

## Abstract

Synonymous codons encoding same amino acids are not used uniformly. This phenomenon known as codon usage bias (CUB) prevails within a genome as well as across different organisms. Since the differential usage of synonymous codons has little influence on amino acid sequence of proteins, the evolution of synonymous codon usage attracted many scientists to study molecular evolution in organisms.

The four-fold degenerate site (FDS) in coding sequences is important for studying the effect of any selection on codon usage bias (CUB) because nucleotide substitution *per se* is not under any selection pressure at the site due to the unaltered amino acid sequence in a protein. We estimated the frequency variation of nucleotides at the FDS across the eight family boxes (FBs) using a method named, unevenness measure of gene,  $Um(g)$ . Our study in total 545 species of bacteria revealed that in many bacteria,  $Um(g)$  correlated strongly with  $Nc'$  - a measure of the CUB. Analysis of the strongly correlated bacteria revealed that the U-ending codons (GGU, CGU) were preferred over the G-ending codons (GGG, CGG) in Gly and Arg FBs in genomes even with G+C% higher than 65.0. Further evidence suggested that these codons can be used as a good indicator of selection on codon usage bias in genomes with higher G+C%.

It is generally believed that the effect of translational selection on codon usage bias is related to the number of transfer RNA genes in bacteria which is more with respect to the high expression genes than the whole genome. Keeping this in the background we analyzed codon usage bias with respect to asparagine, isoleucine, phenylalanine and tyrosine amino acids in seventeen bacteria, whose gene expression data and the tRNA gene numbers were available. In most of the bacteria it was observed that codon usage bias and tRNA gene number were not in agreement, which was unexpected. Our study indicates that tRNA gene numbers may not be the sole determining factor for translational selection of codon usage bias in bacterial genomes.

Constancy in the relative di-nucleotide frequency (RDF) in different parts of a genome is termed as genome signature in bacteria. We further analyzed RDF in different parts of a genome such as the inter-genic and the coding sequences in order to understand its effect on CUB. Our findings indicate that codon usage bias is a major attribute towards the constraint on di-nucleotide in genomes.