

Genomic Distance between Thymidylate Synthase and Dihydrofolate Reductase Genes Does Not Correlate With Phylogenetic Evolution in Bacteria

Vitor Hugo Moreau

Department of Bio-Interaction-ICS, Federal University of Bahia (UFBA), Salvador, BA, Brazil
Nucleus of Biotechnology (NuBiotec), Faculty of Technology and Sciences (FTC), Salvador, BA, Brazil

Abstract: Problem statement: Dihydrofolate Reductase (DHFR) and Thymidylate Synthase (TS) exist as bifunctional enzymes coded into unique polypeptide chain in protozoans. Bifunctional DHFR-TS is associated with an increase in the enzymatic activity by channeling the substrate between the active sites. In some bacteria, DHFR and TS genes are neighbors in the genome, whereas in others, they are located millions of base pairs apart. Gene neighboring gained importance in evolution because it was found to promote the interaction between expressed proteins in gene clusters. Co-expression of neighboring genes might favor protein associations, increasing the enzymatic efficiency. The basis of genomic evolution that leads to gene ordering is not totally understood; however, one could suppose that increasing the efficiency of metabolic pathways could work as an evolutionary pressure to get genes together in the genome of an organism. **Approach:** In this study, phylogenetic analysis of DHFR and TS sequences and the genomic distance between these genes in bacteria were measured. **Results:** No significant correlation was found between genomic distances, in base pairs, of DHFR and TS genes and phylogenetic distance among the studied bacteria. **Conclusion/Recommendations:** This suggested that DHFR and TS enzymes clusters, even if they are coexpressed, might not exert a pivotal role in natural selection of bacteria.

Key words: Dihydrofolate reductase, thymidylate synthase, gene neighboring, gene order, enzyme cluster

INTRODUCTION

Study of gene ordering has become a promising area of genetics since the sequencing of numerous eukaryotic and prokaryotic genomes in the 1990's. In prokaryotes, conservation of gene order follows a common trend in all the species; gene order is generally well preserved at close phylogenetic distances (Tamames *et al.*, 1997). However, it is even possible to observe a lack of gene order conservation as phylogenetic distance increases, mainly because of cluster of genes that remain well conserved during bacterial evolution (Lathe *et al.*, 2000; Huynen and Bork, 1998). Information about co-localized prokaryotic genes can be used to derive functional inferences; for instance, if the function of one gene in a conserved gene cluster is known, the function of a neighboring gene can be inferred (Aravind, 2000). Conservation of gene order can be because of any of the following three reasons: (i) The species have diverged only recently and the gene order has not yet been destroyed; (ii) there has been lateral gene transfer of blocks of genes and (iii) the integrity of the cluster is important to the fitness of the cell (Tamames, 2001).

The biological significance of gene colocalization is well described mainly in prokaryotes. For instance, the operons, clusters of adjacent coexpressed genes that often encode functionally associated proteins, represent the principal form of gene coregulation in prokaryotes (Rogozin *et al.*, 2004). Proposed explanations about selection of the gene ordering include promotion of protein interaction encoded by the neighboring genes in the cluster (Dandekar *et al.*, 1998).

In this study, correlation between the phylogenetic distances of Dihydrofolate Reductase (DHFR) and Thymidylate Synthase (TS) genes has been studied in prokaryotes. Phylogenetic analysis has been performed and the distance between TS and DHFR genes was measured in several bacterial genomes taken from NCBI databank. TS and DHFR are enzymes that work together in successive reactions of folate synthesis pathway. These proteins exist, in some protozoans, as bifunctional enzymes, which are important in channeling the substrate between TS and DHFR active sites (Atreya and Anderson, 2004; Miles *et al.*, 1999; Trujillo *et al.*, 1997). These genes are neighbors in some bacterial genomes, whereas they are located millions of base pairs apart (and even in opposite DNA

strands) in others. Thus, multiple alignments have been performed in this study to calculate the Jones-Taylor-Thornton (JTT) distance matrix between TS and DHFR sequences of 129 bacterial genomes in which TS and DHFR genes are located in the same DNA strand. It has not been possible to observe significant correlation between the TS and DHFR genes distances and its JTT distance from *Pseudomonas syringae*. *P. syringae* displays the longest genomic distance between TS and DHFR genes within the studied bacteria (5.5×10^6 bp).

