

比較ゲノム解析により明らかになった巨大ウイルスの
遺伝子発現制御配列の進化的起源

Evolutionary origins of regulatory motifs in giant viruses
inferred from comparative genomics

京都大学 化学研究所 バイオインフォマティクスセンター 化学生命科学領域 唐 威

研究成果概要

Giant viruses are characterized by their large double-stranded DNA genomes with high coding capacities. These viruses regulate their gene expression using cis-regulatory elements, including promoter sequences and 3'-end stem-loop structures, the latter of which function as polyadenylation signals. However, the knowledge of transcription regulation systems in giant viruses has been limited to few isolated reference viruses. The recent accumulation of viral metagenomic data now provides an opportunity to comprehensively analyze these regulatory sequences across diverse viral lineages. In this study, I performed computational analyses of cis-regulatory elements using a large dataset of 1817 giant virus genomes, spanning the phyla *Nucleocytoviricota* (1706) and *Mirusviricota* (111), to search for promoter motifs in upstream gene regions and stem-loop structures in downstream regions. The analysis of promoter motifs revealed that conserved signals are widespread in *Nucleocytoviricota*, with at least one motif type detected in 1577 genomes (92.4% of the dataset). On average, 3.1 distinct motif types were identified per genome, covering 42.7% of all upstream regions. The distribution of motif types varied significantly by taxa: the order *Pandoravirales* exhibited the highest number of motifs (average 5.1 types, covering 66.5% of upstream sequences), whereas the order *Chitovirales* showed the lowest number of motifs (average 2.0 types, covering 33.0% of upstream sequences). The analysis of 3' non-coding regions revealed that several taxa display a high prevalence of stem-loops. The viral orders whose members showed a downstream stem-loop structure in >10% of genes included *Pimascovirales* (23.5%), *Imitervirales* (20.9%), and *Asfuvirales* (10.5%). Furthermore, analysis to investigate possible presence of ancestral motifs at the root of *Nucleocytoviricota*, I compared all motifs at the order level and revealed four distinct motif clusters whose members were detected across all the six orders of the *Nucleocytoviricota* (*Algavirales*, *Asfuvirales*, *Chitovirales*, *Imitervirales*, *Pandoravirales*, and *Pimascovirales*). The most prominent of these motif clusters was characterized by a consensus sequence 'AAAATTGA' that was found at high frequency in at least four orders (*Imitervirales*, *Pimascovirales*, *Chitovirales*, and *Asfuvirales*). Given its broad phylogenetic distribution, I suggest that this motif as a plausible ancestral regulatory sequence for the phylum.