Co-evolution between codon usage and Protein-Protein Interaction Networks in Bacterial genomes

Maddalena Dilucca^{1*}, Giulio Cimini^{2,3,*}, Sergio Forcelloni¹ and Andrea Giansanti^{1,4}

¹ Dipartimento di Fisica, Sapienza University of Rome, Rome, Italy; maddalena.dilucca@gmail.com

² Dipartimento di Fisica, Tor Vergata University of Rome, Italy; giulio.cimini@gmail.com

³ Istituto dei Sistemi Complessi (CNR) UoS

⁴ INFN Roma1 unit, Rome, Italy; andrea.giansanti.roma1.infn.it

* Correspondence: maddalena.dilucca@gmail.com; giulio.cimini@gmail.com

Received: date; Accepted: date; Published: date

Abstract: In this work, we study the correlation between codon usage and the network features of the PPI in bacteria genomes. We want to extend the information by Dilucca et al. (2015) about E.Coli's genome for a set of other 34 bacteria. We use PCA techniques in the space of codon bias indices (compAI, compAI_w, tAI, NC) and GC content to show that genes with similar patterns of codon usage feature have a significantly higher probability that their encoded proteins interact within the PPI. And vice-versa, we show that interacting in the PPI have a coherent codon usage. This work could allow for future investigations into the possible effects that codon bias signal can have on the topology of protein interaction network and, as such, to improve existing bioinformatics methods for predicting protein interactions.

Keywords: codon usage bias, bacteria, protein-protein interaction networks, interactomes

1. Introduction

In this work, we investigate the relationship between codon bias of coding sequences, at the genetic level [1] and connectivity patterns of protein-protein interaction networks (PPI) [2].

As it is well-known, the genetic code is degenerate: synonymous codons at the genetic level encode for the same amino acid in the translated protein. Moreover, although synonymous codons are indistinguishable in the primary structure of a protein, they are not used randomly, but with different frequencies which may vary across different species, in various parts of one specific genome, and even within different regions of the same gene. This phenomenon, known as Codon Usage Bias (CUB), is well-established in the literature (see, for example, [3-5]). Nevertheless, while remarkable observations have emerged, a general understanding of the biology of CUB still lacks, because of its complex phenomenology [7]. Although CUB does not alter the amino acid sequences of proteins, it is involved in many important cellular processes, including differential gene expression [7], translation efficiency and accuracy [8,9], dynamics of the ribosome, and co-translational folding of the proteins [10,6].

With the present study, we intend to share basic observations of sufficient generality on the co-evolution of CUB and the degree connectivity of bacterial interactomes. The systematic analysis of PPIs is crucial to understand the patterns of chemical reactions within cells, and the role played by proteins in regulative processes. Moreover, on the applicative side, the comparison of the interactomes from different species is fundamental to understand disease-related processes that engage more than one species, such as host-pathogen relationships, to identify clinically relevant host-pathogen PPI, and consequently to develop future therapeutic applications [1]. In previous work, Dilucca et al. showed that translational selection in E. Coli (taken as a case of study)

systematically favors optimal codons in proteins that have a large number of interactors and belong to the most representative communities in the PPI [3]. In the current analysis, we extend those observations to a set of unrelated bacterial species by focusing on the connection between the codon usage of the genes and the role of their corresponding proteins in the protein interaction network. It is well known that codon usage bias is strongly correlated with expression level in many organisms, and it is well structured along the genome, with near neighbor genes having similar usage frequencies of synonymous codons [13]. Accordingly, and considering that gene expression level [14] and proximity between the positions of the genes in the genome are powerful predictors of protein-protein interaction, we analyze how the closeness in codon usage bias of the genes (as measured by the normalized scalar product of RSCU vectors) is reflected into the capacity of the corresponding proteins to make physical contact in the cell. Given the above considerations, we would expect a more similar codon usage bias between interacting proteins than non-interacting ones. We find out that this statement is not valid for all bacteria considered in this analysis, especially if we measure CUB without taking into account the information about tRNA levels. This means that the distance in codon usage bias seems unable to distinguish between proteins that make contacts or not in the protein interaction network. To further investigate this issue, we performed a Principal Component Analysis (PCA) in the space of the codon bias indices (NC, tAI, compAI, GC content), thus using compAI and tAI to add the information about the tRNA pool of the organism. In all bacteria, we observed a clear localization of the most representative communities of the protein interaction networks in the space of the first two PCA components, underlining a more pronounced similarity between proteins belonging to the same community. This observation could be explained considering the fundamental requirement of interacting proteins (especially those belonging to the same community) to be present in the cell according to precise quantities at a given time to form the protein complexes that are necessary upon the ongoing cellular programs. Finally, we show that the distance between a pair of genes in the plane of the first two PCA component is a statistically robust predictor of the likelihood that their corresponding proteins do interact (physically or functionally). In agreement with our previous results, the signal is evident when codon usage frequencies of interacting proteins are far from being random.

