MedCrave
Step into the World of Research

MOJ Proteomics & Bioinformatics

Research Article

Open Access

Analysis of codon usage bias of *Carnobacterium gilichinskyi* strain *WN1359* genome isolated from anoxic environment

Abstract

Earlier studies have shown that genomic features and genomic organization of microbial genome can be illustrated precisely by the study of codon usage bias. The availability of complete genome of *Carnobacterium gilichinskyi* strain WN1359, facilitates to get an insight about the genetic information of this microorganism. *Carnobacterium gilichinskyi* is a gram positive bacteria belongs to the *carnobacteriaceae* family. Present study aims to investigate the influence of mutational pressure and natural selection on the codon usage of this bacterial genome. It has been observed that choice of third synonymous codon position was mainly influenced by mutational bias suggesting that compositional constraint governing the codon usage bias in this bacterial genome. Moreover gene expression and translational efficiency have also been involved shaping the codon usage of this bacterial genome.

Volume 9 Issue 2 - 2020

Tanujaya Taran
Department of Microbiology, Tripura Institute of Paramedical Sciences, India

Correspondence: Tanujaya Taran, Department of Microbiology, Tripura Institute of Paramedical Sciences, India,
Email tanutaran@rediffmail.com

Received: April 15, 2020 | **Published:** May 04, 2020

Introduction

Carnobacterium gilichinskyi strain WN1359 was isolated from Siberian permafrost and was barophilic, psychrophilic and anoxic in nature.[1] So far a total of thirteen validly published members belong to the *Carnobacterium* genus. They are found in polar region, temperate environment due to their tolerance to freezing temperature and also can adapted in harsh environment (UV radiation, and cold temperatures).[2] Leonard et al.[3] first reported the whole genome sequence of *Carnobacterium gilichinskyi* isolated from Siberian permafrost.[3] Present study have analyzed the different influencing factor shaping the codon usage of the whole genome sequence of *Carnobacterium gilichinskyi* strain WN1359 that can grow in a simulated Mars atmosphere.[1]

Codon usage bias refers to a specific codon of a subset of codon used more frequently than other synonymous codons throughout the translation of gene.[4] Studies of codon usage in bacterial genome elucidate the lateral gene transfer (LGT) and horizontal gene transfer (HGT), is important in the movement of genetic material between unicellular and multicellular organism by the transmission of DNA from parent to offspring.[5–8] HGT used as a major evolutionary factor in many organisms. Codon usage gives the insight about the evolution of genome by determining the particular evolutionary forces operating the synonymous codon usage pattern within that genom.[9] Previous studies have demonstrated that codon usage pattern varies between organisms.[9–12] In addition, it has also been found that lateral gene transfer event influence the codon usage pattern of bacterial genomes, suggesting the horizontal gene transfer event could be a major evolutionary force for codon bias.[13] Codon optimization is a phenomenon where codons are replaced with an optimal synonymous codon in order to increase protein expression. This codon optimization relies on the accuracy of codon usage frequencies. Codon usage bias is altered by various factors such as gene expressivity,[14,15] high gene expression level shows greater codon bias than low expressed genes,[16] gene length;[17] tRNA abundanc[18,19] and amino acid composition.[20] This phenomena of codon usage bias can be utilized to characterize the evolutionary pattern of species and to determine the molecular marker. DNA probes and primers can be designed based on codon usage bias.[21] Mutational bias is the major factor responsible for shaping codon usage preference among bacteria. Codon bias is mainly governed by the balance between mutation selection and genetic drift among the organisms,[22] where translational selection pressure appears as a major responsible factor as certain codons used more preferably and are translated more accurately and efficiently.[23]

Methodology

Retrieval of whole genome sequence (WGS):

Whole genome and coding sequences of *Carnobacterium gilichinskyi* strain WN1359 was retrieved from NCBI (www.ncbi.nlm.nih.gov/genbank/genomes). Complete genome contains 2,275 coding sequence (CDS), to avoid sampling error, only those coding sequences (CDS) were selected which have correct start and stop codons with at least 100 codons in a sequence. Coding sequences having uncertain annotation were excluded.

