ViPR News

The ViPR team presented a poster at the American Society for Microbiology Biodefense meeting at the Washington Omni in Washington DC, February 6-9, 2011.

Metadata

Comparative Genomics Tool

There are new enhancements to the Metadata Comparative Genomics tool (MCG) introduced in January 2011. With this tool you can use metadata to partition sequences from a single viral family into 2 or more subsets. Multiple sequence alignments are built, and statistical tests along the alignment determine which position(s) differ significantly between groups. MGC is now accessible from genome search results and working sets. The finished multiple sequence alignment file can be downloaded.

Sequence Feature

Variant Type Analysis

A new capability, sequence feature variant type (SFVT) analysis, has been implemented as a beta release. For a description of what sequence features are, how SFVTs are determined, and how they can be used in viral pathogen research, see an SFVT home page (Dengue). SFVT data and analysis support has been generated for some Pox viruses, Dengue viruses, and Hepatitis C viruses. More will be added in future releases of ViPR. SFVTs can be used in statistical analyses to determine the relationship between sequence variations and virus phenotypic characteristics. Sequence Features are grouped by protein and categorized as structural, functional, immune epitopes and sequence variations. A tool on the Details page, Find a VT(s), helps you find strains bearing a specific variant type. The domain experts in the table to the right validated the sequence features for accuracy.

Phylogenetic Trees

ViPR now provides a utility allowing interactive generation of phylogenetic trees. You can build a tree using search results, working sets, or external files/sequences. Selected sequences are first aligned. You can then generate a tree with either the Quick-tree or Custom tree options. Quick-tree uses the FastME algorithm, a fast, distance-based approach good for generating trees from datasets containing 1) more than 1,000 sequences of short or medium length, 2) more than 100 very long sequences, or 3) to reconstruct a “quick and dirty” tree. When Custom tree is used for nucleotides, the Model Compare feature helps you choose the evolutionary model that best fits your data. The Custom tree option uses the PhyML algorithm to infer a more evolutionarily-accurate phylogenetic topology by applying a substitution model to the nucleotide sequences. This algorithm is best applied to datasets containing 1) fewer than 100 very long sequences, 2) between 100 and 1,000 small or medium length sequences. ViPR uses a customized version of the Archeopteryx tree visualization software that includes a ViPR unique tree decoration capability which allows you to color trees based on isolate characteristics such as year or country of isolation. Tree leaf coloring or decoration can be changed from one characteristic to another with a single mouse click. Dynamic hiding, which can be turned on or off, automatically limits the number of node labels shown based on tree density so that none overlap.

Searches and Data

The speed of QuickText Searches for genomes and genes/proteins has been greatly improved with implementation of the Oracle text indexing feature.

We have integrated experimental epitopes with corresponding 3D structures. When you encounter an experimentally determined epitope associated with a protein for which the structure has been determined, you can go from the Experimentally Determined Epitope Details page to the Protein Structure table. Clicking a View Structure link opens the corresponding structural file in the J Mol viewer where the epitope can be highlighted directly on the structure.

Dengue Virus

Through collaboration with the Broad Institute genome sequencing center and the Genome Resources in Dengue (GRID) consortium, significant quantities of clinical metadata collected from patients from whom Dengue virus has been isolated and sequenced is now available in ViPR. This metadata falls into 3 categories: available immediately upon submission, available upon expiration of a grace period, and available upon request. The first 2 categories may be viewed in ViPR while the third must be requested from the GRID consortium using a link within ViPR. To view this data, select a cohort from the Genome search and click the information icon on the search result to view the Genome Detail page with accompanying clinical data. ViPR facilitates submitting requests for available upon request data to the clinical investigators.

Epidemiological datasets for Dengue virus, assembled under the VBRC project, are now available in ViPR. DengueNet, the Pan American Health Organization (PAHO), and WHO/UNICEF are accessed via the top-level Resources menu (Epidemiological datasets). Datasets are organized by country and year and a heat map is used to rank incidence.

Sequence Feature Expert Panel

<table>
<thead>
<tr>
<th>Virus</th>
<th>Name of scientist</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dengue virus</td>
<td>Dr. Richard Kuhn</td>
<td>Purdue University</td>
</tr>
<tr>
<td>Dengue virus</td>
<td>Dr. Radhakrishnan Padmanabhan</td>
<td>Georgetown University</td>
</tr>
<tr>
<td>Hepatitis C virus</td>
<td>Dr. Carla Kuiken</td>
<td>Los Alamos National Laboratory</td>
</tr>
<tr>
<td>Pox virus</td>
<td>Dr. Elliot Lefkowitz</td>
<td>University of Alabama at Birmingham</td>
</tr>
<tr>
<td>Pox virus</td>
<td>Dr. Bernie Moss</td>
<td>National Institute of Allergy and Infectious Diseases, National Institutes of Health</td>
</tr>
</tbody>
</table>