Although gene neighboring hypothesis of evolution in prokaryotes demands more studies, it is supposed that TS and DHFR genes were not evolutionarily driven to get closer in bacterial genomes in the studied bacterial evolution.

MATERIALS AND METHODS

Protein tables of 385 bacterial genomes were collected from the NCBI databank (www.ncbi.nlm.nih.gov) and the genomic positions, DNA strand signal and identification numbers (gid) of TS and DHFR genes were extracted. Distances between TS and DHFR genes were measured in base pairs (bp) for each bacterium. A total of 129 bacterial genomes displayed TS and DHFR genes in the same DNA strand and were selected to compose the sample set for this study. Phylogenetic analysis with protein sequences of TS and DHFR of the bacteria in the sample set was then performed by multiple alignments, calculation of distance matrix and construction of neighbor-joining phylogenetic tree. Multiple alignments were performed with Clustal W (Higgins *et al.*, 1994) using BLOSUM weight matrix and gap and gap extension penalties of 20 and 0.25, respectively. Multiple alignment results were manually corrected to fix gap and alignment mistakes. Jones-Taylor-Thornton (JTT) distance matrices were calculated using PROTDIST software (PHYLIP package) and NJ phylogenetic trees were then constructed using NEIGHBOR (PHYLIP) (Felsenstein, 2005). Phylogenetic trees were bootstrapped 1000 times to estimate the confidence of each node. JTT distance from *P. syringae* and error bars showed under results section were, respectively, calculated from the mean and standard deviation of 1000 bootstrapped distance matrices.

RESULTS

Gene order is known to be well conserved at close phylogenetic distances. Nevertheless, the evolutionary basis that leads to gene clustering is not totally understood. Phylogenetic distance is a good parameter

to measure the evolutionary distance between organisms and a phylogenetic distance matrix can be constructed from a set of DNA or protein sequences of homologous genes from such organisms (Tamames *et al.*, 1997; Lathe *et al.*, 2000; Huynen and Bork, 1998; Aravind, 2000; Tamames, 2001; Rogozin *et al.*, 2004; Dandekar *et al.*, 1998). Here, a phylogenetic analysis of 129 bacterial species using their TS and DHFR sequences was presented. Distance matrix calculated from DHFR sequences was used to analyze how well the evolutionary distance between the studied species correlated with the distance between TS and DHFR genes in the bacterial genomes. Figure 1 shows genomic representations of TS and DHFR of two *Clostridium acetobutylicum* (Panel A), *Staphylococcus aureus* (Panel B), *Deinococcus radiodurans* (Panel C) and *Pseudomonas syringae* (Panel D). In *C. acetobutylicum*, TS and DHFR genes were contiguous, with no base pairs between them and only the stop codon at the end of the TS gene avoided the putative production of a bifunctional enzyme. In *S. aureus*, TS and DHFR genes were neighboring, but were separated by a 364 bp DNA strand. In *D. radiodurans*, there was a whole expressed DNA sequence between TS and DHFR genes (a putative deoxycytidylate deaminase gene) and, consequently, those genes were not neighbors but were located in close proximity. Last, in *P. syringae*, TS and DHFR genes were located in opposite sides of bacterium genome, being separated by more than 5 million base pairs. Figure 2 shows the distribution of bacterial species as a function of DHFR-TS gene distances in the genome. In species in which DHFR and TS genes were neighbors, DNA sequences between these two genes were no longer than 346 bp, as depicted in Fig. 1. Those sequences were analyzed by multiple alignment have presented lack of similarity with each other (data not shown), with exception of species of the same genus. This, in part, could explain the absence of homology found between junctional peptides of bifunctional DHFR-TS in protozoa (Atreya and Anderson, 2004; Miles *et al.*, 1999; Trujillo *et al.*, 1997).

