NCBI Bioinformatics Resources for Influenza and Other Emerging Infectious Viruses

Yiming Bao[1*], Danso Ako-adjei[1], Olga Blinkova[1], Yuri Ostapchuk[1], Sergey Zhdanov[1], J. Rodney Brister[1]

[1.] NCBI, National Institutes of Health

Background/Objective
The rapid growth of large scale virus genome sequencing projects calls for specialized resources designed to enhance the accessibility and utility of large sequence datasets.

Method
The Influenza Virus Resource (IVR, http://www.ncbi.nlm.nih.gov/genomes/FLU/) was created at NCBI to cope with the rapid increase of influenza virus sequences. It provides a curated database that contains nucleotide, protein and coding sequences of influenza viruses in EMBL/DDBJ/GenBank databases. Metadata are extracted from sequence records and/or generated by computational analyses, and are standardized in the database. Sequences and associated metadata from the database can be searched and downloaded for further analysis. The resource also provides sequence analysis tools integrated with the database, such as multiple sequence alignment and clustering of protein sequences based on different metrics. This allows users to quickly modify a dataset to optimize the analysis. An influenza virus genome annotation tool is included in the resource to validate and predict protein sequences encoded by influenza viruses.

Result
The IVR model has been applied to other important viruses such as dengue virus, West Nile virus, MERS coronavirus and ebolavirus, which are now all under the Virus Variation Resource (http://www.ncbi.nlm.nih.gov/genome/viruses/variation/).

Conclusion
The Virus Variation Resource not only enables scientists to study genetic evolution of the viruses, but also plays an important role in disease surveillance, vaccine selection and development in diagnostic techniques.