ABSTRACT

The flow of information in the cell begins at DNA, which replicates to form more DNA. Information is then "transcribed" (Written/Copied) into RNA, and then it is "translated" (Decoded) into protein. The proteins do most of the work in the cell. Proteins are formed out of 20 amino-acids which are coded in triplets of nucleotides, called codons. The four nucleotides (A;T;C;G) define 64 codons used in the cell. Codons are not uniformly employed in the cell, but at the contrary, certain codons are preferred and we speak about codon bias. Several indices can be used to measure the degree of non-random usage of synonymous codons in a gene, of which effective number of codons (ENC) ($\widehat{N}_c$) is widely used. It is a measure to quantify how far the codon usage of a gene deviates from uniform usage of synonymous codons. As background nucleotide composition is a dominating factor on codon usages, the departures from codon usage is not always desirable. ENC prime ($\widehat{N}_c'$) gives CUB in a gene after filtering out expected CUB due to background nucleotide composition. However it is seen that the $\widehat{N}_c$ value can be erroneous when codon usage is uniform for a less frequently used amino acid in a gene and hence ($\widehat{N}_c'$) can also be erroneous as it is derived from $\widehat{N}_c$ .The objective of our project is to rectify the error in ($\widehat{N}_c'$) and provide a web implementation of improved $\widehat{N}_c'$ for measuring CUB.