Codon usage indices

Use of 59 codons to encode 18 amino acids (Met,Trp and stop codons are excluded) was determined by using a codon indices known as relative synonymous codon usage (RSCU),[24] codon adaptation index (CAI),[25] Effective Number of Codon (ENC)[26] frequency of G+C in CDS (GC content); G+C content at third codon position (GC3s) and occurrence of A, T, G, C at synonymous third codon position (A3,T3,G3,C3). RSCU can be defined as the ratio of the observed frequency of codons to the expected frequency. When RSCU value of a codon is greater than one it denotes that corresponding codon is used more often than the expected frequency. Synonymous codon usage bias refers the differences in the frequency of synonymous codons for individual amino acids in protein coding sequences. Mutation, selection and random genetic drift are the major factors that shape codon usage bias in different organisms. Whereas if the RSCU value is less than one than, reverse is true.[24] ENC analysis is used to quantify the extent of the codon usage bias of the coding sequences. ENC is the measures of synonymous codon bias and ENC values ranges from

20 (when an amino acid uses only one codon) to 61 (when a amino acid is encoded by all the synonymous codon with equal frequency).[26] Codon adaptation index (CAI) is mainly used to measure the extent of bias toward codons and is used as a quantitative method of predicting the level of highly expression of a gene, based on its codon sequence. The CAI value ranges from 0 to 1 where higher CAI indicates stronger codon bias and high gene expression level. CAI is the widely accepted method to measure the gene expression level.[25] These indices of codon usage bias were calculated by using a software Codon W version 1.4.2.[27]

Result and discussion

This study of codon usage on the genome of *Carnobacterium gilichinskyi* strain WN1359 isolated from harsh environment gives the insights about the genetic features of this bacteria. Our study on codon usage pattern provides information about the determining factors playing important role in shaping the codon usage pattern of this bacterial genome. Low genomic GC content (35.26%) predetermined there should be preferred number of A and T ending codon. We have observed seventeen optimal codons in this bacterial genome and among these optimal codons 11 are AT ending codon (4 T- ending and 7 A-ending), whereas in GC ending codon 6 are C ending and no G ending optimal codons were found. The degree of diversity in codon usage in *C. gilichinskyi* strain WN1359 was analyzed using ENC (Effective Number of Codon) and AT3s (AT at third position). It has been examined that ENC value ranges from 22.38 to 61 with an average value of 47.66. This gives an insight about the marked variation of codon usage in the genes in this bacterial Genome. Investigation of the genomic AT3s value also revealed the multiformity of codon usage bias among genes of *C. gilichinskyi* strain WN1359 bacteria. Presence of significant negative correlation (r= - 0.559, P <0.01) between $AT3_S$ and ENC has revealed the profound influence of compositional constraint of codon usage bias in this bacterial genome. These observations justify our hypothesis that variation of codon usage in this bacterial genome is mainly influenced by genomic compositional constraint. ENC vs GC_3s plot determine the various factors accounting the codon usage variation among the genes in this bacterial genome. It has been demonstrated by several earlier studies that heterogeneity of codon usage are due to the compositional constraint or completely dictated by GC3s, the values of ENC should fall on the expected curve and the NC plot of the *C. gilichinskyi* strain WN1359 genome shows that maximum number of genes with higher GC3s with lower NC value and are lying far below the expected curve (Figure 1). This indicates that there must be certain additional factors along with compositional constraint influencing the codon usage bias of this genome. This result promotes the hypothesis that translational selection also contributing the codon usage of this bacterial genome.

