argos, P., 1991. Identification of proteins in sequence databases from amino acid composition data. Anal. Biochem. 198, 330–333.

Smithies, O., Engels, W.R., Devereux, J.R., Slightom, J.L., Shen, S.H., 1981. Base substitutions, length differences, and DNA strand asymmetries in the human Gγ and Aγ fetal globin gene region. Cell 26, 345–353.

Szczepanik, D., Mackiewicz, P., Kowalczuk, M., Gierlik, A., Nowicka, A., Dudek, M.R., Cebrat, S., 2001. Evolution rates of genes on leading and lagging DNA strands. J. Mol. Evol. 52, 426–433.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., Koonin, E.V., 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 29, 22–28.

Tillier, E.R., Collins, R.A., 2000a. The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. J. Mol. Evol. 50, 249–257.

Tillier, E.R., Collins, R.A., 2000b. Replication orientation affects the rate and direction of bacterial gene evolution. J. Mol. Evol. 51, 459–463.

Watanabe, H., Mori, H., Itoh, T., Gojobori, T., 1997. Genome plasticity as a paradigm of eubacterial evolution. J. Mol. Evol. 44 (Suppl. 1), 57–64.

Wilkins, M.R., Williams, K.L., 1997. Cross-species protein identification using amino acid composition, peptide mass fingerprinting, isoelectric point and molecular mass: a theoretical evaluation. J. Theor. Biol. 186, 7–15.

Wilkins, M.R., Gasteiger, E., Wheeler, C.H., Lindskog, I., Sanchez, J.C., Bairoch, A., Appel, R.D., Dunn, M.J., Hochstrasser, D.F., 1998. Multiple parameter cross-species protein identification using MultiIdent – a world-wide web accessible tool. Electrophoresis 19, 3199–3206.

Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., Williams, K.L., Hochstrasser, D.F., 1996. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Biotechnology (New York) 14, 61–65.

Wong, J.T., Cedergren, R., 1986. Natural selection versus primitive gene structure as determinant of codon usage. Eur. J. Biochem. 159, 175–180.