Codon Usage Bias: A Tool for Understanding Molecular Evolution

Arif Uddin*
Department of Zoology, Moinul Hoque Choudhury Memorial Science College, Hailakandi, India

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

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

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


Copyright: © 2017 Uddin A. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

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, hydrophylicity and hydrophobicity 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 3rd 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].