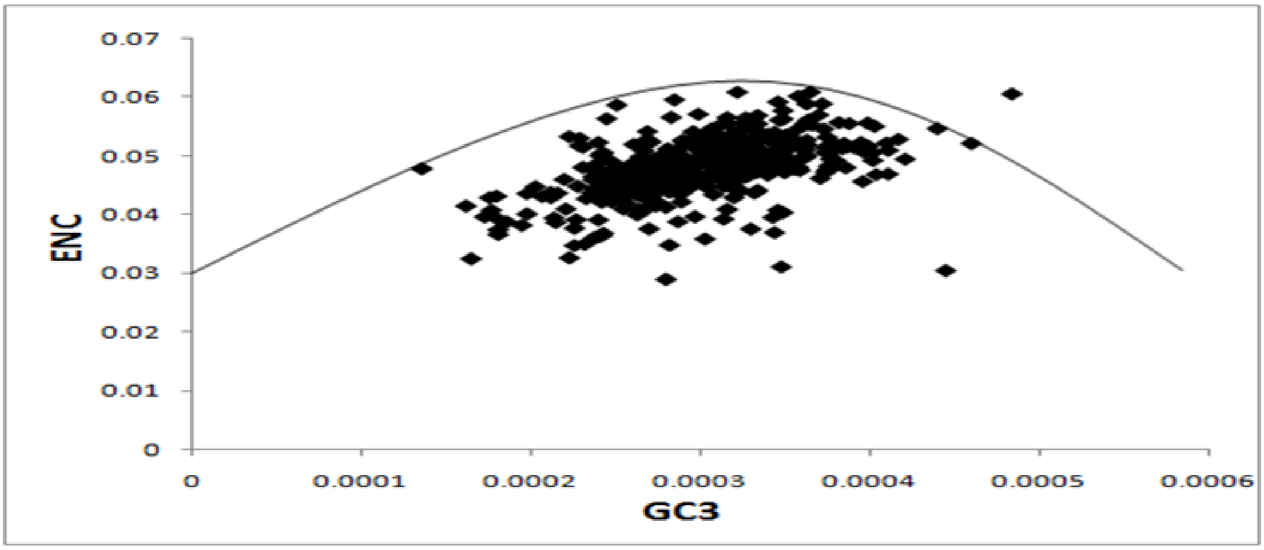

Figure 1 Relationship between ENC and GC3.

Optimal codons

A chi square test was performed between 10% of gene (highly expressed) with higher value of major axis and 10 % of genes (lowly expressed) with lower value of axis. In highly biased genes 17 codons are identified with higher usage which could be referred to as translational optimal codons (10). Optimal codons help to achieve faster translational rate and high accuracy. Among these 17 codons identified as optimal codon, 7 are a ending (41.17%) codon and rest of the 4 codons are T ending (23.52 %) codon and there is no G ending and 6 are C (35.29%) ending codon identified as a optimal codon.

Influence of compositional constraint

We have performed correspondence analysis on the RSCU values and the distribution of genes along the first two major axis has been investigated. The first major axis accounts for 12.21 % of the total variation. The second major axis accounts for 4.15 % of the total variation. Hence the first major axis may be regarded as the single major explanatory axis to account for codon usage variation in this bacterial genome. The changes in third codon position may not be associated with the changes of amino acids encoding by that gene sequences representing that mutational bias may drive different codon choice without affecting the protein sequence.[28] A strong negative correlation between ENC and AT3s also suggest that the genes lies on the right side of the axis 1, has higher AT content and lower ENC values which represent a strong correlation between codon usage preference and nucleotide composition of the genes encoded by this bacterial genome. The ENC showed a significant positive correlation with A3s whereas T3s does not show any correlation but C3s and G3s is negatively correlated with ENC (Table 1).

Distinctive Distribution of Nucleotides at Third codon position

The relationship of the frequencies of A, T, G, C at the third codon position with the ENC values of all the genes has been estimated (Table 1). Several earlier studies have been reported that highly expressed genes are highly biased with low ENC value and vice versa. In this bacterial genome the frequency of A3 increases with decrease ENC value, whereas frequency of G and C at third codon position decreased with higher ENC. The frequency of A3 also positively correlated with decrease ENC but it did not show any strong correlation between T3 and ENC. This result reflects that the mutational bias determine the choice of bases in the third codon position. It assumes that optimal codons are selected based on the mutational bias, so this is expected that translational selection influences the mutational bias to be more prominent in the third codon position in the highly expressed genes (Figure 2).

