Volume 10, Issue 2 February 2017

Bioinformatics up to Date

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

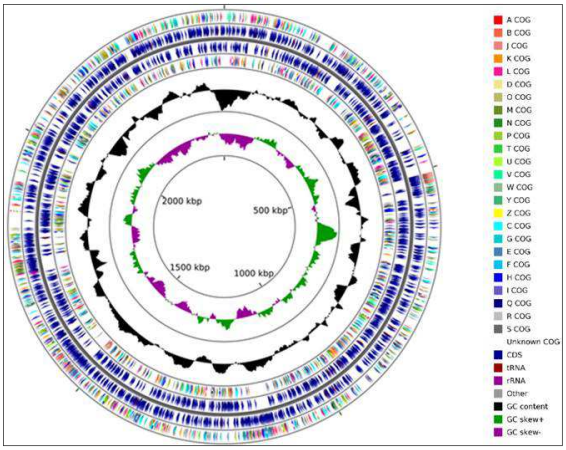
Inside…..

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.

Whole Genome Sequence Analysis of an Alachlor and Endosulfan Degrading *Micrococcus* sp. strain 2385

A research team from Florida A&M University recently isolated Micrococcus sp. strain 2385 from Ochlockonee River, Florida and demonstrated potent biodegradative activity against two commonly used pesticides- alachlor [(2-chloro-2`,6`-diethylphenyl-N (methoxymethyl)acetanilide)] and endosulfan



[(6,7,8,9,10,10-hexachloro-1,5,5a,6,9,9a-hexahydro-6,9methano-2,3,4-benzo(e)di-oxathiepin-3-oxide], respectively. To further identify the repertoire of metabolic functions possessed by strain 2385, a draft genome sequence was obtained, assembled, annotated and analyzed. The genome sequence of Micrococcus sp. strain 2385 consisted of 1,460,461,440 bases which assembled into 175 contigs with an N50 contig length of 50,109 bases and a coverage of 600x. The genome size of this strain was estimated at 2,431,226 base pairs with a G+C content of 72.8 and a total number of 2,268 putative genes. RAST annotated a total of 340 subsystems in the genome of strain 2385 along with the presence of 2,177 coding sequences. A genome wide survey indicated that that strain 2385 harbors a plethora of genes to degrade other pollutants including caprolactam, PAHs (such as naphthalene), styrene, toluene and several chloroaromatic compounds.

[source: *J Genomics* 2016; 4:42-47. doi:10.7150/jgen.16156]

LEAN method: A Network-based analysis of Omics Data:

Most computational approaches for the analysis of omics data in the context of interaction networks have very long running times, provide single or partial, often heuristic, solutions and/or contain user-tuneable parameters. A research group from University of Paris Diderot introduce local enrichment analysis (LEAN) for the identification of dysregulated subnetworks from genome-wide omics datasets. By substituting the common subnetwork model with a simpler *local* subnetwork model, LEAN allows exact, parameter-free, efficient and exhaustive identification of local subnetworks that are statistically dysregulated, and directly implicates single genes for follow-up experiments. The study recently published in journal *Bioinformatics (2017)*

Evaluation on simulated and biological data suggests that LEAN generally detects dysregulated subnetworks better, and reflects biological similarity between experiments more clearly than standard approaches. A strong signal for the local subnetwork around Von Willebrand Factor (VWF), a gene which showed no change on the mRNA level, was identified by LEAN in transcriptome data in the context of the genetic disease Cerebral Cavernous Malformations (CCM). This signal was experimentally found to correspond to an unexpected strong cellular effect on the VWF protein. LEAN can be used to pinpoint statistically significant local subnetworks in any genome-scale dataset. The R-package LEANR implementing LEAN is available on CRAN (https://cran.r-project.org).

