Human and fly protein-coding genes contain more stop resistant codons than random nucleotide sequences

Francisco Prosdocimi[1], J. Miguel Ortega[1]

[1] Lab. Biodados, ICB-UFMG.

It is well known that genetic code minimizes the effect of mutations and similar codons usually codify for the same amino acid, avoiding non-synonymous mutations. Although somewhat allowing mutation resistance, synonymous codons may vary in their mutational distance to stop codons. For example, both UUA and CUU codons codify Leucine residues; however one single mutation in UUA Leucine codon may produce the stop codons UGA or UAA, while CUU would need more mutations to be replaced by a stop codon. Here, we have evaluated the codon usage of human and fly protein-coding genes to check their preference for stop-resistant codons. Analyzing each group of synonymous codons, we have developed a metric to calculate the distance of each codon to a stop. So, we have downloaded all human and fly protein-coding sequences from KEGG Orthology. A FASTA file containing 10,000 sequences composed of 100 random codons was also produced. Considering a given protein, we were able to build a synonymous protein presenting the worst and the best codon configuration in order to avoid mutations to stop. While random sequences have revealed a stop-distance from 28.67 to 157.08, averaged 55.9 ($\pm$ 5.28), human and fly proteins have shown to be more resistant to stop codon mutations, showing stop-distance average indexes of 63.29 ($\pm$ 6.42) and 62.99 ($\pm$ 6.02), respectively. Therefore, actual proteins have been selected to present stop-distant codons.