To confirm if the differences observed in bacterial genomic distances between TS and DHFR genes were not related to evolutionary process, correlation between DHFR/TS gene distances and the phylogenetic distance of each studied bacterium from *P. syringae* were analyzed. Phylogenetic distance was calculated using both DHFR and TS sequences. *P. syringae* presented the longest DNA fragment separating TS and DHFR genes (5,525,240 bp) and was used as a standard for comparison with other species, because it was located in an evolutionary edge of the studied sample.

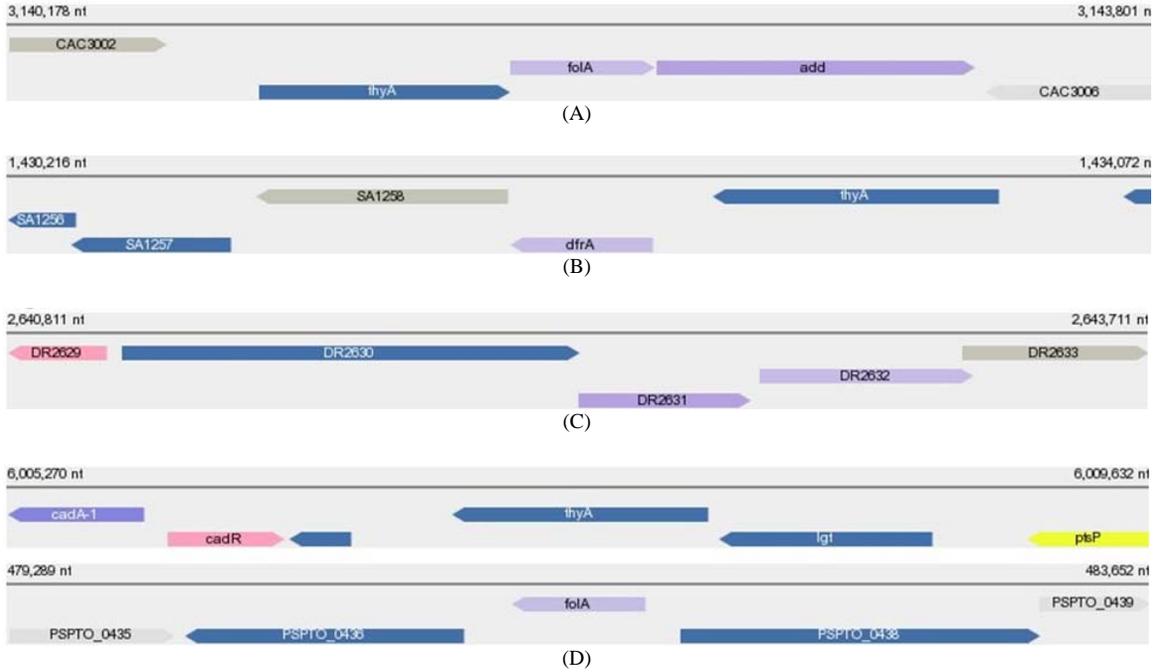


Fig. 1: Genomic representation of TS and DHFR of four bacteria from dataset. (Panel A) TS (*thyA*) and DHFR (*folA*) from *Clostridium acetobutylicum* (neighbor genes with no base pairs in between); (Panel B) TS (*thyA*) and DHFR (*dfrA*) from *Staphylococcus aureus* (Neighbor genes with 200 bp in between); (Panel C) TS (*DR2630*) and DHFR (*DR2632*) from *Deinococcus radiodurans* (Not neighbor genes and 458 bp in between); (Panel D) TS (*thyA*) and DHFR (*folA*) from *Pseudomonas syringae* (not neighbor genes with 5,525,240 bp in between). Figure edited from NCBI genome server images of genomic representations