[Source: Bioinformatics (2017) 33 (5): 701-709. DOI: https://doi.org/10.1093/bioinformatics/btw676]

FARME DB: A functional antibiotic resistance element database

Research team from University of Washington created an annotated database of DNA and protein sequences named "Functional Antibiotic Resistant Metagenomic Element (FARME)" derived exclusively from environmental metagenomics sequences showing Antibiotic resistance in laboratory experiments. The Antibiotic resistance (AR) is a major global public health threat but few resources exist that catalog AR genes outside of a clinical context. Current AR sequence databases are assembled almost exclusively from genomic sequences derived from clinical bacterial isolates and thus do not include many microbial sequences derived from environmental samples that confer resistance in functional metagenomic studies. These environmental metagenomic sequences often show little or no similarity to AR sequences from clinical isolates using standard classification criteria. In addition, existing AR databases provide no information about flanking sequences containing regulatory or mobile genetic elements. The developed database is a compilation of publically available DNA sequences and predicted protein sequences conferring AR as well as regulatory elements, mobile genetic elements and predicted proteins flanking antibiotic resistant genes. FARME is the first database to focus on functional metagenomic AR gene elements and provides a resource to better understand AR in the 99% of bacteria which cannot be cultured and the relationship between environmental AR sequences and antibiotic resistant genes derived from cultured isolates. The url of database is http://staff.washington.edu/jwallace/farme.

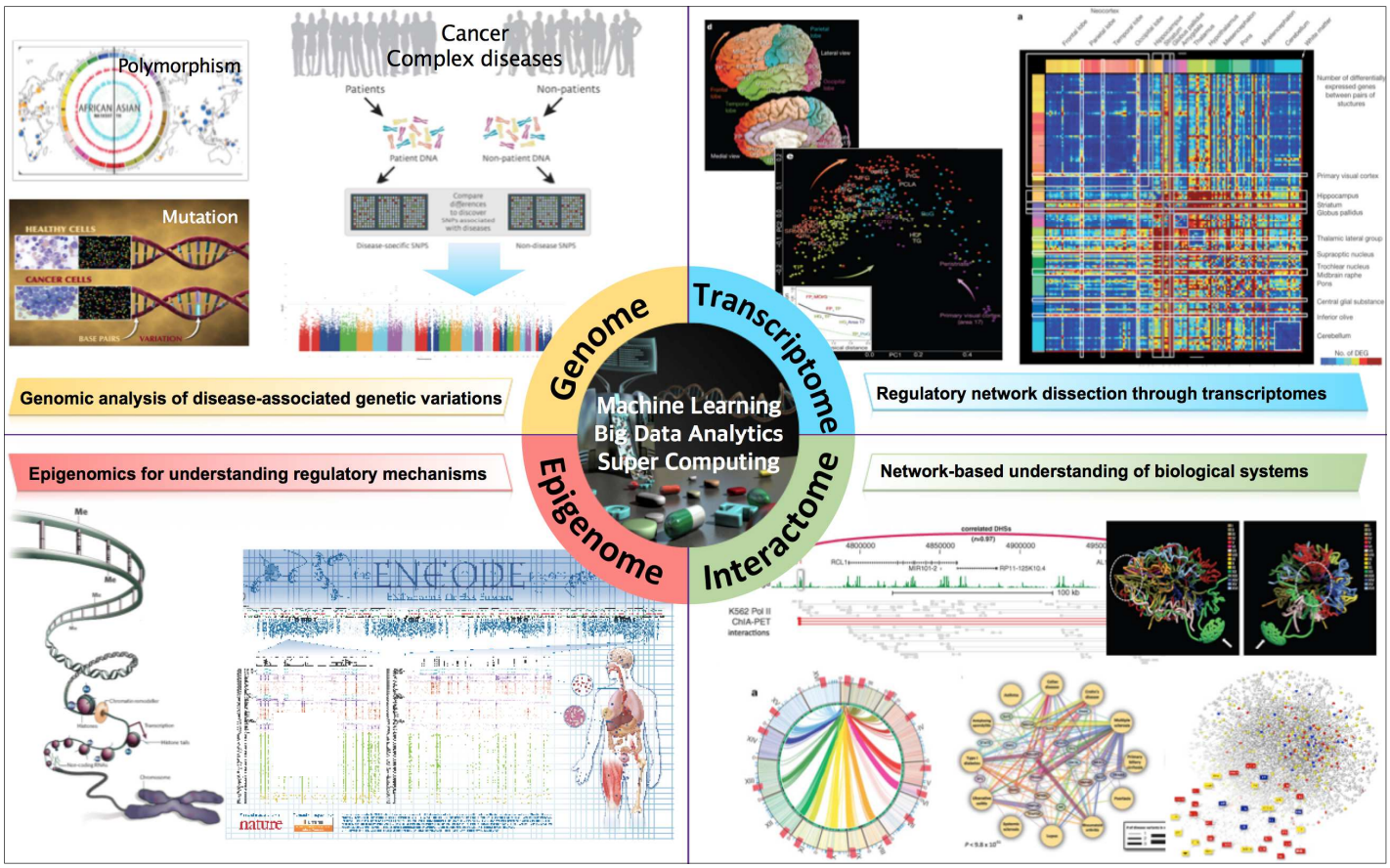
[Source: Database (Oxford) 2017; 2017 (1): baw165. doi: 10.1093/database/baw165]

