

Biochemistry 2018: Genus specific protein patterns of viruses

Sandeep Bansode

Dr D Y Patil Biotechnology and Bioinformatics Institute, India

Email: sandeep.bansode@dpu.edu.in

Abstract

In the era of emerging and re-emerging viral infections, diagnostics and its allied fields have a major role to play in combating the diseases. Enormous amount of the molecular sequence data available in the public domain has the potential to contribute in a major way in the development of novel diagnostic tools. One of the prerequisites for such a study is the identification of signature sequences i.e., small stretches of protein/nucleotide sequences that are unique to a given family/genus/organism. There exist several resources in the public domain archiving signature sequences of proteins based on sequence identity/ similarity. However, these resources do not take into account the taxonomic information which has a significant role to play in viral diagnostics. The present study is an effort to explicitly take into account the taxonomic information and thereby derive genus-specific signature sequences of viral proteins. The preliminary data for obtaining patterns viz., multiple sequence alignment (MSA) is obtained from VirGen database. An in-house developed perl script is used to derive the patterns from the MSA. The patterns are then validated by search against the non-redundant protein sequence database at NCBI, thereby enabling the computation of their sensitivity and specificity. Such a validation requires datasets pertaining to true-positives and true-negatives. True-positive dataset is obtained from the taxonomy database at NCBI by formulating an Entrez query such that the total number of species belonging to a given genus is retrieved. The true-negative dataset constituted of any protein sequence that belongs to genus other than the one in question. Of the 262 proteins belonging to 19 families (RNA viruses) in VirGen, patterns could be detected for 125 proteins, all of which clearly distinguished true-positives and false-positive sequences. These patterns when mapped onto their corresponding 3D structures (25 unique entries of Protein Data Bank) are found to be part of important functional regions like active site and dimerisation interface. The unique viral signature sequences/peptides thus obtained have applications not only in detection assays and as therapeutics but also can serve as putative targets for viral vaccines.

This work is partially presented at International Conference on Biochemistry, Proteomics & Bioinformatics May 16-17, 2018 Singapore