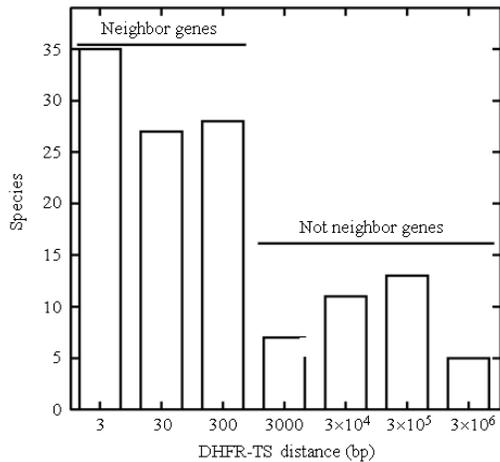


Fig. 2: Distribution of studied bacterial species by DHFR-TS gene distances. Distance ranges are 0-3; 4-30; 31-300; 301-3,000; 3,001-30,000; 30,001-300,000 and 300,001-3,000,000 bp. Bars above the data indicate species in which DHFR and TS genes are neighboring or not neighboring in the bacterial genome

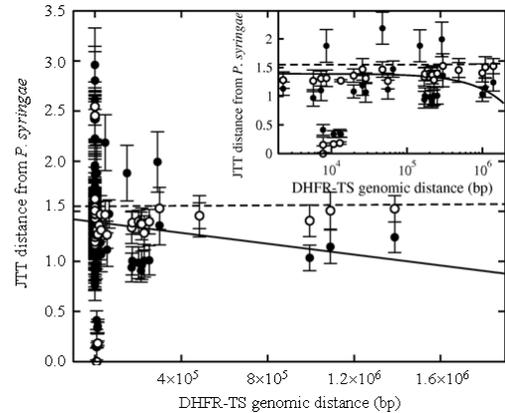


Fig. 3: Correlation plots between TS-DHFR genomic distance and Jones-Taylor-Thornton (JTT) phylogenetic distance from *P. syringae* calculated from the mean of 1000 bootstrapped distance matrix using DHFR sequences. Inset shows x-axes in logarithm scale. Error bars represent the standard deviation between 1000 bootstrapped distance matrices. Correlation coefficients (R) were: -0.12 and 0.01

Figure 3 shows the correlation plot between genomic TS/DHFR distance, in base pairs and phylogenetic distance calculated from Jones-Taylor-Thornton (JTT) distance matrix. No significant correlation was found between DHFR/TS gene distances and JTT distance from *P. syringae* calculated using DHFR (closed circles) or TS (open symbols) sequences. This indicated that there was no significant correlation between TS/DHFR gene distances and the evolutionary distance from *P. syringae* calculated using either DHFR ($R = -0.12$) or TS ($R = 0.01$) sequences. When bacterial groups that contained neighboring DHFR and TS genes (Fig. 2) were compared with those that did not have neighboring DHFR and TS genes, no significant correlation between JTT distance and phylogenetic distance from *P. syringae* could be observed as well (data not shown).

DISCUSSION

Distance matrix calculated from DHFR sequences has been used to analyze how well the evolutionary distance between studied species correlates with the distance between TS and DHFR genes in the bacterial genomes. TS and DHFR exist in some protozoans as bifunctional enzymes. Although these enzymes are expressed, in prokaryotes, by two distinct genes, the fact that some species present these genes located contiguously in the genome suggests that some evolutionary pressure or natural selection would have been applied on these genes to come closer in the genome. Bacterial species studied in this study could be separated into two distinct groups: Those which have DHFR and TS as neighbor genes in the genome and those in which these genes are not neighboring, presenting at least one other expressed gene in between them.