Table 1 Correlation of ENC with T3s, C3s, G3s, A3s

Correlation	T3s	C3s	G3s	A3s
ENC	0.04NS	-0.367**	-0.276**	0.105**

**significant relationships; NS, Non-significant; P<0.001 for significant values

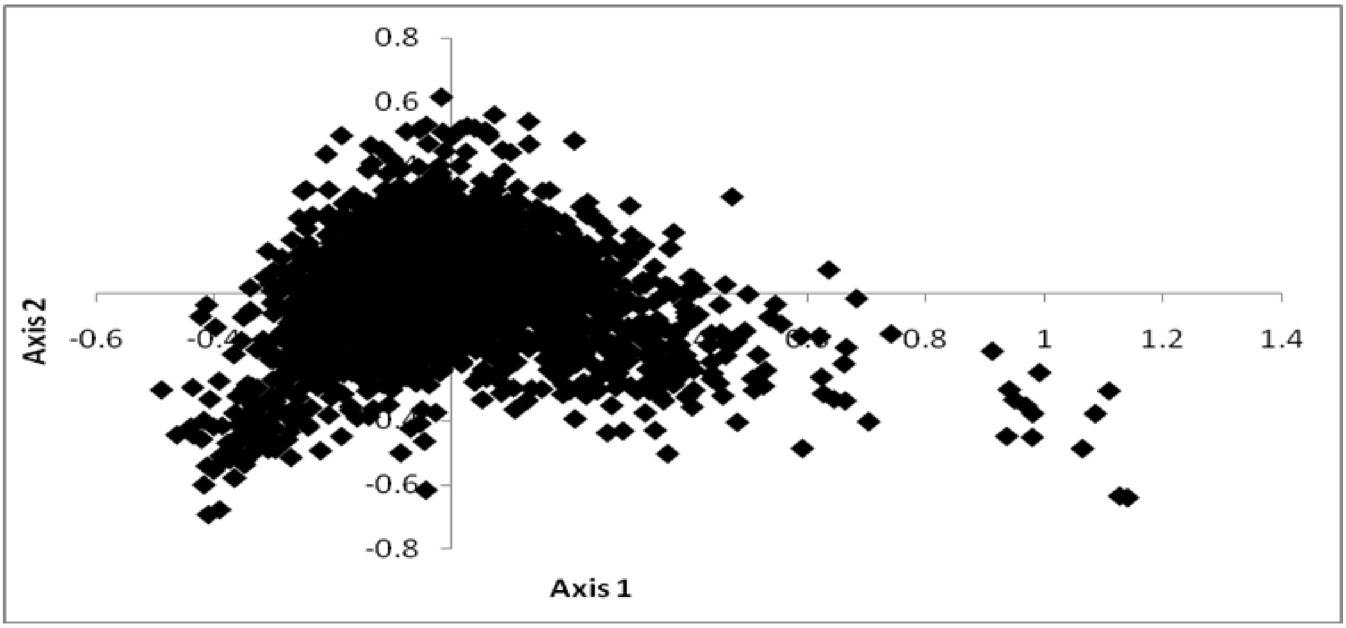

Figure 2 Position of genes along the first two major axis of variation in the correspondence analysis on RSCU.

Gene length influencing Codon bias

Selection pressure speed up the translational process with minimizing proofreading cost and maximizing translation accuracy by using those codons having abundant cognate tRNAs. Several studies have reported that Codon usage bias regulate protein translation efficiency and accuracy and protein folding.[29–35] Translational processivity has a predominant effect on the relationship of gene length and synonymous codon bias.[17] Energy cost and resources consumed for a protein production is proportional to gene length. So that it is expected the selection pressure will be higher in case of longer gene to avoid missense error. It is thought to be there is a positive correlation between codon bias and gene length due to the selection pressure for translational accuracy. A plot has been drawn on ENC values against gene length which depicts a wider variation of ENC values with shorter gene length than that of longer gene. In this bacterial genome the gene lengths are correlated with gene position on axis 1, GC_3 and ENC values (r= -0.252, 0.268 and -0.157 respectively, p <0.01) significantly. ENC (r= - 0.135, p <0.01) shows significant negative correlation with gene length. The significant correlations represent codon usage of this bacterial genome influenced by gene length. So that codon bias is higher in shorter genes than in longer ones (Figure 3).

Role of gene expression level on codon usage

CAI is a widely accepted method to estimate the expression level of the genes in organisms (36). *C. gilichinskyi* strain WN1359 genome represent a significant negative correlation of CAI with axis 1 ($r = -0.476^{**}, P < 0.001$) and a positive correlation with AT3 (and $r = 0.472^{**}$, $P < 0.001$) where as ENC shows a significant negative correlation with CAI ($r = -0.507^{**}, P<0.01$).This observation indicates the effect of gene expressivity on the codon usage in this bacterial genome. Furthermore, gene expressivity (CAI) showed a positive correlation with A3s ($r = 0.333^{**}, p<0.01$) and T3s ($r = 0.241^{**}, p<0.01$), where A3s exhibiting stronger relationship with CAI than T3s suggesting that mutational bias effecting the choice of bases at the third codon position which favors translational selection.

