Synonymous Codon Usage in Bacteria

Maria D. Ermolaeva

The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

Abstract

In most bacteria, synonymous codons are not used with equal frequencies. Different factors have been proposed to contribute to codon usage preference, including translational selection, GC composition, strand-specific mutational bias, amino acid conservation, protein hydropathy, transcriptional selection and even RNA stability. The review discusses these factors and their contribution to bias in synonymous codon usage in bacterial genomes.

Introduction

Most amino acids can be coded by more than one triplet of nucleotides (codon). Such codons are called synonymous and usually differ by one nucleotide in the third codon position (or for some amino acids, in the second position). Synonymous codons are not used with equal frequencies; their usage is different in different genomes and even within the same genome (Grantham *et al.*, 1980; Grantham *et al.*, 1980; Gouy and Gautier, 1982; Ikemura, 1985). Bias in codon usage is an essential feature of most genomes, both bacterial and eukaryotic. One important use of the different patterns in codon usage for different bacterial genomes is to detect and date lateral gene transfer events. At the time of transfer the introduced genes have codon usage that is typical for the donor genome. With time, however, the codon usage evolves to match the codon usage of the host genome. This process of codon usage adjustment is called "amelioration" (Lawrence and Ochman, 1997; Lawrence and Ochman, 1998) and it is sometimes possible to estimate its rate and predict the age of lateral gene transfer events. Lawrence and Ochman (Lawrence and Ochman, 1997; Lawrence and Ochman, 1998) analyzed the *E.coli* genome and found that 755 of 4288 genes have been introduced into this genome, in at least 234 lateral transfer events since this species diverged from the *Salmonella* lineage 100 Myr ago. Other results (Kunst *et al*, 1997; Mosler, 1998) (see (Moszer *et al*, 1999) for review) indicate that about 13% of the *Bacillus subtilis* genes have different codon usage pattern and are likely to have foreign origin. Laterally transferred genes have also been predicted for *Thermotoga maritima* (Nelson *et al.*, 1999) and other genomes.

Different factors have been proposed to contribute to codon usage bias. In this review we discuss three factors that are well documented to shape codon usage bias:

*For correspondence. Email mariae@tigr.org; Tel. 301-610-5942; Fax. 301-838-0208.

translational selection for efficiency, CG composition, strand-specific mutational bias. We also briefly discuss other proposed factors (selection for accuracy of translation, protein hydropathy, transcriptional selection and RNA stability) that are not so extensively studied.

Selection for Translational Efficiency

One of the factors shaping codon usage patterns is the proportion of isoaccepting tRNAs. Relative abundance of tRNA species varies in different genomes and probably even within the same genome during different stages of growth (Dong, 1996). Codon preference is correlated with the composition of the tRNA pool (Ikemura, 1981; Ikemura 1985) although it is not clear whether the pattern in codon usage is a cause or a consequence of the bias in tRNA abundance.

It has been shown experimentally that mRNA consisting of preferred codons is translated faster than mRNA artificially modified to contain rare codons (Robinson *et al.*, 1984; Sorensen *et al.*, 1989). Xia (1998) suggested a mathematical model predicting that the speed of protein synthesis is optimal if the concentration of different tRNA species matches the square-root of the frequency of their corresponding synonymous codons.

The degree of codon usage bias is also shown to depend on the level of gene expression, with highly expressed genes exhibiting greater codon bias than infrequently expressed genes (Gouy and Gautier, 1982; Ikemura, 1985). This correlation can be used to predict highly expressed genes. The phenomena was long known for *E. coli*, but for most other genomes the studies are more preliminary, and sometimes they yield rather contradictory predictions. Different computational approaches that are used to detect codon bias can be a potential source of contradictory results, and we will first discuss the methods and then the predictions obtained using these methods.

CAI (Codon Adaptation Index)

Sharp and Li (1987b) proposed the codon adaptation index (CAI) as a measure of synonymous codon bias, using as a basis the codon preference of genes for highly expressed proteins, such as ribosomal proteins and elongation factors. The CAI value for a gene is a measure of its use of optimal codons, which are the codons commonly used by highly expressed proteins in a given genome.

The codon adaptation index is usually calculated as

$$CAI = \exp\left(\frac{\sum_i \sum_j n_{ij} x_{ij}}{\sum_i \sum_j n_{ij}}\right) \qquad (1)$$

where i denotes amino acid and j denotes codon; n_{ij} is the number of occurrences of the jth codon for the ith amino acid in the given protein; and x_{ij} reflects the relative usage of the jth codon in a reference set of highly expressed

genes. The more the codon usage of a given protein resembles the codon usage of the reference set, the higher its CAI value. x_{ij} can be calculated in the following way:

$$x_{ij} = \frac{y_{ij}}{y_{i\,max}} \qquad (2)$$

where y_{ij} is the number of occurrences of the jth codon for the ith amino acid in the reference set and $y_{i\,max}$ is the value of y_{ij} for the most frequently used codon for the ith amino acid.

