[BIOINFO 2023]

Comprehensive analysis of synonymous codon usage in human Coxsackievirus B3

Xianglan Min[1], Myeongji Cho[1], Mikyung Je[3], and Hyeon S. Son[1,2,3,*]

[1]*Laboratory of Computational Biology & Bioinformatics, Graduate School of Public Health, Seoul National University*

[2]*Institute of Health and Environment, Seoul National University*

[3]*Interdisciplinary Graduate Program in Bioinformatics, College of Natural Science, Seoul National University*

**Corresponding author: hss2003@snu.ac.kr*

Coxsackievirus B3 (CVB3) is a group of enteroviruses belonging to the *Picornaviridae* family, whose infection can cause myocarditis, pancreatitis, aseptic meningitis, and is related to the occurrence of type 1 diabetes. CVB3/GA is a naturally occurring, avirulent strain, while CVB3/28 is a highly virulent strain that has been extensively studied. Previous studies on the 5 'UTR region of two strains showed that alterations in the secondary structure of Internal Ribosome Entry Site (IRES) domain Ⅱ and the compactness of IRES tertiary structure affect the virulence and viral titer of CVB3. It is well known that IRES can initiate cap-independent translation by directly recruiting ribosomes, suggesting that ribosome binding efficiency is a determinant of CVB3 virulence. This led us to wonder whether the CAI and gtAI values of virulent strains in the downstream polyprotein CDS region are also significantly higher than the avirulent strain? Within this study, we performed a comprehensive synonymous codon usage analysis on 23 CVB3 polyprotein coding sequences to explore possible factors associated with CVB3 virulence. The mean ENC value of CVB3 was observed to be 54.48, indicating the overall codon usage bias is not high. Among them, CVB3/GA is the strain that uses synonymous codons most evenly. However, the TTG[L] codon of CVB3/GA was distinguished from the other sequences by abnormally over-represented (RSCU=1.457). Since each sequence has 59 RSCU values, we performed PCA and t-SNE dimensionality reduction analyses and found that the CVB3/GA sequence did not cluster with the sequences of the clade 1, the clade that CVB3/GA located on the phylogenetic tree, but was more similar to the codon usage pattern of Clade 2. In addition, the CAI and gtAI values of virulent strains within Clade 1, such as CVB3/28, were not significantly different from those of CVB3/GA, suggesting that efficient ribosome engaging is

important in the virulence determination of CVB3 but the further improvement of translation efficiency through adaptation to the host tRNA pool is not necessary. These results improve the current understanding of the determinants of CVB3 virulence and will enable the search for new therapeutic targets in the future.