SilkPathDB: a comprehensive resource for the study of silkworm pathogens

Silkworm pathogens have been heavily impeding the development of sericultural industry and play important roles in lepidopteran ecology, and some of which are used as biological insecticides. Rapid advances in studies on the omics of silkworm pathogens have produced a large amount of data, which need to be brought together centrally in a coherent and systematic manner. This will facilitate the reuse of these data for further analysis. A research group from Southwest University, China and USA collected genomic data for 86 silkworm pathogens from 4 taxa (fungi, microsporidia, bacteria and viruses) and from 4 lepidopteran hosts, and developed a open-access Silkworm Pathogen Database (SilkPathDB) to make this information readily available. The implementation of SilkPathDB involves integrating Drupal and GBrowse as a graphic interface for a Chado relational database which houses all of the datasets involved. The genomes have been assembled and annotated for comparative purposes and allow the search and analysis of homologous sequences, transposable elements, protein subcellular locations, including secreted proteins, and gene ontology. The SilkPathDB is built on a platform composed of a Linux Ubuntu Server 14.04, Apache 2.4, MySQL 5.5, PHP 5.5, PostgreSQL 9.3 and BioPerl 1.6. The implementation of SilkPathDB integrates and harnesses three complementary, mature and well supported technologies; Chado, GBrowse and Drupal (http://www.drupal.org). The database URL is http://silkpathdb.swu.edu.cn.

[Source: Database (Oxford) 2017; 2017 (1): bax001. doi: 10.1093/database/bax001]