CAI is used together with other methods to detect laterally transferred genes: such genes usually have strong codon usage bias (which may be detected by χ^2 analysis or other methods), but low values of CAI.

The CAI is a relatively sensitive indicator and is widely used, but it has a number of disadvantages. First, CAI can potentially reflect features other than gene expression level. For example, ribosomal genes are more often located on the leading DNA strand than on the lagging strand (McLean et al., 1998) and therefore CAI may be higher for those genes located on the leading strand if codon bias between leading and lagging strand is significant. Second, it is not clear whether ribosomal proteins are over-expressed in all bacteria. Third, even if all genes would be expressed equally there are always random fluctuations in codon usage that make values of CAI higher for some genes, and the method does not give any guidance as to how to distinguish random fluctuation from real trends.

Codon Bias between Gene Classes

A slight twist on codon bias analysis is the idea of looking at codon bias between gene classes, which was introduced by Karlin and Mrazek (Karlin et al. 1998; Karlin and Mrazek, 2000) and is extensively used by these authors to characterize codon bias in different bacteria. The method is based on comparing codon biases between a given gene and three gene families that are assumed to be over-expressed: ribosomal genes, chaperone-degradation genes and transcription processing factors. This method is similar to CAI and has the same disadvantages.

Correspondence Analysis

Correspondence analysis (Greenacre, 1984) maps the genes in a genome into a 59-dimensional space according to usage of the each of 59 possible codons (codons with a one-to-one mapping to amino acids are ignored). Then it identifies the major trends in codon usage as those axes through the multidimensional hyperspace that account for the largest fractions of the variation among genes. The method does not depend on codon usage preference in ribosomal genes and is widely used, but it seems to have lower sensitivity than the other methods (see Table 1) and it will not find over-expressed genes if their number is small.

χ^2 Analysis

χ^2-based analysis (Shields et al., 1988; Akashi and Shaeffer, 1997; Wernegreen and Moran, 1999) is a statistical method that compares observed codon frequencies to those expected if codon usage reflects only local base composition at synonymous sites. The main advantage of this method is that it is able to distinguish real trends in codon usage from random fluctuations. This method is accurate, sensitive and does not require prior

Table 1. Applying different computational methods to find genes with higher than average codon bias (presumably highly expressed genes). "Yes" notes that the genes were detected. "No" indicates that no such genes were found.

Genome	Method			
	CAI	Correspondence Analysis	Codon Bias Between Gene Classes	χ^2 Analysis
A. aeolicus, A. fulgidus, B. subtilis, C. pneumoniae, D. radiodurans, M. thermoautotrophicum, R. prowazekii			Yes (Karlin and Mrazek, 2000; Karlin and Mrazek, 2001)	
B. burgdorferi, R. prowazekii		No (Lafay et al., 1999; McInerney, 1998; Anderson and Sharp, 1996a)		
Buchnera				Only one gene predicted to be possibly overexpressed (Wernegreen and Moran, 1999)
C. trachomatis		Yes (Romero et al., 2000)	Yes (Karlin and Mrazek, 2001)	
E. coli,	Yes (Sharp and Li, 1987b; Nakamura and Tabata, 1997)		Yes (Karlin and Mrazek, 2001)	Yes (Wernegreen and Moran, 1999)
H. influenzae, M. pneumoniae	Yes (Pan et al., 1998; Nakamura and Tabata, 1997)			
H. pylori		No (Lafay et al., 2000)	Yes (Karlin and Mrazek, 2001)	
M. jannaschii, Synechocystis	Yes (Nakamura and Tabata, 1997)		Yes (Karlin and Mrazek, 2000; Mrazek et al., 2001)	
M. genitalium	No (Nakamura and Tabata, 1997)			
M. tuberculosis	Yes (Pan et al., 1998)	Yes (Anderson and Sharp, 1996a)	Yes (Karlin and Mrazek, 2000)	
T. pallidum	Yes (Fraser et al., 1998)	No (Lafay et al., 1999)	Yes (Karlin and Mrazek, 2001)	
Lactobacillus species		Yes (Pouwels and Leunissen, 1994)		

knowledge of preferred codons, but it is not widely used, perhaps due to its complexity.

Results

Table 1 shows a list of bacterial genomes for which correlation of codon usage bias with the level of gene expression has been studied and computational methods used. "Yes" or "No" in the table indicates whether or not any correlation was detected.

