**ABSTRACT**

The degeneracy in the standard genetic code has allowed all amino acids except Met and Trp to be encoded by more than single codon. The codons encoding the same amino acid *i.e.* synonymous codons are not used with equally in the genome of most of the organisms. Some codons are used preferentially over the others, a phenomenon popular in the name of 'codon usage biases'. Molecular investigations suggest that codon usage bias is common across genomes and may aid in genome evolution in a profound manner. The codon usage is interplay of the equilibrium between natural/translational selection and mutational constraint.

The influenza viruses are a class of single stranded RNA virus with a genome of negative polarity and belong to the family *Orthomyxoviridae.* Influenza A virus (IAV) alone infects roughly 1.5 million people annually across the globe. The genome of IAV is an eight segmented negative-sense RNA, which is required to be converted into positive strand in order to replicate inside the host cells upon infection. In order to evade the host immune response, human seasonal influenza virus utilizes a unique phenomenon called antigenic drift by which it changes its antigenicity by introducing novel mutations in its surface proteins.

The replication cycle of the influenza virus depends on host machinery and the virus utilizes host cellular components for its protein synthesis. Therefore codon usage in this virus and its hosts could be expected to affect viral replication. A detailed understanding of the basic biology of this virus, especially its evolution and methods for host adaptation, is needed to prevent future pandemics.

This study seeks to understand the variation in codon usage of five IAV subtypes that have been found circulating among the humans in varying degree. Sporadic studies have been carried out to study IAV codon usage in human hosts; however, no attempt has been made to perform a comparative study involving these five subtypes together. We believe such studies

would help us gain insights into the evolutionary aspects of this immensely important viral entity that has been a serious global threat both in terms of mortality and morbidity.

Complete coding sequences (cds) of eight different genes *viz.* hemagglutinin (HA), neuraminidase (NA), nucleoprotein (NP), matrix protein (M1 and M2), polymerase acidic protein (PA) polymerase basic proteins (PB1 and PB2), belonging to five IAV subtypes were used in the present study.

Our findings suggested a low codon usage bias prevailing in the IAV genes. The RSCU analysis of the codons offered an intricate picture with the varying partiality of codons in different genes. As the viral genome is AT-rich, the IAV genes showed a preference towards A/T ending codons in majority of the cases. Ironically, the H5N1 subtypes presented slight variation from the rest of the subtypes in preference of A/G at the 3$^{rd}$ codon position as well as in amino acid usage. RSCU analysis presented variations across the subtypes with H2N2, H3N2 and H5N1 presenting different codon usage profiles from the remaining two strains. H1N1 and H1N2 showed somewhat similar codon preferences. NCG and CGN type codons were severely depleted. TTA, GGC and GGT were among the lowly used codons with slight exceptions in certain genes from H1N1 and H1N2 subtypes of IAV. The deviation of H5N1 subtype from the rest was supported by the correspondence analysis. There was a significant positive correlation between GC12 and GC3 which inferred that GC composition was a crucial factor in determining the codon usage in this virus. It appeared that a balance of mutation/selection exists in IAV, which permits the virion to re-adjust its codon usage make-up in response to different host conditions. More extensive inspection of the codon usage profile might facilitate a better comprehension of the diverse aspects of the virulence factors. This would be critical to the screening of suitable drug targets which might pave the way for the development of successful antivirals in tackling the virus.