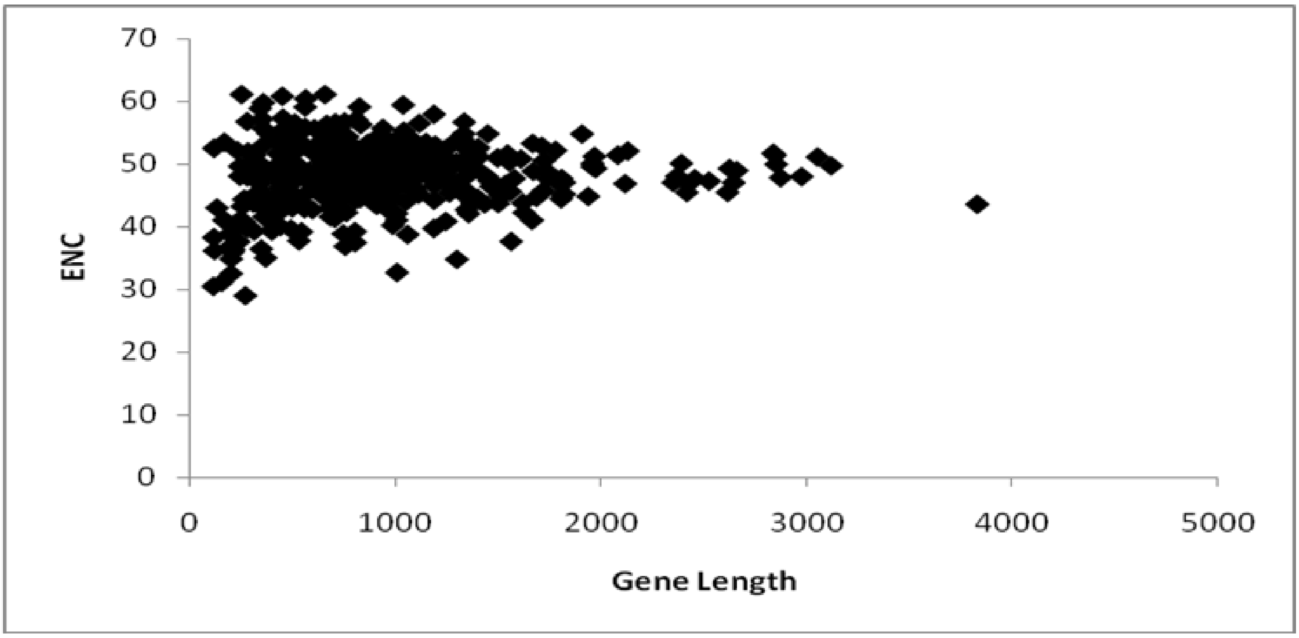

Figure 3 ENC vs Gene length Plot for *C. gilichinskyi* strain WN1359 genome.

Conclusion

This study has been performed to deduce the influencing factors responsible for variation in codon usage preference in *C. gilichinskyi* strain WN1359 strain. From this study, it is easy to perceive that other than the compositional constraint, some other factors also involved in shaping the codon usage pattern in this bacterial genome. From this study it has been observed that mutational pressure plays a significant role in shaping the codon usage pattern in this bacterial genome. Among all the third synonymous codon positions A3s is more predominant in influencing codon usage variation in this genome. This observation suggests that choice of third synonymous codon position is mainly influenced by mutational bias which support our hypothesis that compositional constraint governing the codon usage bias in this genome. In highly expressed genes A ending codons are used most preferably which may be associated with translational efficiency of the gene. It is worth to note that 41.17 % of the most frequently used codon in the highly expressed gene of *C. gilichinskyi* strain WN1359 are A (purine) ending codons suggesting along with the mutational pressure, translational selection pressure also operate the codon usage bias in this bacterial genome. Thus it is important to note that mutational pressure is most predominant in shaping codon usage in *C. gilichinskyi* strain WN1359 genome but some other factors such as gene expression level, translational accuracy and translational efficiency are also involved in codon usage in this bacterial genome.

Acknowledgments

None.

Conflicts of interests

Author declares that there are no conflicts of interests.

Funding

None.

References

1. Nicholson WL, Krivushin K, Gilichinsky D, et al. Growth of Carnobacterium spp. from permafrost under low pressure, temperature, and anoxic atmosphere has implications for Earth microbes on Mars. *Proc Natl Acad Sci U S A*. 2013;110:666–671.
2. Franzmann PD, Höpfl P, Weiss NA. Psychrotrophic, lactic acid-producing bacteria from anoxic waters in Ace Lake, Antarctica; Carnobacterium funditum sp nov and Carnobacterium alterfunditum sp nov. *Archives of microbiology*. 1991;156(4):255–262.
3. Leonard, Michael T, Famerie WG, et al. Complete genome sequence of Carnobacterium gilichinskyi strain WN1359T (DSM 27470T). *Genome Announc*. 2013;1(6):e00985–e01013.
4. Sharp PM, Li WH. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic acids research*. 1987;15(3):1281–1295.
5. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature*. 2000;405(6784):299–304.
6. Dunning Hotopp JC. Horizontal gene transfer between bacteria and animals. *Trends in Genetics*. 2011;27(4):157–163.
7. Robinson, Kelly M, Karsten B. A review of bacteria-animal lateral gene transfer may inform our understanding of diseases like cancer. *PLoS genetics*. 2013;9:10.
8. Keeling, Patrick J, Jeffrey D Palmer. Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics*. 2008;9(8):605–618.
9. Grantham R, Gautier C, Gouy M, et al. Codon catalog usage and the genome hypothesis. *Nucleic acids research*. 1980;8(1):197–197.