One can see that correspondence analysis often does not find a correlation between gene expression level and codon usage; this result can probably be attributed to the lower sensitivity of the method. On the other hand, it is not clear whether all results obtained by the CAI method and its modification (Codon bias between classes) are reliable: some of them may just reflect random fluctuations or difference in nucleotide composition related to particular chromosomal region (see the next two sections about GC composition and strand specific mutational bias). The χ^2-based method does not have the disadvantages mentioned above, but it was applied only to two genomes: *E. coli* and *Buchnera*. For *E. coli*, χ^2 analysis gave results similar to ones obtained using the CAI method: the set of genes predicted to be over-expressed is about the same (Wernegreen and Moran, 1999). In Buchnera almost no genes were detected with codon usage statistically different from the overall codon composition. These results are consistent with *Buchnera* biology: Buchnera is a bacterial endosymbiont of aphids, and it undergoes severe population bottlenecks during maternal transmission through its hosts. Population genetic models indicate that effective population size influences the balance between the effects of any mutational bias and selection for optimal codons (Li, 1987). In particular, translational selection is effective only if the size of the population multiplied by selective coefficient (which characterizes the strength of the selective force) is higher than two.

Thus, despite differences in results for some particular species or genes it is already well-established that for many bacteria there is a positive selection for "popular" codons, especially in highly expressed genes. Does this principle work in the opposite direction; i.e., are "rare" codons also used to regulate the rate of translation? Earlier papers (Konigsberg and Godson, 1983) reported that some *E. coli* genes have an excess of rare codons, which can be explained as a mechanism to maintain low expression. However, later work (Sharp and Li, 1986) showed that the frequency of low codons is not significant and most likely results from the lack of negative selection rather than from the presence of positive selection.

GC Composition

The most obvious factor that determines codon usage is mutational bias that shapes genome GC composition. This factor is most significant in genomes with extreme base composition: species with high GC content use more G- and C-ending codons than species with low GC content. Mutational bias is responsible not only for intergenetic difference in codon usage but also for codon usage bias within the same genome. This can be illustrated with the *M. genitalium* genome (Fraser *et al.*, 1995), which has a very distinctive variation in GC content (Kerr *et al.*, 1997) within the genome (Figure 1). Figure 2 shows codon usage patterns in this genome. Here we used a very simplified method to detect codon usage bias: for each amino acid, we selected the codon that is most frequently used in the whole genome; and then for each gene we calculated the proportion of amino acids that are coded by those frequently used codons (axis Y). Each circle on the plot denotes one gene. The X axis shows the physical position of the gene within the genome, and higher values on the Y axis indicate that the bias in codon usage for a given gene resembles the bias in overall codon usage. Filled circles correspond to genes on "+" strand and open circles denote genes on
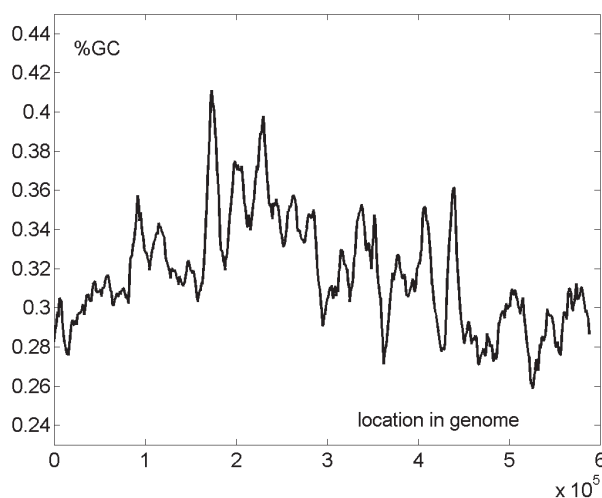

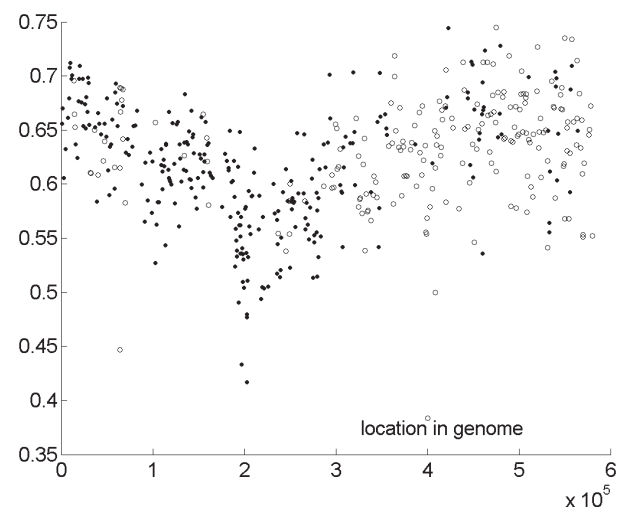Figure 1. GC composition of *M.genitalium* genome. Window size 10000 nucleotides and step size 1000 nucleotides.

