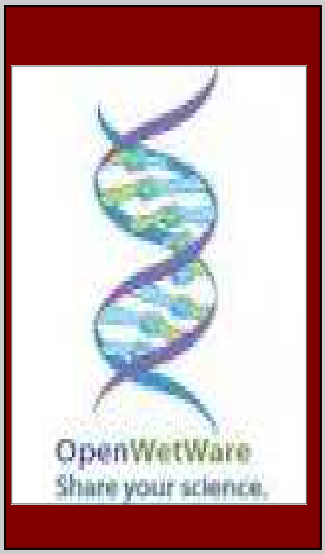
Volume 9, Issue 3 March 2016

Bioinformatics up to Date

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

Inside…..



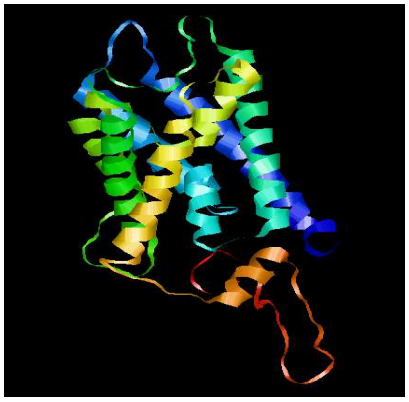
About Us

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 work-shops 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

Pharmacokinetic evaluation of bioactive compounds present in Indian Bay leaf against Human Aquaporin2

The bay leaf, *Cinnamomum tamala* found is a most widely used plant by various ethnic groups in north east (NE) India. There are more than sixty different types oil extracts reported in *Cinnamomum tamala* or Indian Bay leaf plant by various researcher that have anti-diabetic, anti-cough effects. The Aquaporin2 (AQP2) a water channel protein which dysfunction or mutations enhances the failure in water homeostasis in body that leads to nephrogenic diabetes insipidus. Therefore we have carried out an in-silico interaction study with some filtrated compounds of bay leaf against the Aquaporin 2 protein so that the protein will regain its normal function. Due to lack of experimental structure, we have generated a 3D structure of this AQP2 protein through multiple templates, then refined the structure and assessed. *In-silico* docking and pharmacokinetic evaluations have performed with the bay leaf compounds where most of the compounds showed greater results. It could be suggest that the bay leaf compounds may become potential drug candidate for diabetes insipidus in near future.



Draft Genome Sequences of Fungus *Aspergillus calidoustus*

A team from Systems Biology/Bioinformatics, Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, Germany has submitted the draft genome sequence of *Aspergillus calidoustus* (strain SF006504). The work published in s. Genome Announc 4(2):e00102-16.

The functional annotation of *A. calidoustus* predicts a relatively large number of secondary metabolite gene clusters. Aspergillus is the most studied filamentous genus in the ascomycetes division. It contains well-known human pathogens, fermentation agents of Asian food (e.g., A. oryzae), and different industrial producers (e.g., *A. niger* and *A flavus*). The worldwide distributed *Aspergillus calidoustus* was recently separated from the mesophilic species *A. ustus.*

According to the team, the DNA-sequencing resulted in 59,066,664 raw reads, where 50,376,036 reads passed in quality-filter (estimated genome coverage, 300-fold) and have been used for genome assembly. The resulting assembly consists of 78 scaffolds and 41.1 Mbp (N50 3.2Mbp; N90 493 kbp). The total GC content was 51%. RNA sequencing resulted in a total of 393,543,839 raw reads and 352,698,587 preprocessed reads (estimated genome coverage, 850-fold). The final structural gene prediction resulted in 15,139 gene models and 15,537 transcripts.

The genome project was uploaded to DDBJ/ENA/GenBank and is available under accession numbers CDMC01000001 to CDMC01000078. This paper describes the first version of the genome. Genome data and additional information are also available at the HKI Genome Resource (http://www.genome-resource.de/). The presented genome sequence builds the basis for further genome mining.

