

Base Usage and Dinucleotide Frequency of Infectious Bursal Disease Virus

Abstract

Base usage and dinucleotide frequency have been extensively studied in many eukaryotic organisms and bacteria, but not for viruses. In this paper, a comprehensive analysis of these aspects for infectious bursal disease virus (IBDV) was presented. The analysis of base usage indicated that all of the IBDV genes possess equivalent overall nucleotide distributions. However when the base usage at each codon positions was analysed by using cluster analysis, the VP5 open reading frame (ORF) formed a different cluster isolated from the other genes. The unusual base usage of VP5 ORF may indicate that the gene was originated by the virus "overprinting strategy", a strategy in which virus may create novel gene by utilizing the unused reading frames of its existing genes. Meanwhile, the GC content of the IBDV genes and the chicken's coding sequences was comparable; suggesting the virus imitation of the host to increase its translational efficiency. The analysis of dinucleotide frequency indicated that IBDV genome had dinucleotide bias: the frequencies of CpG and TpA were lower and the TpG was higher than the expected. Classical methylation pathway, a process where CpG converted to TpG, may explain the significant correlation between the CpG deficiency and TpG abundance. "Principal component analysis of the dinucleotide frequencies" (DF-PCA) was used to analyse the overall dinucleotide frequencies of IBDV genome. DF-PCA on the hypervariable region and polyprotein (VPX-VP4-VP3) gene showed that the very virulent IBDV (vvIBDV) was segregated from other strains; which meant vvIBDV had a unique dinucleotide pattern. In summary, the study of base usage and dinucleotide frequency had unravelled many overlooked genomic properties of the virus.