Volume 9, Issue 4 April 2016

Bioinformatics up to Date

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

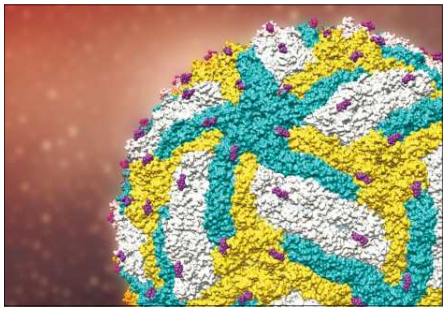
Inside…..

Advisor:
Dr D Ramaiah

Editors:
Mr Robin Das
Dr R Saikia
Dr H P Deka Baruah

Structure of the Zika virus

The Zika virus, a mosquito-borne virus, is mainly transmitted to people through the bite of infected Aedes mosquitoes. It causes Zika virus disease which presents symptoms such as mild fever, skin rashes, conjunctivitis, muscle and joint pain, malaise or headache in infected people. In February 2016, the Zika virus was declared by the World Health Organization (WHO) to be a public health emergency of international concern.



A team lead by Associate Professor Shee-Mei Lok and her team from Duke-NUS breakthrough in understanding the Zika virus structure and its behavior. The research published online on 19 April 2016 in the journal *Nature*. The findings reveal the Zika virus structure and identify potential sites on the virus to target with therapeutics. They have imaged the Zika virus under a cryo-electron microscope from a large number of purified viral particles. By using thousands of images, they reconstructed a high-resolution cryo-electron microscopy structure of the Zika virus.

The high-resolution structure of the Zika virus showed that the overall virus architecture is similar to other flaviviruses such as the West Nile and dengue viruses. The structure also revealed that the Zika virus surface proteins have tighter interactions compared to the dengue virus, therefore making it more stable than the dengue virus. This may explain its ability to survive in harsh conditions such as semen, saliva and urine.

The team targeting next step is to understand the effect of potent antibodies on the Zika virus. By examining the structure presented in this study, they work to determine how the antibodies could be used to kill the virus.

[source: ***Structure of the thermally stable Zika virus.*** *Victor A. Kostyuchenko, Elisa X. Y. Lim et al. Nature (2016)*]

Bioinformatics carrier:

-PhD in computational evolutionary biology; A three-year Ph.D. studentship in evolutionary biology in the laboratory of Andreas Wagner at the University of Zurich. [http://www.naveenbioinformatics.co.in/2016/04/phd-in-computational-evolutionary.html]

-UDSC Genomics Scientist/Technical Officer/SRF Vacancies ; [http://www.du.ac.in/du/index.php?mact=News,cntnt01,detail,0&cntnt01articleid=10524&cntnt01returnid=83]

-WII Biodiversity Project Scientist Walk IN 2016 April [http://www.wii.gov.in/images//images/documents/project_biologist_mehao_wls.pdf]

Draft Genome of Rhodococcus rhodochrous TRN7

Rhodococcus rhodochrousis a metabolically diverse species found in different environmental niches, from soil to waste treatment plants Members of the genus Rhodococcus have been shown to synthesize and accumulate triacyl-glycerol (TAGs) (4), which can be an alternative in the production of biofuels.

Genome sequencing for *R. rhodochrous* TRN7 was performed using the Ion Torrent PGM platform (Life Technologies). Briefly, the genomic DNA was fragmented by using the Bioruptor UCD- 200. The template library was prepared with the Ion Plus fragment library kit and clonally amplified in the One Touch System with the Ion PGM template OT2 400 kit. The amplified library was sequenced using the Ion PGM sequencing 400 kit within the 318 Chip version 2. A total of 5,048,800 reads, ranging from 25 to 484 bp in length, were sequenced. The resulting reads were assembled using the MyPro pipeline software (6). Briefly, sequences were trimmed and filtered, followed by genome assembly. MyPro groups five different genome assemblers: VelvetOptimiser, Edena, Abyss, SOAPdenovo, SPAdes, and SOAP2. Integration of the resulting contigs was performed using CISA and SOAP2. Integration resulted in 173 contigs, totaling 4,871,006 bp, with an average size (N50) of 70,171 bp, longest contig size of 278,361 bp, and GC content of 70.2%. The whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession numbers FBUK01000001 to FBUK01000173.

[Draft genome of Rhodococcus rhodochrous TRN7, isolated from the coast of Trindade Island, Brazil. 2016Genome Announc 4(2):e01707 -15. doi:10.1128/genomeA.01707-15.]

Computational identification of piRNA targets

PIWI-interacting RNAs (piRNAs) are a class of small non-coding RNAs that are highly abundant in the germline. One important role of piRNAs is to defend genome integrity by guiding PIWI proteins to silence transposable elements (TEs), which have a high potential to cause deleterious effects on their host. The mechanism of piRNA-mediated post-transcriptional silencing was also observed to affect mRNAs, suggesting that piRNAs might play a broad role in gene expression regulation. However, there has been no systematic report with regard to how many protein-coding genes might be targeted and regulated by piRNAs.

