Volume 9, Issue 9 September 2016

Bioinformatics up to Date

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

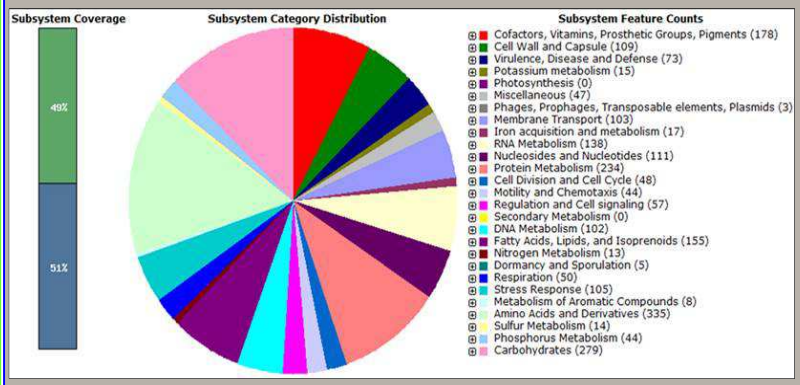
Inside…..

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 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.

Draft Genome Sequence of Carotenoid Producing Yellow Pigmented *Planococcus maritimus* MKU009

Planococcus maritimus MKU009 is a Gram positive cocci and a moderate halophilic bacterium isolated from marine water of Pichavaram, South East Coast of India. Here we report the draft



genome of *Planococcus maritimus* MKU009 with a total size of 3,251,644 bp with N50 value of 1681571 bp. The overall G+C content of the genome was 47.27%. The carotenoid producing crtN, crtB, crtP and crtI genes were located within the first contig of the genome assembly. This genome source will provide insights into functional genomics of carotenoid production and metabolic engineering. The whole-genome shotgun assembly has been deposited at GenBank under the accession no LTZG00000000. Details about the *Planococcus maritimus* MKU009 genome sequencing bioproject, biosample and raw reads submitted to Sequence Read Archive (SRA) were deposited with the following accession numbers: PRJNA313006, SAMN04511041, SRX1598158.

MAPseq (Multiplexed Analysis of Projections by Sequencing

A new computational method called MAPseq (Multiplexed Analysis of Projections by Sequencing) was developed by neuroscientist Anthony Zador, M.D., Ph.D., professor at Cold Spring Harbor Laboratory, which makes it possible to trace the long-range projections of large numbers of individual neurons from a specific region or regions to wherever they lead in the brain in a single experiment. The method is very less expensive, labour-intensive and time-consuming than current mapping technologies allow. The study published in journal *Neuron* in August, 2016.

Although a number of important brain-mapping projects are now under way, all of these efforts to obtain "connectomes," or wiring maps, rely upon microscopes and related optical equipment to trace the myriad thread-like projections that link neurons to other neurons, near and far. For the first time ever, MAPseq "converts the task of brain mapping into one of RNA sequencing," says its inventor, Anthony Zador. MAPseq uses RNA sequencing to rapidly and inexpensively find the diverse destinations of thousands of neurons in a single experiment in a single animal.

MAPseq differs from so-called "bulk tracing" methods now in common use, in which a marker typically a fluorescent protein is expressed by neurons and carried along their axons. In each neuron, it travels to the point where the axon forms a synapse with a projection from another neuron.

[source: Neuron. 2016 Sep 7;91(5):975-87. doi: 10.1016/j.neuron.2016.07.036. Epub 2016 Aug 18]

LocSigDB

LocSigDB is a manually curated database of experimental localization signals for eight distinct subcellular locations; primarily in a eukaryotic cell with brief coverage of bacterial proteins. Proteins must be localized at their appropriate subcellular compartment to perform their desired function. Mislocalization of proteins to unintended locations is a causative factor for many human diseases; therefore, collection of known sorting signals will help support many important areas of biomedical research. By performing a set of 533 experimentally determined localization signals making LocSigDB the most comprehensive compendium of localization signals, to date. Each signal in the LocSigDB is annotated with its localization, source, PubMed references and is linked to the proteins in UniProt database along with the organism information that contain the same amino acid pattern as the given signal. From LocSigDB webserver, users can download the whole database or browse/search for data using an intuitive query interface.

