Abstract

The genetic code is the set of rules by which information encoded within genetic material (DNA or mRNA sequences) is translated into proteins. Protein is a sequence of 20 different amino acids. The Genetic code table represents a mapping of 64 codons to 20 different amino acids. A codon is a combination of three nucleotides (A, U, G, C) that encodes a specific amino acid. As the numbers of codons in the genetic code table are more than the number of amino acids, multiple codons are used for same amino acid. The codons which code for the same amino acid are known as synonymous codons. Synonymous codons are not used with equal frequencies. This unequal usage of synonymous codon is known as Codon Uses Biases (CUB).

Codon Adaption Index (CAI) is the most widespread technique for analyzing Codon Uses Biases. CAI measures the deviation of a given protein coding gene sequence with respect to a reference set of genes. CAI is widely used to theoretically measure the level of gene expression.

In this Project, we have used ROC(Receiver operating characteristics) method to compare CAI method with the experimentally generated gene expression values in 17 organism. We observed that the CAI may be used to predict gene expression in organisms with strong selection pressure on codon usage, whereas the reverse may not be true always.