

Virus Classification using Knowledge Graph

Mikyung Je³, Nara Been¹, Myeongji Cho¹, 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*

Viral taxonomy is essential in understanding viruses and serves as an integrated catalog of their vast diversity and genetic relationships. However, the diversity of viruses is much greater than that of other organisms, with large differences in the genetic material (RNA or DNA) composition (double- or single-stranded), and orientation of the encoded genes. New viruses are constantly being discovered due to advances in science and changes in the environment. The discovery of these viruses challenges traditional virus taxonomy. In general, traditional taxonomy considers various characteristics to understand a virus's characteristics, including replication strategy, virion (virus particle) structure, host (infecting organism) range, geographic distribution, and mode of transmission. However, this attribute information may be missing or incomplete in metagenomic data sets. Therefore, virus classification requires the use of various data and knowledge resources. We focused on knowledge graphs to solve this problem. Knowledge graphs are an important AI approach that integrates different types of complex knowledge and data resources. A wide range of information, including host range, geographic distribution, and genetic characteristics related to virus classification, can be integrated and utilized. The virus knowledge graph is created by collecting data from various fields of biological knowledge. The main sources are NCBI, UniProt, ViralZone, ICTV, and other individual virus data sources. In this study, we built the Virus Classification Knowledge Graph (VCKG), a virus classification knowledge graph that helps users search and analyze information about various viruses in one place. VCKG allows for the effective utilization of a variety of information for virus classification and for exploring the characteristics of viruses based on criteria. We expect these knowledge graphs to provide new insights into virus research and classification.