Volume 8, Issue 12 December 2015

Bioinformatics up to Date

(Bioinformatics Infrastructure Facility, Biotechnology Division)
North-East Institute of Science & Technology
Jorhat - 785 006, Assam

Inside…..

About Us

The Bioinformatics Infrastructure Facility (BIF) at Biotechnology division, CSIR NEIST, Jorhat runs under the Biotechnology Information System Network (BTISnet) programme of DBT, Ministry of Science & Technology, and Government of India. The Centre was established on 2nd February, 2008 to promote innovation in Biological research and education through Bioinformatics accomplishment. The main goal is to facilitate and expose students and researchers from different academic institutions of North East India in Bioinformatics. The center conduct training and workshops for enlightening the use of bioinformatics applications in biological research and development. The Centre has access to global information through 24 hour high speed internet facility, and also e-journal facilities with DeLCON, Science Direct etc. To date the Centre has profoundly extended support in R & D work with a great intensity to different biological discipline including medicinal chemistry, computer aided drug design, genomics and proteomic data analysis etc.

Our Focus

Workshop Cum Training on Biological Data Analysis Through Computational Approach

A 3 days workshop cum hands on training on biological data analysis has been successfully organized at bioinformatics centre, BSTD, CSIR-NEIST under the successful coordination of Dr Ratul Saikia, Senior Scientist and Coordinator of BIF Centre, CSIR-NEIST. The program was held on 8-10 December, 2015 by the support of DBT, Govt of India. There are total 43 participants joined in the workshop coming from different states of North East India viz. Arunachal Pradesh, Manipur, Nagaland etc. The workshop mainly focused on the recent trend on analysis of short-read Next Gene Sequencing data analysis as well as the genomic and transcriptomic data analysis. The workshop gave the participants a deeper understanding of Next-Generation Sequencing (NGS) analysis with a special focus on bioinformatics approaches.

Genome-Wide Analysis of the SBP-Box Gene Family in Chinese Cabbage

The SQUAMOSA PROMOTER BINDING PROTEIN (SBP)-box gene family contains highly conserved plant-specific transcription factors that play an important role in plant development, especially in flowering. A study carried out by researcher from China, aimed to characterize the SBP-box transcription factor genes in Chinese cabbage. Twenty-nine SBP-box genes were identified in the Chinese cabbage genome and classified into six groups. The analysis report published in journal *Genome* by NRC research press 2015.

The researchers identified 23 orthologous and 5 co-orthologous SBP-box gene pairs between Chines cabbage and *Arabidopsis*. An interaction network among these genes was constructed and sixteen SBP-box genes were expressed more abundantly in flowers than in other tissues, suggesting their involvement in flowering. They find the MiR156/157 family members may regulate the coding regions or 3′-UTR regions of Chinese cabbage SBP-box genes. As SBP-box genes were found to potentially participate in some plant development pathways, quantitative real-time PCR analysis was performed and showed that Chinese cabbage SBP-box genes were also sensitive to the exogenous hormones methyl jasmonic acid and salicylic acid. The SBP-box genes have undergone gene duplication and loss, evolving a more refined regulation for diverse stimulation in plant tissues. Their comprehensive genome-wide analysis provides insights into the SBP-box gene family of Chinese cabbage.

[Source: Genome-wide analysis of the SBP-box gene family in Chinese cabbage (Brassica rapa subsp. pekinensis), Genome - 58(11):pp. 463-477]

Digging Deeper into DNA: Efficient Method to Sequence Chloroplast Genomes

Researchers from National Institute of Agricultural Research in Uruguay have developed a chloroplast genome-sequencing strategy that provide a time- and cost-efficient method to assemble a chloroplast genome using whole-genome sequencing. Using the new method, the researchers extracted whole-genome sequence data from red rice (*Oryza sativa L*) and produced a complete chloroplast genome, which is now available on GenBank. The study published in a recent issue of *Applications in Plant Sciences,* 2015.

