Journal of
Proteomics & Bioinformatics

Uddin, J Proteomics Bioinform 2017, 10:5
DOI: 10.4172/jpb.1000e32

Editorial Open Access

# Codon Usage Bias: A Tool for Understanding Molecular Evolution

**Arif Uddin***

*Department of Zoology, Moinul Hoque Choudhury Memorial Science College, Hailakndi, India*

Genetic code is a sequence of three nitrogen bases which encodes a particular amino acid. It is the set of codons which encode twenty amino acids and protein termination signals. In living cells, genetic information in DNA is transcribed into mRNA which subsequently translated into proteins. It is well known that there are 64 codons in standard genetic code, out of which 61 represent 20 standard amino acids and the remaining three are stop codons (TAA, TAG and TGA). Due to the degeneracy of genetic code, most of the codons (except Met, Trp) encode the same amino acid, termed as synonymous codons. The choice of codons encoded same amino acids is species specific and consequently the codons occur at uneven frequencies in genes [1,2].

The phenomenon of unequal usage of codons i.e., some codon are used more frequently than others make codon usage bias [3]. It is a common tendency in a variety of organisms, including prokaryotes as well as eukaryotes [4,5]. The pattern of codon usage bias is a unique property of a genome [6]. Furthermore, within the same organism, different tissues display a different codon usage pattern so the pattern of codon usage was different in different organs [7]. On the other hand, mutations in the third position of codon generally change the synonymous codons with no change in the encoded amino acid thus conserving the primary sequence of the protein [8]. The codon usage bias was first reported to as early as four decades ago. Earlier Clarke [9] and later Ikemura [10], proposed that codon usage adapted to match an organism's tRNA pool [10,11]. Ikemura [3] proved that evolutionary forces acting on the choices of codons marks differences in codon bias between species.

Generally, the codon usage bias is mainly influenced by compositional constraints under mutational pressure and natural selection. These two are the two major evolutionary forces accounting for codon usage variation among genomes [5,12,13]. Apart from these, expression level, gene length, replication, RNA stability, hydrophobicity and hydrophilicity of the [4,14-16], also affect the codon usage bias. In some organisms, codon usage is due to mutation pressure and genetic drift whereas in others, it is due to balance between natural selection and mutational biases [17]. Mutation pressure plays an important role in affecting the synonymous codon usage bias in some genes with very high content of any one of the four nucleobases [5,18-20]. The proportion of extremely high or low Guanine (G) or cytosine (C) nucleobase in the 3$^{rd}$ position of codon in an open reading frame signify mutational bias [21]. The alternations of biochemical mechanism i.e., more recurrent changes of certain bases than others cause mutational biases [22,23]. Mutation pressure is mostly responsible for codon usage bias in some prokaryotes and in many mammals with high AT or GC contents [5,19]. On the other hand, in Drosophila and in some plants, the codon usage bias is primarily caused by translational selection [24]. The non-synonymous substitution is determined by selection because it amends the amino acids and consequently biochemical nature of protein is affected [1].

Some previous reports suggested that in highly expressed genes, the codon usage bias is due to translational selection. In highly expressed genes, favored codons are easily recognized by the abundant tRNA molecules [25,26]. The relationship between codon bias and the level of gene expression has been experimentally established in *Escherichia coli* [27].

Analysis of codon usage bias is important in understanding the molecular biology, genetics and genome evolution [28,29]. It also helps in new gene discovery [29], design of primers [30], design of transgenes [29], determining the origin of species [31], and prediction of expression level [32], heterologous gene expression [33], and prediction of gene function [34].