[Sources: Draft genome sequences of fungus Aspergillus calidoustus. Genome Announc 4(2):e00102-16. doi:10.1128/genomeA.00102-16]

Glycosyltransferases Identified as Significant Mutational Targets in Colon Cancer

In a new study published in Scientific Reports, an online journal of the Nature Publishing Group, scientists at Case Western Reserve University School of Medicine have successfully characterized the mutational landscapes of glycosylation-associated genes in colon cancer, identifying three glycosyltransferases as significant mutational targets in cancer. These findings are significant and suggest the functionally deleterious mutations in glycosyltransferase genes and contribution to the pathogenesis of molecular subsets of colon and other gastrointestinal malignancies.

Dr. Kishore Guda, assistant professor of general medical sciences (oncology) at the School of Medicine, led this critical research, involving the targeted re-sequencing of 430 glycosylation-associated genes and matched primary tumour tissues. Through this process, Guda and his team identified three glycosyltransferases (B3GNT2, B4GALT2, ST6GALNAC2) as significant mutational targets in CRCs. Analysis of independent large-scale tumour tissue datasets confirmed recurrent mutations within these genes in colon and other gastrointestinal cancers.

The study lays important groundwork for the future characterization of these glycosyltransferases that may provide additional insights into the biologic role of these genes in colon cancer progression.

[Biochemical and functional characterization of glycosylation-associated mutational landscapes in colon cancer. *Venkitachalam, S et al. Scientific Reports (23 March, 2016)*]

EaSeq: Bioinformatics Platform

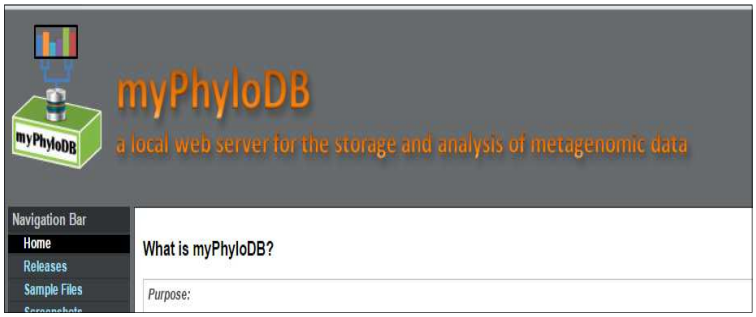
EaSeq is an agile computational environment which provides extensive interactive visualization, combines the exploratory power of genome browsers with a comprehensive set of user-friendly tools for genome-wide abstraction and visualization, and also allows experimentalists to easily extract knowledge from hundreds of genome-wide data-sets. Combined with a comprehensive toolset, this can accelerate genome-wide interpretation and understanding of genome-wide sequencing data – mainly Chromatin Immuno precipitation Sequencing (ChIP-seq) data.

EaSeq contains a variety of analysis and visualization tools, such as peak-finding, quantitation, normalization, clustering, distance analysis, randomization, scoring etc., so the amount of file transfers and format conversions between different tools is reduced. EaSeq is freely available and works on a standard personal computer, can substantially increase the throughput of many analysis workflows, facilitate transparency and reproducibility by automatically documenting and organizing analyses, and enable a broader group of scientists to gain insights from ChIP-seq data. EaSeq offers a far more visual and intuitive alternative, which makes it possible for biomedical researchers to study and test hypotheses using their own data. This means that instead of waiting for weeks for others to carry out an analysis, researchers will be able to perform the analyses themselves in a matter of hours.

[http://www.nature.com/nsmb/journal/vaop/ncurrent/full/nsmb.3180.html]