Figure 2. *M. genitalium.* Filled circles shows genes on "+" strand and open circles denote "-" strand. Axis X shows physical location of the gene in the genome and axis Y indicates portion of codons in this gene that are most used in the genome as a whole (see more detailed explanation in the text).
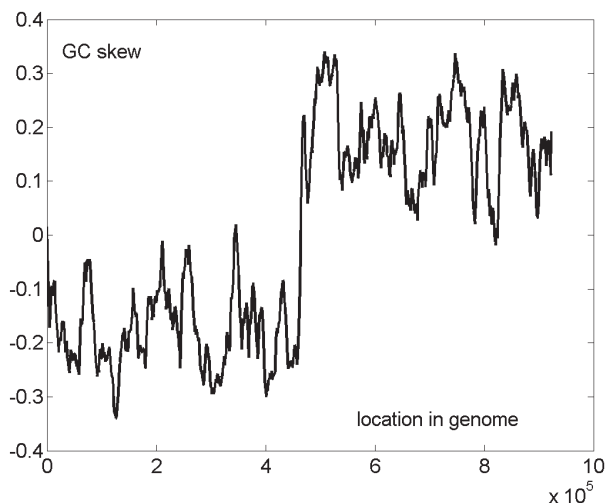
Figure 3. GC skew for the main chromosome of the B. burgoferri genome. GC skew was calculated as (G-C)/(G+C) with window size 10000 nucleotides and step size 1000. Switch in GC skew sign corresponds to the origin of replication.

the "-" DNA strand. Comparison between Figure1 and Figure 2 shows that codon usage in *M. genitalium* is correlated with GC composition.

Strand-Specific Mutational Bias

Nucleotide composition of leading and lagging DNA strands is usually different due to asymmetry in biochemical processes such as DNA replication and repair (Sueoka 1962). This creates a variation in synonymous codon usage between leading and lagging strands.

The distribution of different oligomers is skewed between leading and lagging DNA strands (Salzberg *et al.*, 1998), but the trend is most obvious when considering
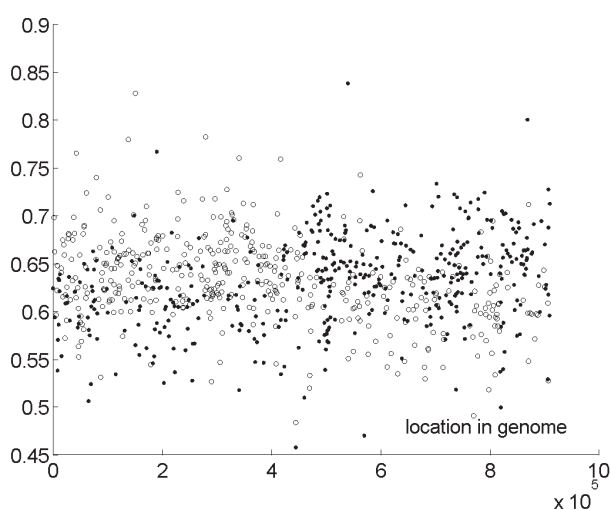

Figure 4. *B. burgdorferi*, main chromosome. Filled circles shows genes on "+" strand and open circles denote "-" strand. Axis X shows physical location of the gene in the chromosome and axis Y indicates portion of codons in this gene that are most used in the genome as a whole (see more detailed explanation in the text).

a difference in the number of Gs and Cs along a single DNA strand. This difference is usually measured as GC skew or (G-C)/(G+C). AT skew is defined similarly and also varies between leading and lagging strand. Figure 3 shows GC skew for the *B. burgdorferi* genome (Fraser *et al.*, 1997). The switch in the sign of GC skew corresponds to the origin of replication. To the left of this point, the minus strand is the leading strand; to the right, the plus strand is the leading strand.

A comparison between Figure 3 and Figure 4 illustrates that codon usage in *B. burgdorferi* correlates with GC skew. Codon usage was calculated in the same way as for *M.genitalium* (see the previous section). GC skew varies for different bacteria, with the highest (up to 30%) for *B. burgdorferi* and *T. pallidum,* and consequently the contribution of strand-specific mutational bias is also different. The skews can be also partly caused by a tendency of genes to be transcribed in the same direction that they are replicated, i.e. on the leading strand. This component of skew presumably does not affect synonymous codon usage, for example in some bacteria (*M. genitalium*, *M.pneumoniae* and *S. subtilis*) the skew for total genomic DNA is opposite to that for codon position 3 (McLean, 1998).