According to the researchers, several fragments and, in some cases, nearly entire copies of the chloroplast genome may be found within the nuclear genomes of plants. This is because genetic material from chloroplasts has been continuously transferring to the nucleus through years of evolution. Chloroplasts are understood to have originated from prokaryotes that were engulfed by eukaryotes millions of years ago.

The new method developed by Garaycochea and his colleagues will enable researchers to strategically analyze this whole-genome sequence data and assemble the chloroplast genome for their plant of interest. It is less costly and less time consuming than other methods. Certain tedious lab procedures such as prior plastid DNA isolation, plastid DNA enrichment, and reliance on a reference genome are not required. The researchers have developed a bioinformatics strategy to recover and assemble a chloroplast genome using data derived from low-coverage 454 GS FLX/ Roche whole-genome sequencing. A comparative genomics approach was applied to obtain the complete chloroplast genome from a weedy biotype of rice from Uruguay.

[Source: A strategy to recover a high-quality, complete plastid sequence from low-coverage whole-genome sequencing.Applications in Plant Sciences 3(10): 1500022]

NanoOK: Quality Control for Portable, Rapid, Low-Cost DNA Sequencing

Scientists at The Genome Analysis Centre (TGAC) have been putting Oxford Nanopore's MinION sequencer through its paces with an open-source, sequence alignment-based genome analysis tool called 'NanoOK'. NanoOK is the first open-source tool that provides comprehensive alignment-based quality control and error profile analysis for the MinION platform. The NanoOK's main output is a detailed PDF report featuring graphs and tables of sample analysis data. Individual graphs are also available to include in publications and presentations and the raw data is available for users to perform additional custom analysis. The work is published in paper, titled: NanoOK: Multi-reference alignment analysis of nanopore sequencing data, quality and error profiles; in Oxford Journals *Bioinformatics* 2015.

NanoOK's comprehensive alignment-based error profiling enables researchers to understand data quality, the effect of different alignment tools and to understand the effect of updates to the MinION's chemistry and software. The tool currently supports four popular Nanopore aligners but is easily extensible through a Java programming interface. It also handles metagenomic sampling gracefully, due to support for multiple reference sequences and the output report PDF benefits from programming language R's graphical capabilities, for at-a-glance reporting of large data volumes. NanoOK will extract reads as FASTA or FASTQ files, align them (with a choice of alignment tools), then generate a comprehensive multi-page PDF report containing yield, accuracy and quality analysis. NanoOK has a number of dependencies - Perl, LaTeX, R and an alignment tool - which means it works best on Linux and Mac OS platforms. It can be downloaded from https://github.com/TGAC/NanoOK.

[Source: http://www.tgac.ac.uk/news/252/68/NanoOK-Quality-Control-for-portable-rapid-low-cost-DNA-sequencing/]