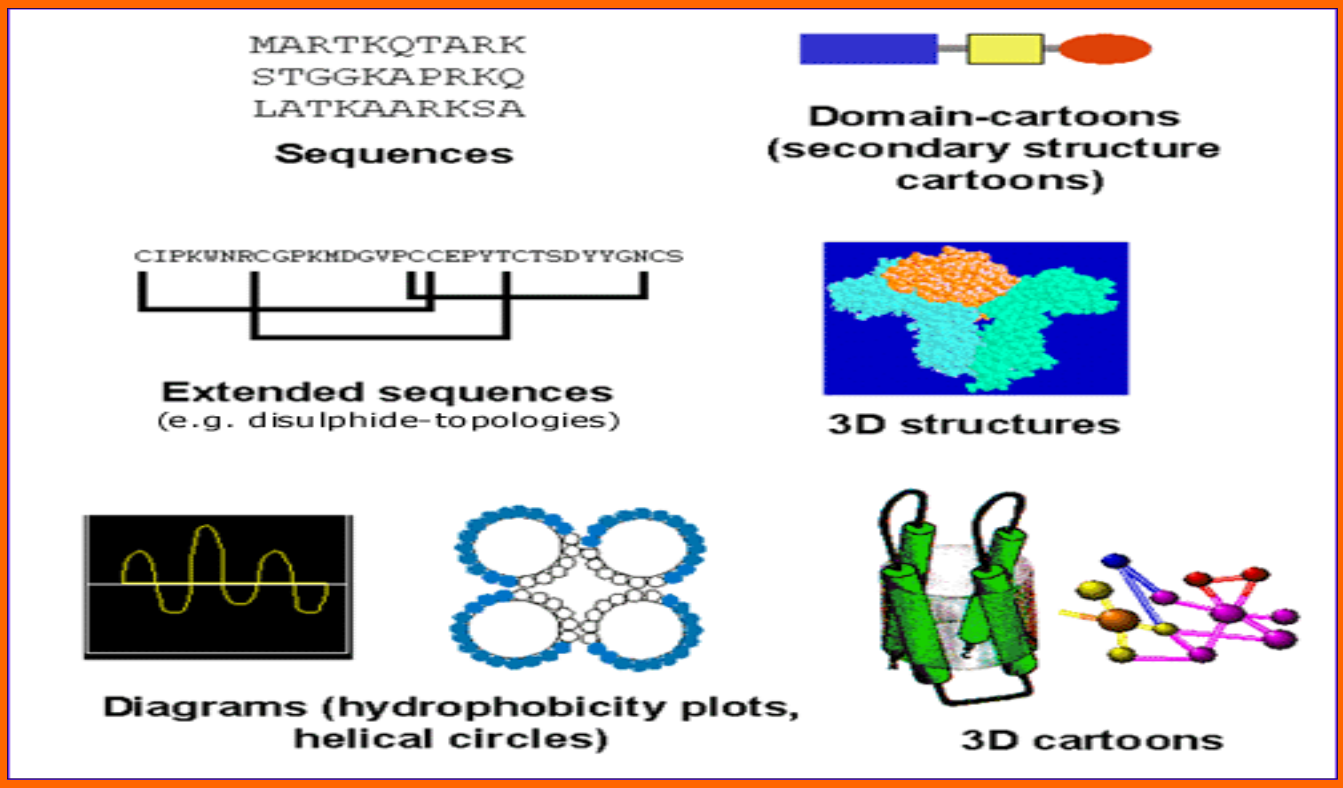
myPhyloDB: a local web server for the storage and analysis of metagenomic data

myPhyloDB v.1.1.2 is a user-friendly personal database with a browser-interface designed to facilitate the storage, processing, analysis, and distribution of microbial community populations (e.g. 16S metagenomics data). MyPhyloDB archives raw sequencing files, and allows for easy selection of project(s)/sample(s) of any combination from all available data in the database. The data processing capabilities of myPhyloDB are also flexible enough to allow the upload and storage of pre-processed data, or use the built-in Mothur pipeline to automate the processing of raw sequencing data. myPhyloDB provides several analytical (e.g. analysis of covariance, t-tests, linear regression, differential abundance (DESeq2), and principal coordinates analysis (PCoA)) and normalization (rarefaction, DESeq2, and proportion) tools for the comparative analysis of taxonomic abundance, species richness and species diversity for projects of various types (e.g. humanassociated, human gut microbiome, air, soil, and water) for any taxonomic level(s) desired. It is a local web-server, users can quickly distribute data between colleagues and end-users by simply granting others access to their personal myPhyloDB database. The database is available at http://www.ars.usda.gov/services/soft ware/download.htm?softwareid¼472.



