Volume 7, Issue 10 October 2014

Bioinformatics up to Date

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

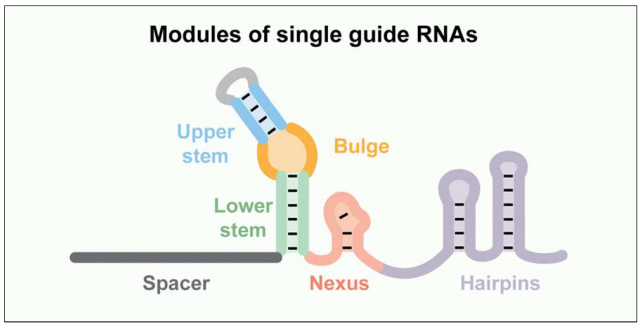
Inside…..

5 days Hands-on Workshop on Molecular Biotechnology and Bioinformatics

Venue: ICSCCB, R.H. 2, Ujwal Regalia, Near Prabhavee Tech Park, Baner Road, Pune – 411045, India

Researchers Advance Genome Editing Technique

Customized genome editing – the ability to edit desired DNA sequences to add, delete, activate or suppress specific genes – has major potential for application in medicine, biotechnology, food and agriculture.



Now, in a paper published in Molecular Cell, North Carolina State University researchers and colleagues examine six key molecular elements that help drive this genome editing system, which is known as CRISPR-Cas.

NC State's Dr. Rodolphe Barrangou, an associate professor of food, bioprocessing and nutrition sciences, and Dr. Chase Beisel, an assistant professor of chemical and biomolecular engineering, use CRISPR-Cas to take aim at certain DNA sequences in bacteria and in human cells. CRISPR stands for "clustered regularly interspaced short palindromic repeats," and Cas is a family of genes and corresponding proteins associated with the CRISPR system that specifically target and cut DNA in a sequence-dependent manner.

Essentially, the authors say, bacteria use the system as a defense mechanism and immune system against unwanted invaders such as viruses. Now that same system is being harnessed by researchers to quickly and more precisely target certain genes for editing.

"This paper sheds light on how CRISPR-Cas works," Barrangou said. "If we liken this system to a puzzle, this paper shows what some of the system's pieces are and how they interlock with one another. More importantly, we find which pieces are important structurally or functionally – and which ones are not."

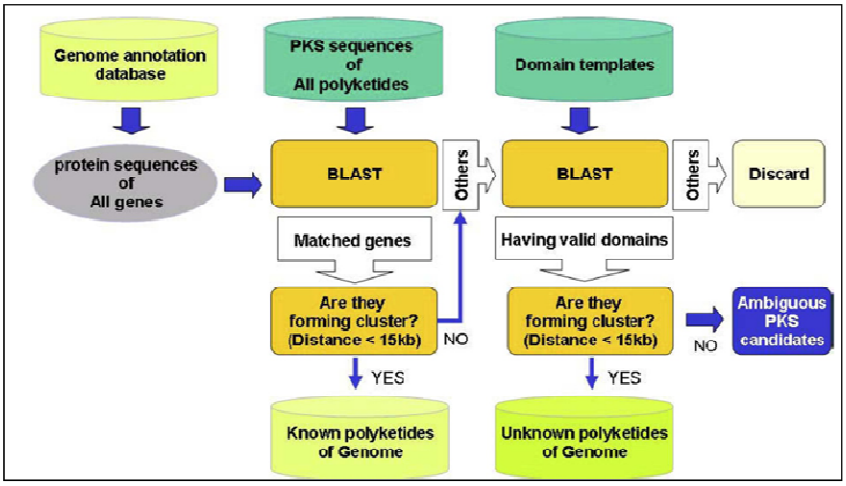
The CRISPR-Cas system is spreading like wildfire among researchers across the globe who are searching for new ways to manipulate genes. Barrangou says that the paper's findings will allow researchers to increase the specificity and efficiency in targeting DNA, setting the stage for more precise genetic modifications.

The work by Barrangou and Beisel holds promise in manipulating relevant bacteria for use in food – think of safer and more effective probiotics for your yogurt, for example – and in model organisms used in agriculture, including gene editing in crops to make them less susceptible to disease. The collaborative effort with Caribou Biosciences, a start-up biotechnology company in California, illustrates the focus of these two NC State laboratories on bridging the gap between industry and academia, and the commercial potential of CRISPR technologies, the researchers say.

[http://scicasts.com/genomics/8551-researchers-advance-genome-editing-technique/]

Computational analysis of Polyketide Synthases through ASMPKS

Polyketides are secondary metabolites of microorganisms having diverse biological activities, including pharmacological functions such as antibiotic, antitumor and agrochemical properties. Polyketides are synthesized by a set of enzymatic reactions and create multi domain mega syntheses called as Polyketide synthase (PKS)s, which coordinate the elongation of carbon skeletons by the stepwise condensation of short carbon precursors. Due to their importance as drugs, the volume of data on polyketidesis rapidly increasing and creating a need for computational analysis methods for efficient Polyketide research.