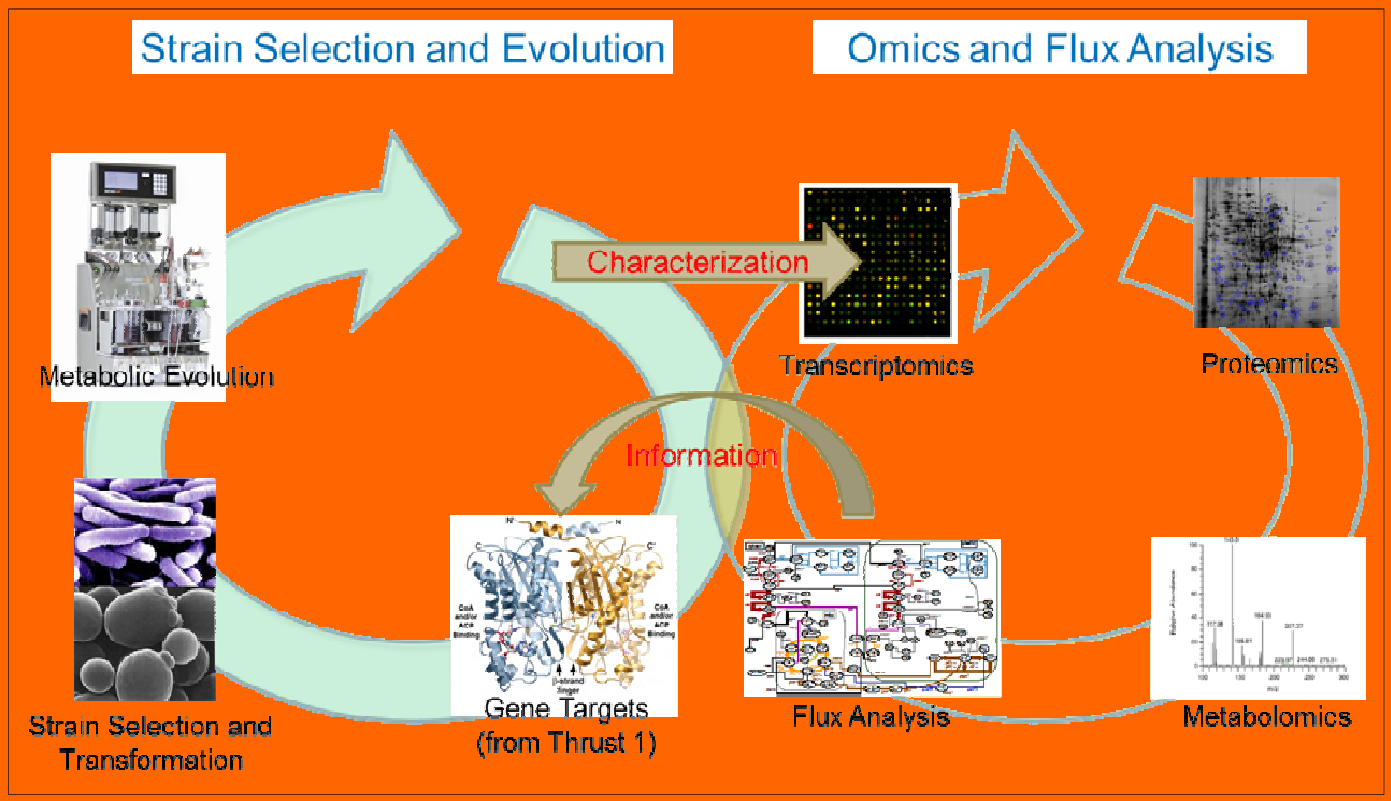
CerebralWeb: A Cytoscape.js Plug-in to Visualize Networks

CerebralWeb is a light-weight JavaScript plug-in that extends Cytoscape.js to enable fast and interactive visualisation of molecular interaction networks stratified based on subcellular localisation or other custom annotation. CerebralWeb also supports the automatic retrieval of compatible localisations for human, mouse and bovine genes via a Web Service and also supports the automated parsing of Cytoscape compatible XGMML network files.

CerebralWeb can be used to visualize any large network (up to ~2000 nodes for optimal performance given current JavaScript engines and hardware speeds) where the nodes have an attribute value matching one of Cerebral's subcellular localizations. The user may specify subcellular localization via the 'localization' node attribute in an uploaded network file. Alternatively, CerebralWeb may fetch this information in real-time from InnateDB.com via a web service that provides Cerebral subcellular localization annotations for human, mouse and bovine genes. To access this feature of CerebralWeb the user need only provide Ensembl, Entrez Gene, UniProt or InnateDB identifiers as node attributes. The web service query will then look-up the gene ontology. CerebralWeb currently supports embedded network visualization on the InnateDB (www.innatedb.com) and Allergy and Asthma Portal (allergen.innatedb.com) database and analysis resources.