Figure 3 shows that no significant correlation was found between DHFR/TS gene distances and JTT distance from *P. syringae* calculated using DHFR (closed circles) or TS (open symbols) sequences. This indicates that there is no significant correlation between TS/DHFR gene distances and the evolutionary distance from *P. syringae* calculated using either DHFR ($R = -0.12$) or TS ($R = 0.01$) sequences.

CONCLUSION

Results of this study suggest that bacteria may not suffer evolutionary pressure to get TS and DHFR genes closer in genome and the optimization of performance of folate metabolism shall not be important for

evolution in such species. Expression of DHFR and TS as bifunctional enzymes in some protozoans is associated with an increase in the efficiency of folate metabolism in such organisms. This increased efficiency is reported to be important to protozoan life cycle and pathogenesis. Folate metabolism is a target for several antibiotics commonly used in clinic and may be an important factor of selection and evolution, mainly in pathogenic bacteria. Lack of significant correlation between JTT distance from *P. syringae* and the genomic distance between DHFR and TS genes, in bacteria, strongly suggests that the coexpression of DHFR and TS in these organisms may not be an important factor for the evolutionary success.

ACKNOWLEDGMENT

This study was supported by Fundacao de Amparo a Pesquisa do Estado da Bahia (FAPESB).

REFERENCES

- Aravind, L., 2000. Guilt by association: Contextual information in genome analysis. *Genome Res.*, 10: 1074-1077. DOI: 10.1101/gr.10.8.1074
- Atreya, C.E. and K.S. Anderson, 2004. Kinetic characterization of bifunctional thymidylate synthase-dihydrofolate reductase (TS-DHFR) from *Cryptosporidium hominis*. *J. Biol. Chem.*, 18: 18314-18322. DOI: 10.1074/jbc.M400009200
- Dandekar, T., B. Snel, M. Huynen and P. Bork, 1998. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.*, 23: 324-328. DOI: 10.1016/S0968-0004(98)01274-2
- Felsenstein, J., 2005. Phylogeny Inference Package (PHYLIP) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
<http://evolution.genetics.washington.edu/phylip.html>
- Higgins, D., J. Thompson, T. Gibson, J.D. Thompson and D.G. Higgins, 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22: 4673-4680.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC308517/>
- Huynen, M.A. and P. Bork, 1998. Measuring genome evolution. *Proc. Natl. Acad. Sci. USA.*, 95: 5849-5856.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC34486/>
- Lathe, W.C., B. Snel and P. Bork, 2000. Gene context conservation of a higher order than operons. *Trends Biochem. Sci.*, 25: 474-479. DOI: 10.1016/S0968-0004(00)01663-7

- Miles, E.W., S. Rhee and D.R. Davies, 1999. The molecular basis of substrate channeling. *J. Biol. Chem.*, 18: 12193-12196. DOI: 10.1074/jbc.274.18.12193
- Rogozin, I.B., K.S. Makarova, Y.I. Wolf and E.V. Koonin, 2004. Computational approaches for the analysis of gene neighborhoods in prokaryotic genomes. *Brief Bioinform.*, 2: 131-149. <http://bib.oxfordjournals.org/cgi/reprint/5/2/131>
- Tamames, J., 2001. Evolution of gene order conservation in prokaryotes. *Genome Biol.*, 6: 1-11. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC33396/>
- Tamames, J., G. Casari, C. Ouzounis and A. Valencia, 1997. Conserved clusters of functionally related genes in two bacterial genomes. *J. Mol. Evol.*, 44: 66-73. DOI: 10.1007/PL00006122
- Trujillo, M., R. Duncan and D.V. Santi, 1997. Construction of a homodimeric dihydrofolate reductase-thymidylate synthase bifunctional enzyme. *Protein Eng.*, 5: 567-573. <http://peds.oxfordjournals.org/cgi/reprint/10/5/567>