EGEE USER FORUM

Contribution ID: **158**   Type: **Oral**

Grid-based International Network for Flu Observation

Tuesday, 13 April 2010 15:00 (20 minutes)

Since the H1N1 outbreak lately, there has been a worldwide effort to isolate and sequence flu virus genomes. Specimens with a positive result are sequenced and deposited in influenza databases. The present EUAsiaGrid application, called g-INFO (Grid-based International Network for Flu Observation), shows the integration of existing data sources towards a global surveillance network for molecular epidemiology, based on Service Oriented Architecture and Grid technologies. Its relevance is being tested through the current H1N1 outbreak.

Impact

Results are made available to the research community in the corresponding website: http://g-info.healthgrid.org/, providing a real identity card of the concerned virus strains. Thanks to the molecular specificities highlighted (site for protease cleavage, glycosylation sites, epitopes and binding site), experts have in their possession promptly elements allowing them to take the most appropriate decisions relating to the transmission and the geographical expansion of the epidemic.

Keywords

flu, surveillance network, grid, Service Oriented Architecture, epidemiology

URL for further information

http://g-info.healthgrid.org/

Detailed analysis

The current prototype is using the NCBI database (National Center for Biotechnology Information). Everyday the NCBI-FTP server is updated with new sequences of H1N1 segments, with 7 files: sequences of nucleotide, protein and coding region and corresponding metadata. A grid database (AMGA) is populated with such data through an automatic synchronisation. The pipeline starts with a sequence preparation in correct format, then a multiple alignment using Muscle followed by a curation with G-blocks to identify conserved blocks. From this step a phylogenetic analysis is performed to obtain a branching diagram. Based on virologists requirements, the selected sequences from this diagram are subjected to further analysis in order to identify key features related to pathogenicity such as the site for protease cleavage, the glycosylation sites, the epitopes or the binding site.

Conclusions and Future Work

Future developments will involve additional influenza databases within the network. By being constantly attentive to the virologists and epidemiologists requirements, the data processing can be adapted accordingly.

The final goal is to have the grid-based surveillance network ready to impact the next pandemics.

Primary authors: Ms DA COSTA, Ana Lucia (HealthGrid); Mr DOAN TRUNG, Tung (IFI, Vietnam)

Co-authors: Mr BERNARD, Aurélien (CNRS-IN2P3, France); Dr NGUYEN HONG, Quang (IFI, Vietnam); Dr LE, Thanh-hoa (Institute of Biotechnology, Vietnam); Mr BRETON, Vincent (CNRS-IN2P3, France); Mr LEGRE, Yannick (HeallthGrid, France)

Presenter: Ms DA COSTA, Ana Lucia (HealthGrid)

Session Classification: Bioinformatics

Track Classification: Scientific results obtained using distributed computing technologies