To summarize, in line with what observed in the first work [3], our main result indicates that the functional structuring of the protein interaction network has interfered with the peculiar codon choice of the genes over evolutionary time. These results point out that codon bias evolution should be a relevant parameter in the prediction of unknown protein-protein interactions from only genomic information. Last but not least, we claim that using the codon usage bias as an additional level of information in the study of protein interaction networks could be useful to provide alternative treatments in the light of growing resistance to antibiotics and the propagation of numerous infectious agents.

2. Materials and Methods

2.1 Data sources

In this work, we select a set of 34 bacterial genomes with different behavior, environment and taxonomy (see Table 1 for details). Each bacteria represents a specific clade in the phylogenetic tree by Plata et al. [14]. Nucleotide sequences were downloaded from the FTP server of the National

86 Center for Biotechnology Information
87 (ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/old_genbank/Bacteria/) [2].

88 **Table 1.** Summary of the 34 bacterial dataset considered in this work. In table, we report the
89 organism name, abbreviation, RefSeq, STRING code, Size of genome (number of genes), and the
90 density of the protein interaction network defined as ratio between the number of links in the real
91 interactome and the maximum number of possible links ($n(n-1)/2$, where n is the number of nodes).

Organisms	Abbrev.	RefSeq	STRING code	Size	Density of PPI
Agrobacterium fabrum str. C58	agtu	NC_003062	176299	2765	0.008
Anabaena variabilis ATCC 29413	anva	NC_007413	240292	5043	0.005
Aquifex aeolicus VF5	aqae	NC_000918	224324	1497	0.009
Bifidobacterium longum NCC2705	bilo	NC_004307	216816	1726	0.004
Bordetella bronchiseptica RB50	bobr	NC_002927	257310	4994	0.005
Bordetella parapertussis 12822	bopa	NC_002928	360910	4185	0.008
Brucella melitensis bv. 1 str. 16M	brme	NC_003317	224914	2059	0.006
Buchnera aphidicola str. Bp	buap	NC_004545	224915	504	0.008
Burkholderia pseudomallei K96243	bups	NC_006350	272560	3398	0.002
Buchnera aphidicola Sg uid57913	busg	NC_004061	198804	546	0.002
Burkholderia thailandensis E264	buth	NC_007651	271848	3276	0.001
Caulobacter crescentus	cacr	NC_011916	565050	3885	0.002
Campylobacter jejuni	caje	NC_002163	192222	1572	0.004
Corynebacterium efficiens YS-314	coef	NC_004369	196164	2938	0.006
Corynebacterium glutamicum ATCC 13032	cogl	NC_003450		2959	0.005
Chlamydia trachomatis D/UW-3/CX	chtr	NC_000117.1	272561	894	0.008
Clostridium acetobutylicum ATCC 824	clac	NC_003030.1	272562	3602	0.005

Francisella novicida U112	frno	NC_008601	401614	1719	0.007
Fusobacterium nucleatum ATCC 25586	funu	NC_003454.1	190304	1983	0.002
Haemophilus ducreyi 35000HP	hadu	NC_002940	233412	1717	0.004
Klebsiella pneumoniae	klpn	NC_009648	272620	4775	0.005
Listeria monocytogenes EGD	limo	NC_003210	169963	2867	0.003
Mesorhizobium loti MAFF303099	melo	NC_002678.2	266835	6743	0.0001
Mycoplasma genitalium G37	myge	NC_000908	243273	475	0.005
Mycoplasma pneumoniae M129	mypn	NC_000912.1	272634	648	0.006
Mycobacterium tuberculosis H37Rv	mytu	NC_000962.3	83332	3936	0.006
Porphyromonas gingivalis ATCC 33277	pogi	NC_010729	431947	2089	0.001
Ralstonia solanacearum GMI1000	raso	NC_003295.1	267608	3436	0.002
Sphingomonas wittichii RW1	spwi	NC_009511	392499	4850	0.007
Staphylococcus aureus NCTC 8325	stau	NC_007795	93061	2767	0.004
Synechocystis sp. PCC 6803	syp	NC_000911.1	1148	3179	0.004
Thermotoga maritima MSB8	thma	NC_000853.1	243274	1858	0.001
Vibrio cholerae N16961	vich	NC_002505	243277	2534	0.001
Xylella fastidiosa 9a5c	xyfa	NC_002488	160492	2766	0.002

