

ABSTRACT

Gene is a DNA fragment, which is a sequence of four different nucleotides: Adenine (A), Guanine (G), Cytosine (C) and Thymine (T). These sequences are used in the form of triplets known as codons to produce different proteins so that they can be utilized in the body functions in the organism. Protein is a sequence of 20 different amino acids. Genetic code table represents a mapping of 64 codons to 20 different amino acids. As the numbers of codons in the genetic code table are more than the number of amino acids, multiple codons code for same amino acids. The codons which code for the same amino acids are known as synonymous codons. Synonymous codons are not used with equal frequencies. This unequal usage of synonymous codons is known as Codon Usage Bias (CUB). Stronger CUB is observed in high expression genes for efficient translation in comparison to the whole gene set in an organism.

Several methods have been proposed to measure CUB. Few widely used methods are CAI, Nc, MELP etc. In this report we have studied the applicability of these methods in predicting gene expression. We used ROC (Receiver Operating Characteristics) method to compare CUB measures with the experimentally generated gene expression values in eighteen organisms. As per our observation CAI is the best known measure to predict gene expression in most of the organism. Further the method gives better result in larger genes in comparison to smaller genes within an organism.