Journal of Biochemistry and Physiology

Genus specifi c protein patterns of viruses

Sandeep Bansode

Dr D Y Patil Biotechnology and Bioinformatics Institute, India

Abstract:

In the era of emerging and re-emerging viral infections, diagnostics and its allied fi elds 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 perquisites for such a study is the identifi cation of signature sequences i.e., small stretches of protein/nucleotide sequences that are unique to a given family/genus/organism. Th ere 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 signifi cant role to play in viral diagnostics. Th e present study is an eff ort to explicitly take into account the taxonomic information and thereby derive genus-specifi c signature sequences of viral proteins. Th e 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. Th e patterns are then validated by search against the non-redundant protein sequence database at NCBI, thereby enabling the computation of their sensitivity and specifi city. 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. Th e 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. Th ese 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. Th e 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.

Biography:

Sandeep Bansode has completed his Bachelor's degree in Agricultural Science, Master's degree in Bioinformatics. He has worked as a Research Associate/Lecturer in Vidya Pratishthan School of Biotechnology, Programmer at National Institute of Virology, research Associate at National Research Centre on Plant Biotechnology and as Software Engineer at Trance Technologies. He is presently a Senior Scientist/Head of Bioinformatics at Vidya Pratishthan's School of Biotechnology, India.