10. Stenico M, Lloyd AT, Sharp PM. Codon usage in Caenorhabditis elegans: delineation of translational selection and mutational biases. *Nucleic Acids aResearch*. 1994;22(13):2437–2446.
11. Gupta SK, Bhattacharyya TK, Ghosh TC. Synonymous codon usage in Lactococcus lactis: mutational bias versus translational selection. *Journal of Biomolecular Structure and Dynamics*. 2004;21(4):527–535.
12. Hsiao, Yi-Yuong, Liu K, et al. Analysis of codon usage patterns in toxic dinoflagellate Alexandrium tamarense through expressed sequence tag data. *Comparative and functional genomics*. 2010;2010:138538.
13. Chen SL, Lee W, Hottes AK, et al. Codon usage between genomes is constrained by genome-wide mutational processes. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(10):3480–3485.
14. Grantham R, Gautier C, Gouy M, et al. Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic acids research*. 1981;9(1):213–213.
15. Gouy M, Gautier C. Codon usage in bacteria: correlation with gene expressivity. *Nucleic acids research*. 1982;10(22):7055–7074.
16. Roymondal U, Das S, Sahoo S. Predicting gene expression level from relative codon usage bias: an application to Escherichia coli genome. *DNA research*. 2009;16(1):13–30.
17. Eyre-Walker A. Synonymous codon bias is related to gene length in Escherichia coli: selection for translational accuracy? *Molecular biology and evolution*. 1996;13(6):864–872.
18. Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular biology and evolution*. 1985;2(1):13–34.
19. Mitreva M, Wendl MC, Martin J, et al. Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. *Genome Biology*. 2006;7(8):R75.
20. Lobry JR, Gautier C. Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 Escherichia coli chromosome-encoded genes. *Nucleic Acids Research*. 1994;22(15):3174–3180.
21. Supek F, Vlahoviček K. Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity. *BMC bioinformatics*. 2005;6(1):182.
22. Bulmer M. Coevolution of codon usage and transfer RNA abundance. *Nature*. 1987;325(6106):728–730.
23. Hershberg R, Petrov DA. Selection on codon bias. *Annual review of genetics*. 2008;42:287–299.
24. Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. *Journal of molecular evolution*. 1986;24(1–2):28–38.
25. Sharp, Paul M, Wen-Hsiung Li. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic acids research*. 1987;15(3):1281–1295.
26. Wright F. The ‘effective number of codons’ used in a gene. *Gene*. 1990;87(1):23–29.
27. Peden JF. *Analysis of codon usage* (Doctoral dissertation, University of Nottingham). 2000.
28. Bulmer M. Coevolution of codon usage and transfer RNA abundance. *Nature*. 1987;325(6106):728–730.
29. H Akashi. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics*. 1994;136:927–935.
30. DA Drummond, CO Wilke. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*. 2008;134:341–352.
31. H Gingold, Y Pilpel. Determinants of translation efficiency and accuracy. *Mol Syst Biol*. 2011;7:481.
32. R Hershberg, DA Petrov. Selection on codon bias. *Annu Rev Genet*. 2008;42:287–299.
33. JB Plotkin, G Kudl. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*. 2011;12:32–42.
34. PS Spencer, E Siller, JF Anderson, et al. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J Mol Biol*. 2012;422:328–335.
35. Y Xu, P Ma, P Shah, et al. Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature*. 2013;495:116–120.
36. Gerstein, Bussemaker, Harmen J, et al. Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Research*. 2003;31(8):2242–2251.

Citation: Taran T. Analysis of codon usage bias of *Carnobacterium gilichinskyi* strain *WN1359* genome isolated from anoxic environment. *MOJ Proteomics Bioinform.* 2020;9(2):36–40. DOI: 10.15406/mojpb.2020.09.00276