2.2 Codon Usage Bias Measures

In this work, we used four measures to quantify the extent of CUB: 1) the tRNA Adaptation index (tAI) [5], which is based on the assumptions that tRNA availability is the driving force for translational selection; 2) CompAI and CompAI_w [3], which refine tAI using the competition between cognate and near-cognate tRNA; 3) the Effective Number of Codons (NC) [18], which is a statistical measure of the number of codons used in the sequence; and 4) the GC content, namely the

percentage of guanine and cytosine in the RNA molecules. Moreover, we use the Relative Synonymous Codon usage (RSCU) to have a finer description of the non-uniform usage of synonymous codons in each coding sequence in the form of a 61-component vector.

2.3 Protein-Protein Interaction Network

PPI networks of the 34 bacterial genomes were retrieved from the STRING database (Known and Predicted Protein-Protein Interactions) [11]. Community detection in PIN was performed by using the Molecular Complex Detection (MCODE) software. In this study, we evaluate communities based on the k-core concept, i.e., the largest subgraph within which each node has at least K connections. In line with our previous study [3], we consider only the first eight communities with a degree $k > 10$.

2.4 Analyses of the RSCU vectors of genes encoding for interacting proteins in the PPI

The main advantage of RSCU is to allow quantifying the similarity between two genes in their usage of synonymous codons using the classical operations between vectors. This feature makes it particularly suitable for the study of how the similarity between two genes in terms of codon usage bias is reflected in the ability of their encoded proteins to make contact in the protein interaction network. Specifically, a vector of RSCU for each gene of a genome was constructed with a homemade Python script. We computed the normalized distribution of the scalar products between RSCU vectors for genes whose encoded proteins are in contact in the protein interaction network. We then compared this distribution with that obtained by considering a random model of the protein interaction network, in which all connections are randomly shuffled, but the number of connections of each protein is kept constant. For each pairwise comparison, we used the Mann-Whitney test for assessing if distributions are different (p-value < 0.001).

2.5 Principal Component Analysis

Principal Component Analysis (PCA) is a multivariate statistical method that transforms a set of possibly correlated variables into a set of linearly uncorrelated ones (called principal components, spanning a space of lower dimensionality) [6].

We use PCA over the space of the five codon bias indices described above. Thus, each gene is represented as a 5-dimensional vector with coordinates (compAI, compAI_w, tAI, NC, GC).

The first two principal components (PC_1 and PC_2) turn out to represent as much as 65% of the total variance of codon bias over the genomes. Thus, we focus on the plane defined by these two vectors.

We calculate this plane for each bacterial genome (see Figure 2 in Supplementary Materials), reporting the associated variances, eigenvalues and eigenvectors in Supplementary Materials. We then project the first eight communities extrapolated with MCODE in the $PC_1 - PC_2$ plane, and we calculate their centroids (with error bars defined by standard deviations) to check if these communities are well separated on this plane.

2.6 Z-score analysis.

We define d as the Euclidean distance between two genes on the $PC_1 - PC_2$ plane, and $\text{Pr}(\text{link}|d)$ as the fraction of gene pairs, among those localized within a distance d , whose encoded proteins are connected in the PPI network. We then compare this conditional probability estimated on the real

interactomes with $\Pr(< link >_{\Omega} | d)$, namely the same probability estimated on the configuration model (CM) of the networks (a degree-conserving rewiring of the network links) [22]. The CM is used as the null hypothesis that no relation exists between the codon usage of two genes and a physical interactions between the encoded proteins. The significance of $\Pr(link|d)$ with respect to the null hypothesis is quantified through the Z-score values:

$$Z_d = \frac{\Pr(link|d) - \Pr(< link >_{\Omega} | d)}{\sigma[< link >_{\Omega} | d]}$$

We computed the function of Z-score as a function of the gene distance d for each genome.