[source: Database (Oxford). 2015 Feb 27;2015. pii: bav003. doi: 10.1093/database/bav003]

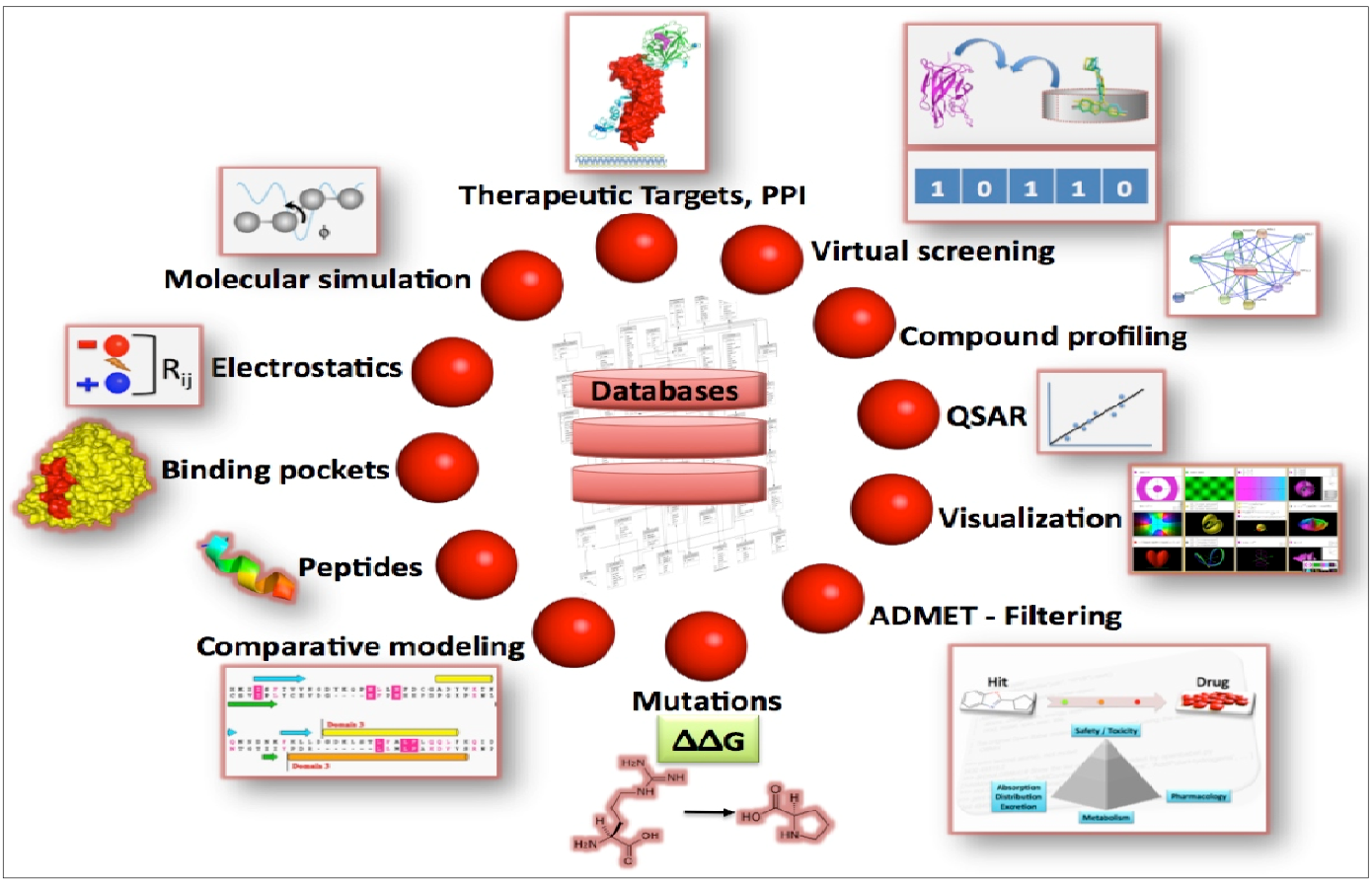
UniCon3D

A de novo protein structure prediction method that performs stepwise synthesis and assembly of foldon units via conditional sampling from a novel united-residue probabilistic model, which captures local conformational bias of backbone and side chain simultaneously in a united residue representation. The rationale for choosing united-residue representation is to integrate both backbone and side chain during structure modeling. It is found that (1) stepwise sampling produces lower energy conformations with higher accuracy than random sampling when everything else remains the same; (2) UniCon3D attains comparable performance with top five automated methods of CASP11 and CASP10 in a dataset of 30 and 15 difficult target domains, respectively; and (3) UniCon3D outperforms a baseline counterpart of UniCon3D that performs traditional random sampling as well as GDFuzz3D and FT-COMAR, two state-of-the-art approaches for de novo protein structure prediction aided by residue-residue contacts in a dataset containing 45 CASP10 targets.

The experimentally motivated stepwise, probabilistic sampling can lead to improvements during *de novo* conformational sampling of united-residue polypeptide chains by generating lower energy conformation with higher accuracy than traditional random sampling approaches.

[Source: Bioinformatics. 2016 Sep 15;32(18):2791-9. doi: 10.1093/bioinformatics/btw316]

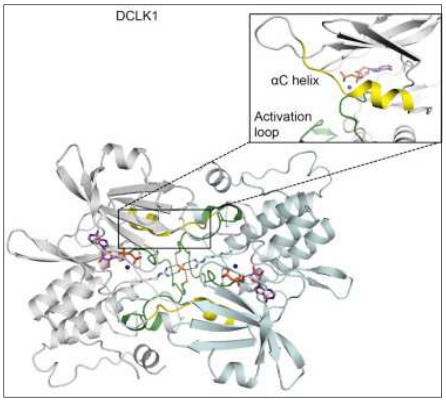
Bioinfo. Animation



Bioinformatics drug discovery

Doublecortin-like kinase 1

Doublecortin-like kinase 1 (DCLK1) is a serine/threonine kinase belongs to the family of microtubule-associated proteins. Originally identified for its role in neurogenesis, DCLK1 has recently been shown to regulate biological processes outside of the CNS. DCLK1 is among the 15 most common putative driver genes for gastric cancers and is highly mutated across various other human cancers. The protein encoded by this DCLK1 gene contains two N -terminal doublecortin domains, which bind microtubules and regulate microtubule polymerization, a C-terminal serine/threonine protein kinase domain, which shows substantial homology to Ca2+/calmodulin-dependent protein kinase, and a serine/proline-rich domain in between the doublecortin and the protein kinase domains, which mediates multiple protein-protein interactions.