There are also other trends in synonymous codon usage that may be related to differential DNA repair, for example genes close to the origin of replication tend to use G- and C-ending codons whereas those close to the terminus use A- and T-ending codons. (Sharp 1991, Deschavanne and Filipski, 1995).

Other Factors that May Influence Codon Usage

Accuracy of Translation

Another hypothesis suggests that there is also selection for translational accuracy (Bulmer 1991). Such selection as well as selection for translational speed favors codons that correspond to the most abundant tRNA species, because this minimizes number of incorrect tRNAs that have to be rejected before the arrival of the right tRNA. It is difficult to completely separate the effect of maximizing translational accuracy on codon usage bias from that of maximizing translational efficiency, because highly expressed proteins are also more conserved. In addition, both the rate of synonymous substitution and the degree of codon usage bias reflect the intensity of selection at the translational level (Sharp and Li, 1987a). The main argument in favor of translational accuracy selection is that the level of synonymous codon bias is positively correlated to gene length (Eyre-Walker, 1996a; Hooper and Berg, 2000): since the cost of producing a protein in terms of the energy used and the resources consumed is proportional to gene length, selection to avoid errors should be stronger in longer genes. This correlation however is rather weak, especially as reported in (Hooper and Berg, 2000). In these analyses the very start and end of the genes were not taken into account because they have lower codon bias, due to constraints imposed by overlapping genes and signal sequences (Bulmer, 1988; Chen and Innouye, 1990; Eyre-Walker and Bulmer 1993; Eyre-Walker 1996b).

There is also an argument against translational accuracy selection in bacteria. If the effect of maximizing translational accuracy exists, then within-gene codon usage should be correlated with variation of amino acid conservation: gene regions of greater amino acid conservation should exhibit more dramatic codon usage bias than do regions of lower amino acid conservation. This correlation was shown for eukaryotes (Akashi, 1994), but was not observed in *E. coli* and *S. typhimurium* (Hartil *et al.*, 1994). From all the above, we can conclude that to date it has not been unambiguously shown that translational accuracy plays a role in bacterial evolution, although we cannot yet completely exclude this possibility.

Protein Hydropathy

A recent study (Romero *et al.*, 2000) reports that there is a correlation between protein hydropathy and codon usage pattern in *Chlamydia trachomatis*. The source of this phenomenon is not clear. One possible explanation is that there is a positive correlation between the frequency of G and C nucleotides in the 3rd codon positions and the hydropathy of the corresponding proteins (D'Onofrio *et al.*, 1999). A second explanation, suggested by Romero *et al.*, is related to the speed of protein folding. It is possible that the process of folding is slower for hydrophobic proteins and therefore they sometimes prefer 'non-optimal' codons. This hypothesis can be only verified by analyzing at least a few other bacterial genomes, which has not yet been done.

Transcriptional Selection and RNA Stability

Some authors suggest that bias in codon usage can be also caused by selection for transcription effectiveness and RNA stability. Xia (1996) reports that genes from organisms with a higher metabolic rate use more A-ending codons (presumably because in such organisms adenine is more abundant). Another hypothesis (Lao and Forsdyke, 2000) suggests that evolutionary pressure favors purine-rich (i.e., G- and A-rich) codons because this prevents distracting RNA-RNA interactions (the "RNA politeness" hypothesis). Codon usage may also define RNA secondary structure, which results in different RNA stability (Cohen B. and Skiena S., submitted).

Conclusion

Three factors (GC composition, GC (AT) skew and translational selection for efficiency) were shown to play a role in shaping codon bias. The relative contribution of these factors varies for different genomes and depends on genome composition and biology. For example, GC skew is very strong in the *B. subtilus* and *T. pallidum* genomes, and the correlation between GC skew and codon usage bias is easily visible in Figures 3 and 4. For other genomes sequenced until now, the correlation is more subtle. Codon usage in genomes with extreme composition (very high or very low GC content) is mostly shaped by mutational bias. Effect of translational selection may depend on genome biology. First, this effect varies with effective population size, and second, it may depend on how often the organism goes through the exponential growth phase when the speed of translation is especially important.

A few other factors were proposed to contribute to synonymous codon usage bias, but they are less well-studied. It is possible that these factors have less contribution, or that their effects apply only to a few genomes. For example, even if the "RNA politeness" hypothesis is correct, it clearly does not apply to all genomes. This hypothesis implies that selectional pressure should increase the purine load in the 3rd codon position; however, some genomes have GC and AT skews in the 3rd codon position that are opposite to those in the 1st and 2nd position (McLean *et al.*, 1998), which means that the third codon base actually decreases purine load.

