g-VATA: A Grid-Enabled Application for Viral Antigenic Transmission Analysis

Tuesday 17 Mar 2015 at 14:20 (00h20')

Content:
g-VATA is a grid-enabled web-based application to study antigenic transmission of viruses from its non-human animal reservoir pool to the human host. The application only requires the viral proteome sequences in FASTA format from both the host and the reservoir as an input. The application aligns these sequences from the two sets collectively for a unified alignment position, and then separates the aligned sequences of the sets for comparative analysis position-by-position. The analysis is done through application of a k-mer window approach of size nine for immunological reasons. The detailed method is described in Tan Swan et al. (manuscript in preparation) [1]. Briefly, each of the overlapping nonamer amino acid positions of the viral proteome represent a possible core binding domain for human leukocyte antigen molecules and T-cell receptors, and are classified as four patterns of antigenic sequence motifs [2]: (1) "index", the most prevalent sequence; (2) "major" variant, the most common variant sequence; (3) "minor" variants, multiple different sequences, each with an incidence less than that of the major variant; and (4) "unique" variants, each observed only once in the alignment. These motifs describing the incidences of the nonamer sequences allow evaluation of the specificity of the sequences to the reservoir or the human host and aid in the identification of sequences highly selective to the human host. This characterization of viral antigenic transmission provides a tool for surveillance of viruses with potential for adaptation in human, and this may contribute to efforts in prevention or treatment of viral infection.

g-VATA greatly benefits from the grid back-end for massively large datasets, which is common in current times, given the low cost for sequencing small viral genomes and also the advent of affordable next-generation sequencing. For example, there are more than 400,000 sequences of Influenza A virus alone reported in public NCBI Entrez Protein Database. The g-VATA pipeline is a combination of a parallel compute intensive execution and also distributed compute less-intensive execution. The parallel compute intensive execution job is sent to a dedicated computing cluster (e.g. HPC/Grid), while the distributed compute less-intensive execution jobs are sent to Desktop Grid based on BOINC. g-VATA is built on top of gUSE/WS-PGRADE [3-7] platform that make a complex process pipeline become more manageable and make distributing the job to the distributed computing infrastructure (e.g. HPC, Grid).
g-VATA is publicly available at https://bioinfo.perdanauniversity.edu.my/liferay-portal-6.1.0/. g-VATA was applied for transmission analysis of H7N9 influenza A virus, the recent pandemic threat, with a longstanding global distribution in multiple avian hosts [8]. This resulted in the identification of nonamer sequences that were selective for the human host (Tan Swan et al., manuscript in preparation) [1].

Reference:
1. Tan S, Mohamed NE, Raman HSA, Khan AM, and August JT (2015). H7N9 Influenza Virus Mutations as Signatures of Human Transmission. [manuscript in preparation]

Primary authors : Mr. SJAUGI, Muhammad Farhan (Perdana University - Centre for Bioinformatics) ; Dr. KHAN, Mohammad Asif (Perdana University - Centre for Bioinformatics)

Co-authors : Ms. TAN, Swan (Perdana University - Centre for Bioinformatics) ; Ms. RAMAN, Hadia Syahirah (Perdana University - Centre for Bioinformatics) ; Ms. MOHAMED, Nik Elena (Perdana University - Centre for Bioinformatics)

Presenter : Mr. SJAUGI, Muhammad Farhan (Perdana University - Centre for Bioinformatics)

Session classification : Biomedicine & Life Science

Track classification : Biomedicine & Life Sciences Applications

Type : Oral