DCAMKL1 protein is involved in several different cellular processes, including neuronal migration, retrograde transport, neuronal apoptosis and neurogenesis. This gene is up-regulated by brain-derived neurotrophic factor and associated with memory and general cognitive abilities. Multiple transcript variants generated by two alternative promoter usage and alternative splicing have been reported, but the full-length nature and biological validity of some variants have not been defined. These variants encode different isoforms, which are differentially expressed and have different kinase activities.

A study provide evidences that DCLK1 kinase activity negatively regulates microtubule polymerization. They present the crystal structure of the DCLK1 kinase domain at 1.7 Å resolution, providing detailed insight into the ATP-binding site that will serve as a framework for future drug design. This structure also allowed for the mapping of cancer-causing mutations within the kinase domain, suggesting that a loss of kinase function may contribute to tumorigenesis.

Upcoming events

Patents

Methods of identifying point mutations in a genome

US6994962B1

Inventor: William G. Thilly

Abstract

The invention relates to a method for identifying inherited point mutations in a targeted region of the genome in a large population of individuals and determining which inherited point mutations are deleterious, harmful or beneficial. Deleterious mutation are identified directly by a method of recognition using the set of point mutations observed in a large population of juveniles. Harmful mutations are identified by comparison of the set of point mutation observed in a large set of juveniles and a large set of aged individuals of the same population. Beneficial mutations are similarly identified.

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