3. Results

3.1 Interacting proteins do not share a common codon usage statistics

We analyze how the closeness in codon usage bias of two genes (as measured by the normalized scalar product of their RSCU vectors) is reflected in the capacity of their proteins to make physical contact in the cell. According to what stated in the Introduction section, we expect a more similar codon usage bias between interacting proteins than non-interacting ones. Noteworthy, we find that this statement is not valid for all bacteria considered in this analysis if we measure CUB by not taking into account the information about tRNA levels. In other words, the pure codon statistics is not able to distinguish between proteins that are in contact or not in the protein interaction network. In Table 2, we report in boldface the 12 bacteria that pass the test (p-values < 0.001, Mann-Whitney test).

Table 2. P-values of the pairwise comparisons between the normalized distribution of scalar products of RSCU vectors for genes corresponding to interacting proteins in the PPI and distribution obtained by shuffling the PPI. For each bacterial dataset, we report the organism name, abbreviation, and p-value. The bacterial species for which the difference is statistically significant, according to the Mann-Whitney test (p-value < 0.001), are reported in boldface.

Organisms	Abbr.	p-value
Agrobacterium tumefaciens	agtu	1.28*10 ⁻⁶
Anabaena variabilis ATCC 29413	anva	1.02*10 ⁻⁴
Aquifex aeolicus VF5	aqae	0.20
Bifidobacterium longum NCC2705	bilu	0.15
Bordetella bronchiseptica RB50	bobr	0.27
Bordetella parapertussis 12822	bopa	1.06*10 ⁻³
Brucella melitensis bv. 1 str. 16M	brme	1.56*10 ⁻⁷
Buchnera aphidicola str. Bp	buap	0.13
Burkholderia pseudomallei K96243	bups	1.16*10 ⁻⁶
Buchnera aphidicola Sg uid57913	busg	0.035

Burkholderia thailandensis E264	buth	9.1*10 ⁻¹⁵
Caulobacter crescentus	cacr	0.15
Campylobacter jejuni	caje	0.10
Corynebacterium efficiens YS-314	coef	0.27
Corynebacterium glutamicum ATCC 13032	cogl	6*10 ⁻³
Chlamydia trachomatis D/UW-3/CX	chtr	0.18
Clostridium acetobutylicum ATCC 824	clac	7*10 ⁻¹¹
Francisella novicida U112	frno	0.66
Fusobacterium nucleatum ATCC 25586	funu	1.1*10 ⁻¹¹
Haemophilus ducreyi 35000HP	hadu	0.57
Klebsiella pneumoniae	klpn	0.89
Listeria monocytogenes EGD	limo	1.3*10 ⁻⁵
Mesorhizobium loti MAFF303099	melo	0.01
Mycoplasma genitalium G37	myge	5.8*10 ⁻⁵
Mycoplasma pneumoniae M129	mypn	5.2*10 ⁻⁵
Mycobacterium tuberculosis H37Rv	mytu	1.6*10 ⁻³
Porphyromonas gingivalis ATCC 33277	pogi	4.5*10 ⁻⁶
Pseudomonas aeruginosa UCBPP-PA14	psae	0.016
Ralstonia solanacearum GMI1000	raso	0.022
Rickettsia prowazekii str. Madrid E	ripr	0.18
Sphingomonas wittichii RW1	spwi	8.9*10 ⁻⁹
Staphylococcus aureus NCTC 8325	stau	0.02
Synechocystis sp. PCC 6803	sysp	3*10 ⁻⁴
Thermotoga maritima MSB8	thma	1.6*10 ⁻³
Vibrio cholerae N16961	vich	2.1*10 ⁻⁸
Xylella fastidiosa 9a5c	xyfa	5.1*10 ⁻⁴

168

169 3.2 Principal Component Analysis over the space of the codon bias indices

170 We then perform PCA over the space of the five codon bias indices (CompAI, CompAI_w, tAI, NC
171 and GC content) measured separately for each gene in each genome (see Figure 1 for an example).
172 The other plots are shown in Supplementary Material. The two first principal components (PC_1 and
173 PC_2) turn out to represent for as much as 65% of the total variance of codon bias over each genome