Acknowledgements

References

Akashi, H. 1994. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. Genetics 136: 927-935.

Akashi, H., and Schaeffer, S.W. 1997. Natural selection and the frequency distributions of "silent" DNA polymorphism in *Drosophila*. Genetics 146: 295-307.

Andersson, S.G.E. and Sharp, P.M. 1996a. Codon usage and base composition in *Rickettsia prowazekii*. J. Mol. Evol., 42: 525-536.

Andersson, S.G.E. and Sharp, P.M. 1996b. Codon usage in the *Mycobacterium tuberculosis* complex. Microbiology, 142: 915-925

Bulmer, M. 1988. Codon usage and intragenic position. J. Theor. Biol. 133: 67-71.

Bulmer, M. 1991. The selection mutation drift theory of synonymous codon usage. Genetics. 129: 897-907.

Chen, G.-F.T. and Innouye M. 1990. Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of *Escherichia coli*. Nucleic Acids Res. 18: 1465-1473.

Deschavanne, P. and Filipski, J. 1995. Correlation of GC content with replication timing and repair mechanisms in weakly expressed *E. coli* genes. Nucleic Acids Res. 23:1350-1353.

Dong, H., Nilsson, L., and Kurland, C.G. 1996. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rate. J. Mol. Biol. 260: 649-663.

Eyre-Walker, A. and Bulmer, M. 1993. Reduced synonymous substitution rate at the start of enterobacterial genes. Nucleic Acids Res. 21: 4599-4603.

Eyre-Walker, A. 1996a. Synonymous codon bias is related to gene length in Escherichia coli: selection for translational accuracy? Mol. Biol. Evol. 13: 864-872.

Eyre-Walker, A. 1996b. The close proximity of Escherichia coli genes: implications for stop codon and synonymous

codon usage. J. Mol. Evol. 42: 73-78.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.-F., Dougherty, B.A., Bott, K.F., Hu, P.-C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, III, C.A., and Venter, J.C. 1995. The minimum Sgene complement of mycoplasma genitalium. Science. 270: 397-408.

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., Gwinn, M., Dougherty, B., Tomb, J.-F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S.,Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J., Weidman, J., Utterback, T., Watthey, L., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K.,Hatch, B., Smith, H.O., and Venter, J.C. 1997. Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. Nature. 390: 580-586

Fraser, C.M., Norris, S.J., Weinstock, G.M., White, O., Sutton, G., Clayton, R., Dodson, R., Gwinn, M., Hickey, E., Ketchum, K.A, Sodergren, E., Hardham, J, McLeod, M., Salzberg, S., Khalak, H., Weidman, J., Howell, J.K., Chidambaram, M., Utterback, T., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O., and Venter, J.C. 1998. Complete genomic sequence of *Treponema pallidum,* the syphilis spirochete. Science. 281: 375-388.

Gouy, M. and Gautier, C. 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 10: 7055-7064.

Grantham, R.C., Gautier, C., Gouy, M., Mercier, R., and Pave A. 1980. Codon catalog usage and the genome hypothesis. Nucleic Acids Res. 8: 49-79.

Grantham, R.C., Gautier, C., Gouy, M., Jacobzone, M., and Mercier, R. 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. 9: 43-79.

Greenacre, M.J. (1984) Theory and applications of correspopondence analysis. London: Academic Press.

Hartil, D.L., Moriyama, E.N., and Sawyer, S.A. 1994. Selection intensity for codon bias. Genetics 1138: 227-234.

Hooper, S.D. and Berg, O.G. 2000. Gradients in nucleotide and codon usage along *Escherichia coli* genes. Nucleic Acids Res. 28: 3517-3523.

Ikemura, T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. 151: 389-409.

Ikemura, T. 1985 Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2:13-34.

Karlin, S., Mrazek, J., and Campbell, A.M. 1998. Codon usages in different gene classes of the *Escherichia coli* genome. Mol. Microbiol. 29: 1341-1355.

Karlin, S. and Mrazek, J. 2000. Predicted highly expressed genes of diverse prokaryotic genomes. J. Bacteriol. 182: 5238-5250.

Karlin, S. and Mrazek, J. 2001. Predicted highly expressed and putative alien genes of Deinococcus radiodurans and implications for resistance to ionizing radiation damage. Proc. Natl. Acad. Sci. USA 98: 5240-5945.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Danchin, A, *et al.*1997. The complete genome sequence of the Gram-positive bacterium Bacillus subtilis. Nature. 390: 249-256.