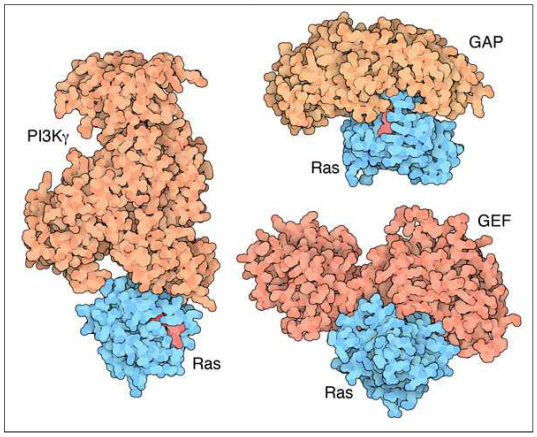
Bioinfo. Animation

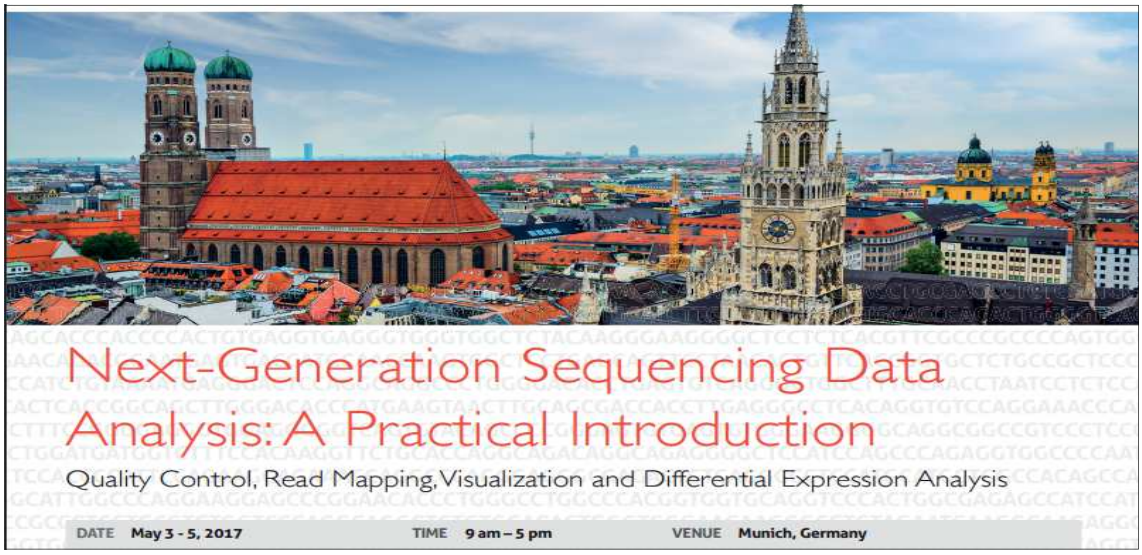


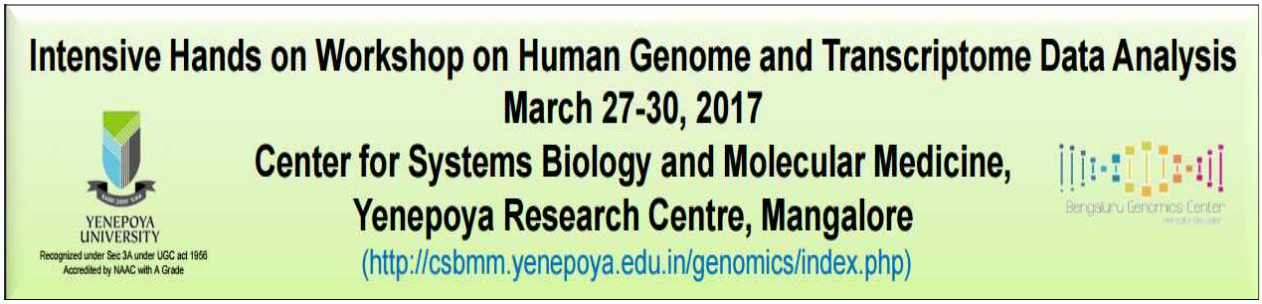
OMICS data analysis

Ras Protein

Oncogenes are genes that are closely linked to cancer, and the gene that encodes Ras was among the first to be discovered. Mutation of an oncogene changes the function of the encoded protein, creating the malignant properties that are needed for cancer to grow and spread. For instance, some oncogenes encode proteins involved in programmed cell death, such as p53 tumor suppressor, and the oncogenic mutation allows the cancer cell to evade our powerful protections against abnormal growth. Ras is involved in the signals passed between cells that control the amount of growth that is allowed at any time. Cancer-causing mutation of Ras creates a form of the protein that is always on. This is a disaster, because the mutated Ras continually tells the cancer cells that it is okay to multiply, without the normal limits that control cell growth. Ras is at the middle of a complex signaling network that delivers messages about growth, and is assisted by many different proteins. GEF proteins (guanine nucleotide exchange factors), such as Sos-1, turn the Ras switch on. Once GTP has bound in the empty site, the on state interacts with effectors like PI3Kγ a lipid kinase that phosphorylates lipids in the signaling network. Ras will then slowly break the GTP, but the process is accelerated by GAPs (GTPase-activating proteins), such as p120GAP.







Patents

Modular bioinformatics platform

US 20030177143 A1

Inventors: Steve Gardner

ABSTRACT

A bioinformatics system and method is provided for integrated processing of biological data. According to one embodiment, the invention provides an interlocking series of target identification, target validation, lead identification, and lead optimization modules in a discovery platform oriented around specific components of the drug discovery process. The discovery platform of the invention utilizes genomic, proteomic, and other biological data stored in structured as well as unstructured databases. According to another embodiment, the invention provides overall platform/architecture with integration approach for searching and processing the data stored in the structured as well as unstructured databases. According to another embodiment, the invention provides a user interface, affording users the ability to access and process tasks for the drug discovery process.

Kindly send us your feedback to

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