(see plot of Figure 2 for example). The other plots are shown in Supplementary Material. Projection of the first two principal components on the individual codon bias indices (loadings) for each genome shows that none of the five indices predominantly contributes to the data variability (see plot Figure 2 for example). Thus, the placement of a gene in the $PC_1 - PC_2$ plane depends on a weighted contribution of all the indices. So, we calculate the centroids of the eight top MCODE communities in this plane. We show that the genes encoding for the respective proteins are well localized in this reduced space (see Figure 1). For each genome, the first community (composed overall by 97 % from genes class COG J, that are translational, ribosomal structure and biogenesis) is always separated and isolated from the others. For the other communities, it depends on the bacteria: some bacteria such as *raso*, *ripr* or *bups* have all communities well separated, whereas other bacteria such as *caje*, *chtr* or *pogi* do not have separated centroids (see Figure 2 of Supplementary Material). In other words, for some bacteria, if a set of proteins are physically and functionally connected in a module, then their corresponding genes should share common codon bias features.

Figure 1. Example of plane $PC_1 - PC_2$. We show the example of *agtu* species. In the inset, there are centroids of first eight communities. We show that first community is always well separated from the others. The other bacteria are shown in Supplementary Material.

Figure 2. Example of eigenvalues of the correlation matrix. We show the example of *agtu* species. The first and second principal components ($PC_1 - PC_2$) turn out to represent as much as 68% of the total variance of codon bias. The other bacteria are shown in Supplementary Materials.

3.3 Z-score profiles: the closer is the codon usage of genes, the higher is the probability of protein interaction

Finally, for each bacterial genome, we calculate a conditional probability $Pr(link|d)$ of a functional interaction between proteins, given that their relative genes fall within a distance d in the plane of the two principal components $PC_1 - PC_2$ [2]. We compare $Pr(link|d)$ estimated on the real interactomes downloaded from STRING with $Pr(<link>_{\Omega}|d)$ estimated using the Configuration Model (CM) of the network, in which links are shuffled by preserving the interaction degree of each protein (see Materials and Methods for details).

To compare these two conditional probability, we study the Z-scores for $Pr(link|d)$ as a function of the Euclidean distance d between the codon usage bias of pair of genes in the PCA plane (see section . In Figure 3, we show the averaged Z-score profile over the set of bacteria considered in this analysis.

Figure 3. Z-scores of $Pr(link|d)$ as a function of the Euclidean distance d between the codon usage bias of pair of genes (computed via PCA), averaged over the set of chosen bacteria. The horizontal dashed lines mark the significance interval of 3σ standard deviations. We show the example of *agtu* species.

Interestingly, for small distances ($d \leq 3$), the probability of finding a connection between two proteins in the real interactome is much higher than in the CM ($Z\text{-score} > 1.96$). Conversely, for larger distances ($d > 3$) the real network and the CM become compatible ($-1.96 < Z\text{-score} < 1.96$).

Looking at the individual profiles in Supplementary Material, we note that this profile is common among different bacterial species. Thus, as a general rule, we conclude that sets of genes sharing similar codon usage patterns encode for proteins that are more likely to interact in the protein interaction network than expected by chance. Notably, for all the bacteria that pass the test for RSCU distributions in Table 2, the individual Z-score profiles are similar to Figure 3. This is also valid for some bacteria (i.e., *busg*, *cacr*, *caje*, *coef*, *hadu*, *kpn*, *raso*, and *stau*) that instead do not pass the test in Table 2. In contrast, the rest of the bacteria (i.e., *aqae*, *bilo*, *bobr*, *buap*, *chr*, *frno* and *melo*) are characterized by a different Z-score profile (See Supplementary Material), probably due to the low density of their PPI (See Table 1).

4. Discussion

In this work, we study how the coherence in codon usage among genes is reflected in the capacity of the encoded proteins to interact in the protein interaction network. For this purpose, we have extended our previous work taking *E. Coli* as a case of study [3] to other 34 bacterial genomes