Kerr, A.R., Peden, J.F., Sharp, P.M. 1997. Systematic base composition variation around the genome of *Mycoplasma genitalium*, but not *Mycoplasma pneumoniae*. Mol. Microbiol. 25: 1177-1179.

Konigsberg, W. and Godson, G.N. 1983. Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 80:687-691.

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

Lafay, B., Atherton, J.C., and Sharp, P.M. 2000. Absence of translationally selected synonymous codon usage bias in Helicobacter pylori. Microbiology. 146: 851-860.

Lawrence, J.F. and Ochman, H. 1997. Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. 44: 383-397.

Lawrence, J.F. and Ochman, H. 1998. Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA 95: 9413-9417.

Lao, P.J. and Forsdyke, D.R. 2000. Thermophilic bacteria strictly obey Szybalski's transcription direction rule and politely purine-load RNAs with both adenine and guanine. Genome Research. 10: 228-236.

Li, W.-H. 1987. Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol. 24: 337-345.

McInerney, J.O. 1998. Replicational and transcriptional selection on codon usage in *Borrelia burgorferi*. Proc. Natl. Acad. Sci. USA 95: 10698-10703.

McLean, M.J., Wolfe, K.H., and Devine, K.M. 1998. Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes.

Moszler, I. 1998. The complete genome of Bacillus subtilis: from sequence annotation to data management and analysis. FEBS Lett, 430: 28-36.

Moszler, I., Rocha, E. PC, Danchin, A. 1999. Codon usage and lateral gene transfer in *Bacillus subtilis*. Current Opinion in Microbiology. 2: 524-528.

Mrazek, J., Bhaya D., Grossman, A.R. and Karlin S. 2001. Highly expressed and alien genes of the *Synechocystis* genome. Nucleic Acids Res. 29: 1590-1601.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., Mcdonald, L., Utterback,

T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.R., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M. 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima. Nature 399, 323 - 329 (1999).

D'Onofrio, G., Jabbari, K., Musto, H., Bernardi, G. 1999. The correlation of protein hydropathy with the base composition of coding sequences. Gene. 238: 3-14.

Pan, A., Dutta, C., and Das, J. 1998. Codon usage in highly expressed genes of *Haemophillus influenzae* and *Mycobacterium tuberculosis*: translational selection versus mutational bias. Gene. 215: 405-413.

Pouwels, P.H. and Leunissen, A.M. 1994. Divergence in codon usage of Lactobacillus species. Nucleic Acids Res. 22: 929-936.

Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M., Humphreys, G. 1984. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. Nucleic Acids Res. 12: 6663-6671.

Romero, H., Zavala, A., and Musto, H. 2000. Codon usage in Chlamidia trachomatis is the result of strand-specific mutational biases and a complex pattern of selective forces. Nucleic Acids Res. 28: 2084-2090.

Salzberg, S.L., Salzberg, A.J., Kerlavage, A.R., Tomb, J.-F. 1998. Skewed oligomers and origins of replication. Gene. 217: 57-67.

Sharp, P.M. and Li, W.H. 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for "rare" codons. Nucleic Acids Res. 14:7737-7749.

Sharp P.M. and Li, W.H. 1987a. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. 4: 222-230.

Sharp P.M. and Li W.H. 1987b. The codon usage adaptation index – a mesure of directional synonymous codon usage bias, and its potential application. Nucleic Acids Res. 15: 1281-1295.

Sharp, P.M. 1991. Determinants of DNA sequence divergence between Escherichia coli and Salmonella typhimurium: codon usage, map position, and concerted evolution. J. Mol. Evol. 33: 23-33.

Shields, D.C., Sharp, P.M., Higgins, D.G., and Wright, F. 1988. "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. 5: 704-716.

Sorensen, M.A., Kurland, C.G., and Pedersen, S. 1989. Codon usage determines translation rate in *Escherichia coli*. J. Mol. Biol. 207: 365-377.

Sueoka, N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition. Proc. Natl. Acad. Sci USA 48: 1141-1149.

Wernegreen, J.J. and Moran, N.A. 1999. Evidence for genetic drift in endosymbionts (Buchnera): analyses of protein-coding genes. Mol. Biol. Evol. 16: 83-97.

Xia, X. 1996. Maximizing transcription efficiency causes codon usage bias. Genetics. 144: 1309-1320.

Xia, X. 1998. How optimized is the translational machinery in *Escherichia coli*, *Salmonella typhimurium* and *Saccharomyces cerevisiae*? Genetics. 149: 37-44.

