

Codon selection pattern in *Thermosynechococcus elongatus* genome

Ratna Prabha¹, Dhananjaya Pratap Singh¹, Alok K. Srivastava¹, Dilip K. Arora¹

¹National Bureau of Agriculturally Important Microorganisms,

Kushmaur, Maunath Bhanjan, 275101 India

Background: Codon usage pattern was analyzed in *Thermosynechococcus elongatus*, a thermophilic cyanobacterium and a model organism for the study of photosynthesis. Codon usage is mainly a genome strategy. Each genome has a particular codon usage signature that reflects particular evolutionary forces acting within that genome. Codon bias is maintained by a balance between selection, mutation, and genetic drift. Codon usage has direct utility in molecular characterization of species. Codon usage variations are widely used for quantitative prediction of gene expressivity.

Method: The frequency of 59 codons (except Met, Trp, and stop codon) was examined by using different codon indices, which are relative synonymous codon usage (RSCU); effective number of codons (Nc); codon adaptation index (CAI); frequency of G+C in a coding sequence (GC content); frequency of G+C at third position of codons (GC3s); frequency of A, T, G, C at the synonymous third positions of codons (A3, T3, G3, C3 respectively). Correspondence Analysis (CA) and Chi-square test was performed to identify the relation between various factors.

Result & Conclusion: In *T. elongatus*, there is a single major explanatory axis which accounted for codon usage variation. CA of *T. elongatus* genes reveals a strong correlation of GC3s with this axis. In addition, a negative correlation was observed between Nc and GC3s. These results suggest that mutational bias is the major factor in shaping codon usage in this organism. Among all the four nucleotides at third position, C3s is found to directly influence codon usage variation. Highly expressed genes tend to use C or G at the synonymous positions compared with lowly expressed genes which might be related to the translational efficiency of the genes as it has been reported that RNY (R- purine, N- any nucleotide base, and Y-pyrimidine) codons are more advantageous for translation. It is interesting to note that 78% of the preferred triplets in the highly expressed genes of *T. elongatus* are pyrimidine ending codons. This result suggests that besides, mutational bias, translational selections are also operating in determining the codon usage variation in this organism. This hypothesis is also suggested by the fact that among the 18 most used codons in entire genome, 13 are pyrimidine ending codons. A role of translational accuracy was also identified in dictating the codon usage variation of this genome.

Our results suggest that mutational bias is the major factor in shaping codon usage in *T. elongatus* genome, but other factors, namely, translational selection and translational accuracy also influenced the codon usage variation.