characterized by different taxonomy [14]. As a general rule, we found that the only codon statistics based on the frequencies of occurrence of synonymous codons are not able to distinguish between proteins that make contacts or not in the protein interaction network. Conversely, by inserting the information contained in the tRNA levels, as expressed by tAI, CompAI, and CompAI_w, we observed that highly connected proteins belonging to the same communities in the protein interaction network are encoded by genes that are coherent in their codon choices. Specifically, our results provide evidence that if two genes have similar codon usage patterns, then the corresponding proteins have a significant probability of being functionally connected and interacting. Consequently, this study provides new information based on the similarity in codon usage of genes that can be potentially integrated into existing computational prediction methods of protein-protein interaction.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Behura SK, Severson DW, Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes, *Biol Rev Camb Philos Soc.* 2013 Feb; 88(1): 49-61. Doi: 10.1111/j. 1469-185X.2012.00242.x Epub 2012 Aug 14
2. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW, GenBank, *Nucleic Acids Res.* 2012;40 (Database issue):D48--D53
3. Dilucca M, Cimini G, Giansanti A, Essentiality, conservation, evolutionary pressure and codon bias in bacterial genes, *Gene* 2018 DOI:10.1016/j.gene.2018.04.0174.
4. Dilucca M, Cimini G, Semmoloni A, Deiana A, Giansanti A, Codon Bias Patterns of E. coli's Interacting Proteins. *PLoS ONE* 10(11), 2015: e0142127, <https://doi.org/10.1371/journal.pone.0142127>
5. dos Reis M, Savva R, Wernisch L., Solving the riddle of codon usage preferences: a test for translational selection, *Nucleic Acids Res.* 2004;32(17):5036—5044
6. Gavin CKW, Porras P, Aranda B, Hermjakob H, and Orchard SE, Analyzing Protein-Protein Interaction Networks, *J. Proteome Res.* 2012, 11, 4, 2014-2031
7. Hanson G and Collier J, Codon optimality, bias and usage in translation and mRNA decay, *Nat Rev Mol Cell Biol.* 2018 Jan; 19(1): 20–30. doi: [10.1038/nrm.2017.91](https://doi.org/10.1038/nrm.2017.91)
8. Hunter B. Fraser, Aaron E. Hirsh, Dennis P. Wall, and Michael B. Eisen, Coevolution of gene expression among interacting proteins, *PNAS* June 15, 2004 101 (24) 9033-9038; first published June 2, 2004 <https://doi.org/10.1073/pnas.0402591101>
9. Jolliffe IT, Principal component analysis, Second Edition, Springer
10. Kober KM, Pogson GH, Genome-wide patterns of codon bias are shaped by natural selection in the purple sea urchin, *strongylocentrotus purpuratus*, *G3: Genes, Genomes, Genetics* 3 (7), 1069-1083, 2013
11. Najafabadi HS, Goodarzi H, Salavati R, Universal function-specificity of codon usage, *Nucleic Acids Res.* 2009 Nov;37(21):7014-23. doi: 10.1093/nar/gkp792.
12. Park K, Lai YC,, Characterization of weighted complex networks, *Phys Rev E Stat Nonlin Soft Matter Phys.* 2004 Aug;70(2 Pt 2):026109. Epub 2004 Aug 23.
13. Plotkin JB, Kudla G, Synonymous but not the same: the causes and consequences of codon bias, *Nat.Rev.Genet* 2011, Jan; 12(1):32-42. doi:10.1038/nrg2899. Epub 2010 Nov 23
14. Plata G., Long-term phenotypic evolution of bacteria, *Nature* Vol 000 Month 2014 doi:10.1038/nature13827

15. Quax TE, Claassens NJ, Soll D, van der Oost J, Codon bias as a means to fine-tune gene expression, *Mol Cell*. 2015 Jul 16;59(2):149-61. Doi:10.1016/j.molcel.2015.05.035
16. Sabi R, Tuller T., Modelling the efficiency of codon-tRNA interactions based on codon usage bias. *DNA Res*. 2014;21(5):511–26
17. Szklarczyk D, Franceschini A, Wyder S, von Mering C et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life, October 2014, DOI: 10.1093/nar/gku1003
18. Tuller T, Challenges and obstacles related to solving the codon bias riddles, *Biochem Soc Trans*. 2014;42(1):155–159
19. Zhao F, Yu Ch, Liu Y, Codon usage regulates protein structure and function by affecting translation elongation speed in *Drosophila* cells, *Nucleic Acids Research*, Volume 45, Issue 14, 21 August 2017, pag: 8484-8492 doi.org/10.1093/nar/gkx501
20. Zou Z, Codon usage biases co-evolve with transcription termination machinery to suppress premature cleavage and polyadenylation, Research article Mar 16, 2018, doi:10.7554/eLife.33569
21. Wright F, The effective number of codons' used in a gene. *Gene*. 1990;87(1):23--29

© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).