Further Reading

- **MALDI-TOF Mass Spectrometry in Microbiology**
  Edited by: M Kostrzewa, S Schubert (2016)
  www.caister.com/malditof

- *Aspergillus* and *Penicillium* in the Post-genomic Era
  Edited by: RP Vries, IB Gelber, MR Andersen (2016)
  www.caister.com/aspergillus2

- **The Bacteriocins: Current Knowledge and Future Prospects**
  Edited by: RL Dorit, SM Roy, MA Riley (2016)
  www.caister.com/bacteriocins

- **Omics in Plant Disease Resistance**
  Edited by: V Bhadauria (2016)
  www.caister.com/opdr

- **Acidophiles: Life in Extremely Acidic Environments**
  Edited by: R Quatrini, DB Johnson (2016)
  www.caister.com/acidophiles

- **Climate Change and Microbial Ecology: Current Research and Future Trends**
  Edited by: J Marxsen (2016)
  www.caister.com/climate

- **Biofilms in Bioremediation: Current Research and Emerging Technologies**
  Edited by: G Lear (2016)
  www.caister.com/biorem

- **Microalgae: Current Research and Applications**
  Edited by: MN Tsaloglou (2016)
  www.caister.com/microalgae

- **Gas Plasma Sterilization in Microbiology: Theory, Applications, Pitfalls and New Perspectives**
  Edited by: H Shintani, A Sakudo (2016)
  www.caister.com/gasplasma

- **Virus Evolution: Current Research and Future Directions**
  Edited by: SC Weaver, M Denison, M Roossinck, et al. (2016)
  www.caister.com/virusevol

- **Arboviruses: Molecular Biology, Evolution and Control**
  Edited by: N Vasilakis, DJ Gubler (2016)
  www.caister.com/arbo

- *Shigella*: **Molecular and Cellular Biology**
  Edited by: WD Picking, WL Picking (2016)
  www.caister.com/shigella

- **Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment**
  Edited by: AM Romaní, H Guasch, MD Balaguer (2016)
  www.caister.com/aquaticbiofilms

- **Alphaviruses: Current Biology**
  Edited by: S Mahalingam, L Herrero, B Herring (2016)
  www.caister.com/alpha

- **Thermophilic Microorganisms**
  Edited by: F Li (2015)
  www.caister.com/thermophile
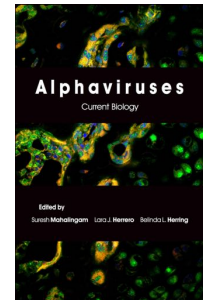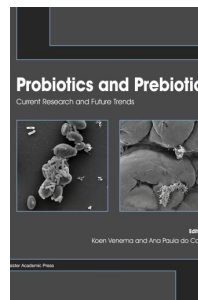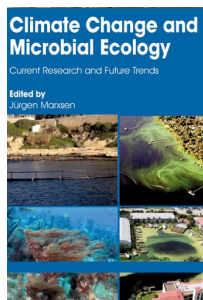
- **Flow Cytometry in Microbiology: Technology and Applications**
  Edited by: MG Wilkinson (2015)
  www.caister.com/flow

- **Probiotics and Prebiotics: Current Research and Future Trends**
  Edited by: K Venema, AP Carmo (2015)
  www.caister.com/probiotics

- **Epigenetics: Current Research and Emerging Trends**
  Edited by: BP Chadwick (2015)
  www.caister.com/epigenetics2015

- *Corynebacterium glutamicum*: **From Systems Biology to Biotechnological Applications**
  Edited by: A Burkovski (2015)
  www.caister.com/cory2

- **Advanced Vaccine Research Methods for the Decade of Vaccines**
  Edited by: F Bagnoli, R Rappuoli (2015)
  www.caister.com/vaccines

- **Antifungals: From Genomics to Resistance and the Development of Novel Agents**
  Edited by: AT Coste, P Vandeputte (2015)
  www.caister.com/antifungals

- **Bacteria-Plant Interactions: Advanced Research and Future Trends**
  Edited by: J Murillo, BA Vinatzer, RW Jackson, et al. (2015)
  www.caister.com/bacteria-plant

- *Aeromonas*
  Edited by: J Graf (2015)
  www.caister.com/aeromonas

- **Antibiotics: Current Innovations and Future Trends**
  Edited by: S Sánchez, AL Demain (2015)
  www.caister.com/antibiotics

- *Leishmania*: **Current Biology and Control**
  Edited by: S Adak, R Datta (2015)
  www.caister.com/leish2

- *Acanthamoeba*: **Biology and Pathogenesis (2nd edition)**
  Author: NA Khan (2015)
  www.caister.com/acanthamoeba2

- **Microarrays: Current Technology, Innovations and Applications**
  Edited by: Z He (2014)
  www.caister.com/microarrays2

- **Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications**
  Edited by: D Marco (2014)
  www.caister.com/n2

Order from caister.com/order