[Source: http://www.ncbi.nlm.nih.gov/pubmed/25953080]

Bioinfo. Animation



Patent News

System and Method for Creating Dynamic Workflows Using Web Service Signature Matching

US20050234964A1

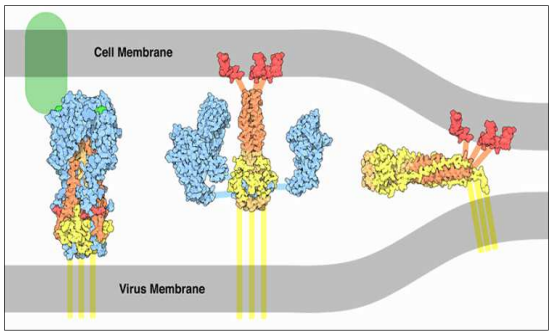
Inventor : Virinder Batra *et al.*

Abstract

A system and method for dynamically implementing a chain of Web services from a client on the World Wide Web to execute a workflow. The described system includes: a database for storing a list of available Web services, wherein each listed Web service includes a description of a task performed by the Web service, and an input and output signature of the Web service; and a selecting system for forming the chain of Web services by selecting a Web service for each of a plurality of tasks in the workflow, wherein the selecting system matches input and output signatures to ensure that each selected Web service is compatible with adjacent Web services in the chain of Web services.

Hemagglutinin

Hemagglutinin is a spike-shaped protein that extends from the surface of the Influenza virus. Hemagglutinin is composed of two different types of chains, one playing targeting mechanism: they search for specific sugar chains on our cellular proteins. When they find the proper one, Hemagglutinin binds to the cell and the orange chains initiate the attack. The name Hemagglutinin refers to the ability of influenza to agglutinate red blood cells: the virus is covered with many hemagglutinin molecules, which together can glue many red blood cells together into a visible clump.



Hemagglutinin is a deadly molecular machine that targets and attacks cells. This occurs in several steps. First, the three binding sites at the top of the spike bind to sugars on cellular proteins, shown in green at the top left. Then, the whole virus is carried inside the cell into the endosome and the cell adds acid, which normally digests the stuff inside the endosome. In acid, hemagglutinin unfolds and then refolds into an entirely different shape. The portions shown in orange and red are normally folded against the protein, but in acid, they pop out and point upward. The red portion, termed the fusion peptide, has a strong affinity for membranes, so it inserts into the cell membrane and locks the virus to the cell.

[Source: http://pdb101.rcsb.org/motm/motm-by-category]

Upcoming Events

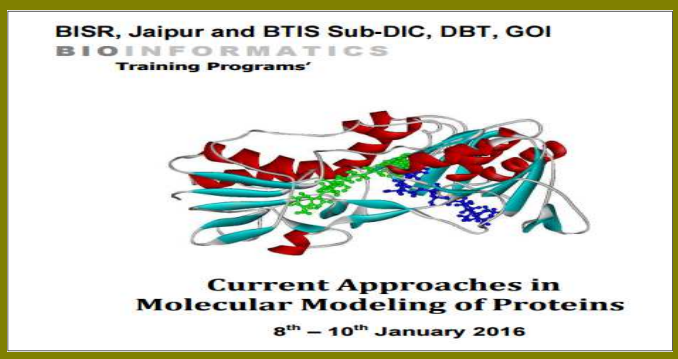
International Conference on Bioinformatics and Systems Biology (BSB-2016)"

International Conference on Bioinformatics and Systems Biology
4th to 6th March 2016

Department of Applied Sciences,
Indian Institute of Information Technology-Allahabad

Current Approaches in Molecular Modeling of Proteins

8 th – 10th January 2016
Birla Institute of Scientific Research Jaipur



Kindly send us your feedback to

Dr Ratul Saikia, Robin Das
BIF Center, Biotechnology Group, BSTD
CSIR-North East Institute of Science and Technology, Jorhat, Assam
E-mail: rsaikia19@gmail.com, robindas460@gmail.com