A team from Chinese Academy of Sciences, Beijing, China trained a support vector machine classifier based on a combination of Miwi CLIP-Seq-derived features and position-derived features to predict the potential targets of piRNAs on mRNAs in the mouse. Reanalysis of a published microarray dataset suggested that the expression level of the 2587 protein-coding genes predicted as piRNA targets showed significant upregulation as a whole after abolishing the slicer activity of Miwi, supporting the conclusion that they are subject to piRNA-mediated regulation. The web version of the method called pirnaPre as well as our results for browse is available at http://www.regulatoryrna.org/software/piRNA/piRNA_target_mRNA/index.ph

[source: bioinformatics.oxfordjournals.org/content/32/8/1170.abstract]

Bloom Filter Trie: A data structure for pan-genome storage

High throughput sequencing technologies have become fast and cheap technique that resulting, large-scale projects started to sequence tens to several thousands of genomes per species, producing a high number of sequences sampled from each genome. Such a highly redundant collection of very similar sequences is called a pan-genome. It can be transformed into a set of sequences "colored" by the genomes to which they belong. A colored de Bruijn graph (C-DBG) extracts from the sequences all colored *k*-mers, strings of length *k*, and stores them in vertices.

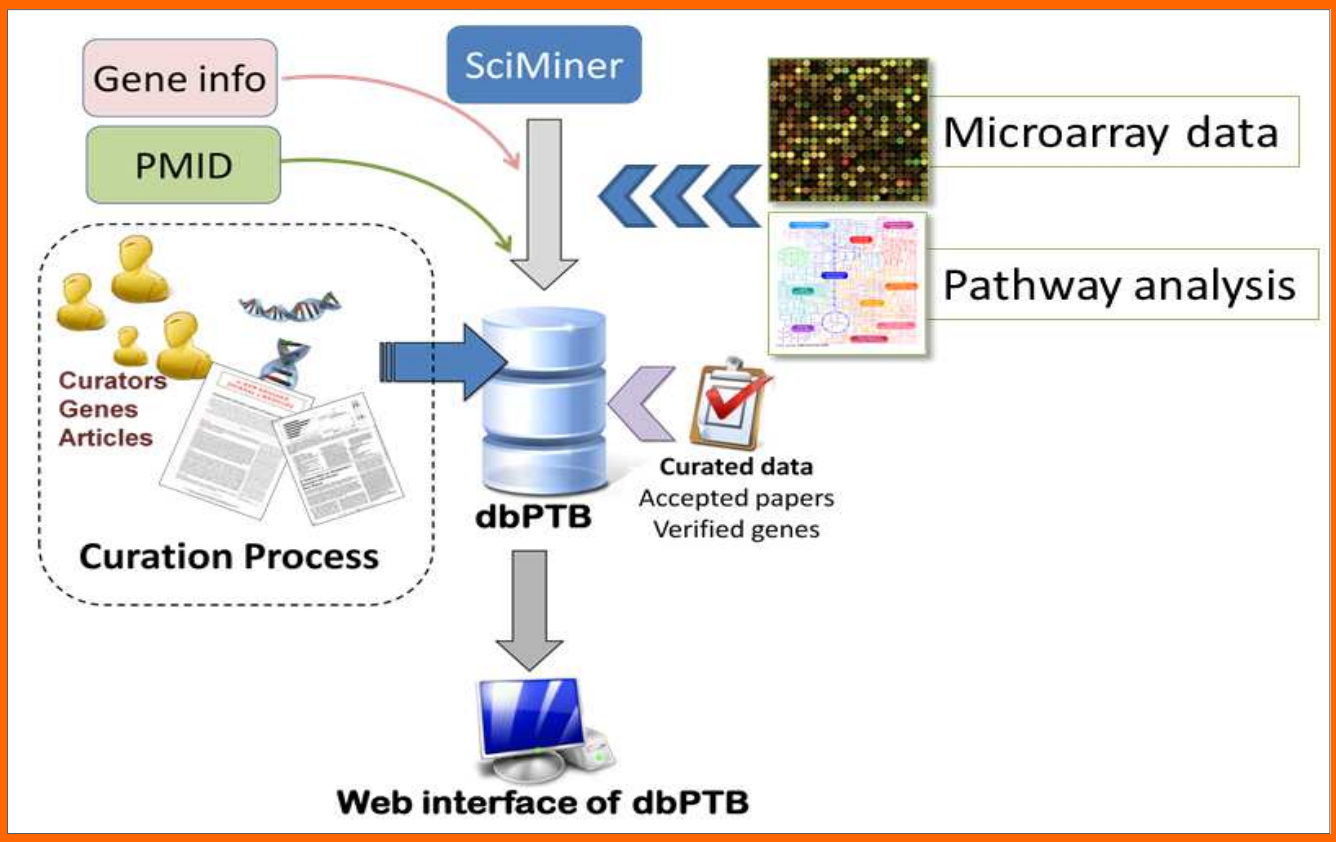
The bloom filter trie (BFT), an alignment-free, reference-free and incremental data structure for storing a pan-genome as a C-DBG. The data structure allows to store and compress a set of colored *k*-mers, and also to efficiently traverse the graph. Bloom filter trie was used to index and query different pan genome datasets. Compared to another state-of-the-art data structure, BFT was up to two times faster to build while using about the same amount of main memory. For querying *k*-mers, BFT was about 52–66 times faster while using about 5.5–14.3 times less memory. The trie stores *k*-mers and their colors based on a new representation of vertices that compress and index shared substrings. Vertices use basic data structures for lightweight substrings storage as well as Bloom filters for efficient trie and graph traversals. Experimental results prove better performance compared to another state-of-the-art data structure. The tool available at https://www.github.com/GuillaumeHolley/BloomFilterTrie.