## References

1. Plotkin JB, Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet 12: 32-42.
2. Plotkin JB, Dushoff J, Desai MM, Fraser HB (2006) Codon usage and selection on proteins. J Mol Evol 63: 635-653.
3. Ikemura T (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J Mol Biol 146: 1-21.
4. Akashi H (1997) Codon bias evolution in Drosophila. Population genetics of mutation-selection drift. Gene 205: 269-278.
5. Sharp PM, Stenico M, Peden JF, Lloyd AT (1993) Codon usage: mutational bias, translational selection, or both? Biochem Soci Trans 21: 835.
6. Grantham R, Gautier C, Gouy M (1980) Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. Nucleic Acids Res 8: 1893-1912.
7. Plotkin JB, Robins H, Levine AJ (2004) Tissue-specific codon usage and the expression of human genes. Proc Natl Acad Sci U S A 101: 12588-12591.
8. Biro JC (2008) Studies on the origin and evolution of codon bias. Biomolecules 0807: 3901.
9. Clarke B (1970) Darwinian evolution of proteins. Science 168: 1009-1011.
10. Clarke TF, Clark PL (2010) Increased incidence of rare codon clusters at 5'and 3'gene termini: implications for function. BMC Genomics 11: 1.
11. Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biology Evol 2: 13-34.
12. Shields DC, Sharp PM, Higgins DG, Wright F (1988) "Silent" sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. Mol Bio Evol 5: 704-716.
13. Stenico M, Lloyd AT, Sharp PM (1994) Codon usage in Caenorhabditis elegans: delineation of translational selection and mutational biases. Nucleic Acids Res 22: 2437-2446.
14. Moriyama EN, Powell JR (1998) Gene length and codon usage bias in Drosophila melanogaster, Saccharomyces cerevisiae and Escherichia coli. Nucl Acids Res 26: 3188-3193.
15. Powell JR, Moriyama EN (1997) Evolution of codon usage bias in Drosophila. Proc Natl Acad Sci U S A 94: 7784-7790.
16. Powell JR, Sezzi E, Moriyama EN, Gleason JM, Caccone A (2003) Analysis of a shift in codon usage in Drosophila. J Mol Evol 57: S214-S225.

***Corresponding author:** Uddin A, Department of Zoology, Moinul Hoque Choudhury Memorial Science College, Algapur, Hailakndi-788150, India, Tel: 09613554108; E-mail: arif.uddin29@gmail.com

**Received** May 27, 2017; **Accepted** May 29, 2017; **Published** May 31, 2017

**Citation:** Uddin A (2017) Codon Usage Bias: A Tool for Understanding Molecular Evolution. J Proteomics Bioinform 10: e32. doi:10.4172/jpb.1000e32

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

18. Karlin S, Mrazek J (1996) What drives codon choices in human genes? J Mol Bio 262: 459-472.

19. Zhao S, Zhang Q, Chen Z, Zhao Y, Zhong J (2007) The factors shaping synonymous codon usage in the genome of *Burkholderia mallei.* J Genet Genomics 34: 362-372.

20. Zhong J, Li Y, Zhao S, Liu S, Zhang Z (2007) Mutation pressure shapes codon usage in the GC-Rich genome of foot-and-mouth disease virus. Virus Genes 35: 767-776.

21. Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. Proc Natl Acad Sci 85: 2653-2657.

22. Francino MP, Ochman H (2001) Deamination as the basis of strand-asymmetric evolution in transcribed *Escherichia coli* sequences. Mol Biol Evol 18: 1147-1150.

23. Green P, Ewing B, Miller W, Thomas PJ, Green ED (2003) Transcription-associated mutational asymmetry in mammalian evolution. Nat Genet 33: 514-517.

24. Liu Q, Feng Y, Zhao Xa, Dong H, Xue Q (2004) Synonymous codon usage bias in *Oryza sativa*. Plant Sci 167: 101-105.

25. Bibb M, Findlay P, Johnson M (1984) The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene 30: 157-166.

26. McEwan NR, Gatherer D (1999) Codon indices as a predictor of gene functionality in a Frankia operon. Can J Bot 77: 1287-1292.

27. Andersson S, Kurland C (1990) Codon preferences in free-living microorganisms. Microbiol Rev 54: 198-210.

28. Sharp PM, Matassi G (1994) Codon usage and genome evolution. Curr Opin Genet Dev 4: 851-860.

29. Yang X, Luo X, Cai X (2014) Analysis of codon usage pattern in *Taenia saginata* based on a transcriptome dataset. Parasites Vect 7: 1-11.

30. Zheng Y, Zhao WM, Wang H, Zhou YB, Luan Y, et al. (2007) Codon usage bias in *Chlamydia trachomatis* and the effect of codon modification in the MOMP gene on immune responses to vaccination. Biochem Cell Biol 85: 218-226.

31. Ahn I, Jeong BJ, Bae SE, Jung J, Son HS (2006) Genomic analysis of influenza A viruses, including avian flu (H5N1) strains. Eur J Epidemiol 21: 511-519.

32. Gupta S, Bhattacharyya T, Ghosh TC (2004) Synonymous codon usage in *Lactococcus lactis*: mutational bias versus translational selection. J Biomol Struct Dyn 21: 527-535.

33. Kane JF (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. Cur Opinion Biotechnol 6: 494-500.

34. Lin K, Kuang Y, Joseph JS, Kolatkar PR (2002) Conserved codon composition of ribosomal protein coding genes in *Escherichia coli*, *Mycobacterium tuberculosis* and *Saccharomyces cerevisiae*: lessons from supervised machine learning in functional genomics. Nucleic acids Res 30: 2599-2607.