[source: http://www.myphylodb.org/]

Bioinfo. Animation



Types of data

Patent

Method and system for providing a probe array chip design database

US6188783B1

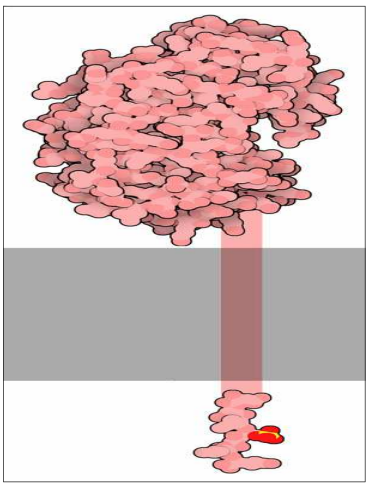
Inventor: David J. Balaban Earl A. Hubbell Michael P. Mittmann Gloria Cheung Josie Dai

Abstract

Systems and method for organizing information relating to the design of polymer probe array chips including oligonucleotide array chips. A database model is provided which organizes information interrelating probes on a chip, genomic items investigated by the chip, and sequence information relating to the design of the chip. The model is readily translatable into database languages such as SQL. The database model scales to permit storage of information about large numbers of chips having complex designs.

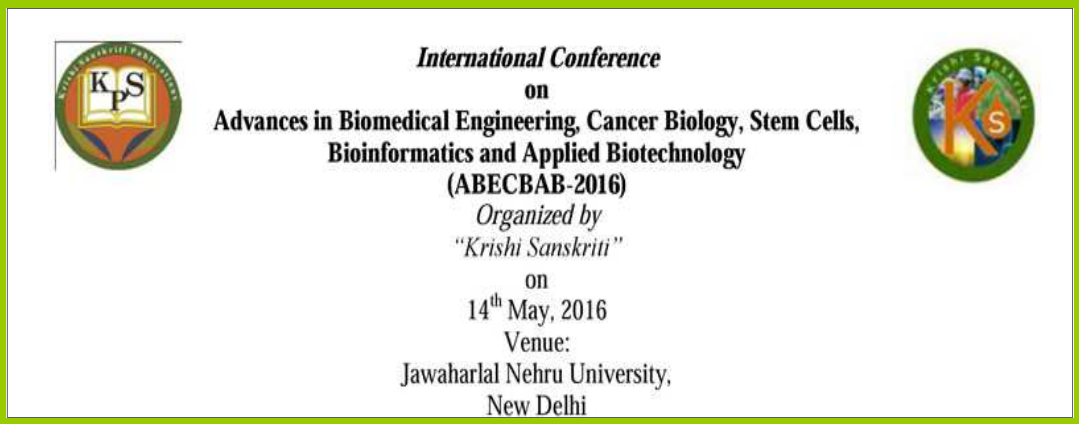
Beta-secretase

Beta-secretase, also known as BACE1 or memapsin-2, is a protease that makes specific cuts during the maturation of some protein chains. It is normally found in the endoplasmic reticulum and the Golgi, where it trims a few proteins that are particularly important in neural function. These include neuregulin, a protein that helps control the formation of myelin sheaths around nerve axons, and voltage-gated sodium channels, which are important for the transmission of nerve signals.



Beta-secretase is very similar to the digestive protease pepsin. Like pepsin, it has a deep active site cleft that grips protein chains, and a pair of aspartate amino acids that make the cut. Beta-secretase is different, however, in that it has a long tail that tethers the enzyme to the membrane surface. This tail localizes the enzyme in the proper place, so that it doesn't float freely through the cell and wreak havoc on other proteins. Two structures in the PDB show two portions of the enzyme: 1sgz (at the top) is the catalytic domain that cleaves proteins, and 1py1 contains a small portion of the chain on the other side of the membrane (at the bottom) which is important for regulating the activity of the enzyme and directing it towards the proper site in the cell.

Upcoming Events



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