[source: Algorithms for Molecular Biology201611:3, DOI: 10.1186/s13015-016-0066-8]

2P2Idb v2: a structural database of protein–protein interactions

2P2Idb is a hand-curated structural database dedicated to protein–protein interactions with known small molecule orthosteric modulators. It compiles the structural information related to orthosteric inhibitors and their target and provides links to other useful databases. 2P2Idb includes all interactions for which both the protein–protein and protein–inhibitor complexes have been structurally characterized. Since its first release in 2010, the database has grown constantly and the current version contains 27 protein–protein complexes and 274 protein–inhibitor complexes corresponding to 242 unique small molecule inhibitors which represent almost a 5-fold increase compared to the previous version. A number of new data have been added, including new protein–protein complexes, binding affinities, molecular descriptors, precalculated interface parameters and links to other webservers. A new query tool has been implemented to search for inhibitors within the database using standard molecular descriptors. A novel version of the 2P2I-inspector tool has been implemented to calculate a series of physical and chemical parameters of the protein interfaces. Several geometrical parameters including planarity, eccentricity and circularity have been added as well as customizable distance cutoffs. This tool has also been extended to protein–ligand interfaces. The 2P2I database thus represents a wealth of structural source of information for scientists interested in the properties of protein–protein interactions and the design of protein–protein interaction modulators.

[Source: Database, 012016, 1–6 doi: 10.1093/database/baw007]



patent

Modular bioinformatics platform

US 20030177143 A1

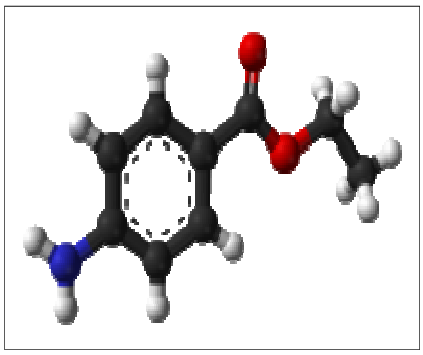
Inventors : Steve Gardner

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.

Benzocaine

Benzocaine is a local anesthetic commonly used as a topical pain reliever or in cough drops. It is the active ingredient in many over-the-counter anesthetic ointments such as products for oral ulcers. It is also combined with antipyrine to form A/B otic drops to relieve ear pain and remove earwax. Benzocaine is the ethyl ester of *p*-aminobenzoic acid (PABA). It can be prepared from PABA and ethanol by Fischer esterification or via the reduction of ethyl *p*-nitrobenzoate. Benzocaine is sparingly soluble in water; it is more soluble in dilute acids and very soluble in ethanol, chloroform and ethyl ether. The melting point of benzocaine is 88 – 90 °C, and the boiling point is about 310 °C. The density of benzocaine is 1.17 g/cm^3.



Benzocaine is generally well-tolerated and non-toxic when applied topically as recommended. Benzocaine topical is used to reduce pain or discomfort caused by minor skin irritations, sore throat, sunburn, teething pain, vaginal or rectal irritation, ingrown toenails, hemorrhoids, and many other sources of minor pain on a surface of the body. Benzocaine is also used to numb the skin or surfaces inside the mouth, nose, throat, vagina, or rectum to lessen the pain of inserting a medical instrument such as a tube or speculum.

Upcoming Events

Department of Biotechnology, Govt. of India Sponsored

Training on

Advanced methods for molecular typing of microbes

Date - 9th to 13th May, 2016

State Biotech Hub Training

SBTHub (Assam), College of Veterinary Science

AAU, Khanapara, Guwahati - 781022

Symposium on

Genomics in clinical practice: Future of precision medicine

June 1-2, 2016, Yenepoya University, Mangalore, India

YENEPOYA UNIVERSITY

YU-IOB Center for Systems Biology and Molecular Medicine

Institute of Bioinformatics

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