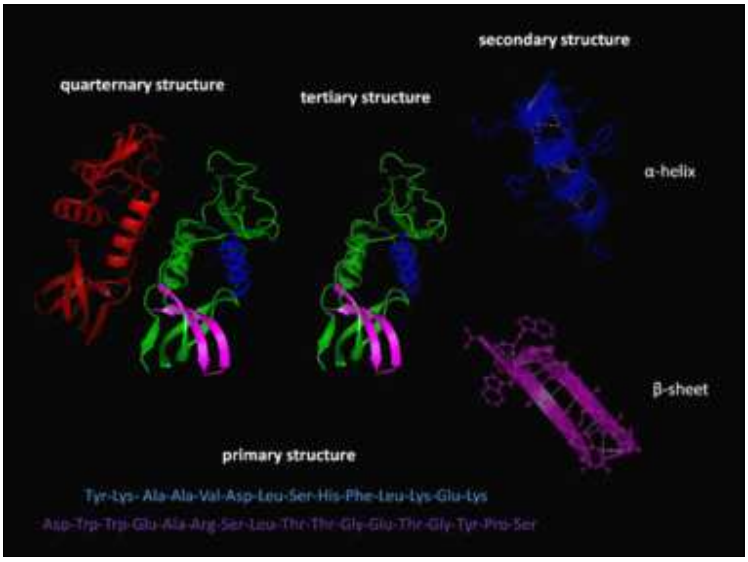
Protocol of automated polyketide annotation

ASMPKS (Analysis System for Modular Polyketide Synthesis) a system described by Hongseok Tae and his colleges for computational analysis of PKSs. ASMPKS predicts domain information from protein sequences based on the homology search method with template sequences of domains. ASMPKS is user-friendly and operates on a web interface, useful system to analyze known polyketides and to predict new polyketides.

(Anwesha Gohain; BIOTECHNOLOGY, CSIR-NEIST)

Protein folding

Protein folding a process by which a protein structure assumes its functional shape or conformation. It is the physical process by which a polypeptide folds into its characteristic and functional three-dimensional structure from random coil. Each protein exists as an unfolded polypeptide or random coil when translated from a sequence of mRNA to a linear chain of amino acids. This polypeptide lacks any stable (long-lasting) three-dimensional structure. Amino acids interact with each other to produce a well-defined three-dimensional structure, the folded protein known as the native state. The resulting three-dimensional structure is determined by the amino acid sequence (Anfinsen's dogma).



The correct three-dimensional structure is essential to function, although some parts of functional proteins may remain unfolded. Failure to fold into native structure generally produces inactive proteins, but in some instances misfolded proteins have modified or toxic functionality. Several neurodegenerative and other diseases are believed to result from the accumulation of amyloid fibrils formed by misfolded proteins. Many allergies are caused by incorrect folding of some proteins, for the immune system does not produce antibodies for certain protein structures.

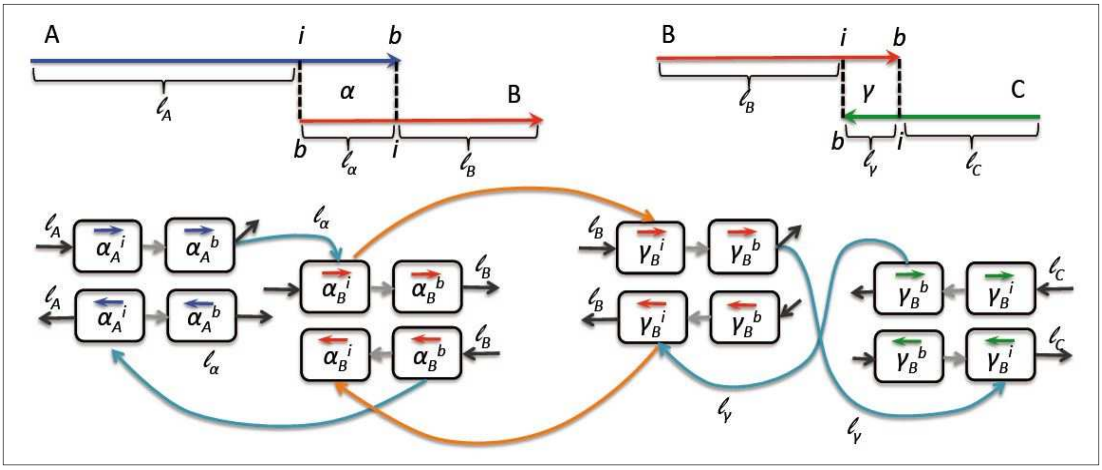
3DNALandscapes: A database for exploring conformational features of DNA

3DNALandscapes, located at: http://3DNAscapes.rutgers.edu, is a new database for exploring the conformational features of DNA. In contrast to most structural databases, which archive the Cartesian coordinates and/or derived parameters and images for individual structures, 3DNALandscapes enables searches of conformational information across multiple structures. The database contains a wide variety of structural parameters and molecular images, computed with the 3DNA software package and known to be useful for characterizing and understanding the sequence-dependent spatial arrangements of the DNA sugar-phosphate backbone, sugar-base side groups, base pairs, base-pair steps, groove structure, etc. The data comprise all DNA-containing structures--both free and bound to proteins, drugs and other ligands--currently available in the Protein Data Bank. The web interface allows the user to link, report, plot and analyze this information from numerous perspectives and thereby gain insight into DNA conformation, deformability and interactions in different sequence and structural contexts. The data accumulated from known, well-resolved DNA structures can serve as useful benchmarks for the analysis and simulation of new structures. The collective data can also help to understand how DNA deforms in response to proteins and other molecules and undergoes conformational rearrangements.

[Nucleic Acids Res. 2010 Jan;38(Database issue):D267-74. doi: 10.1093/nar/gkp959. Epub 2009 Nov 11.]

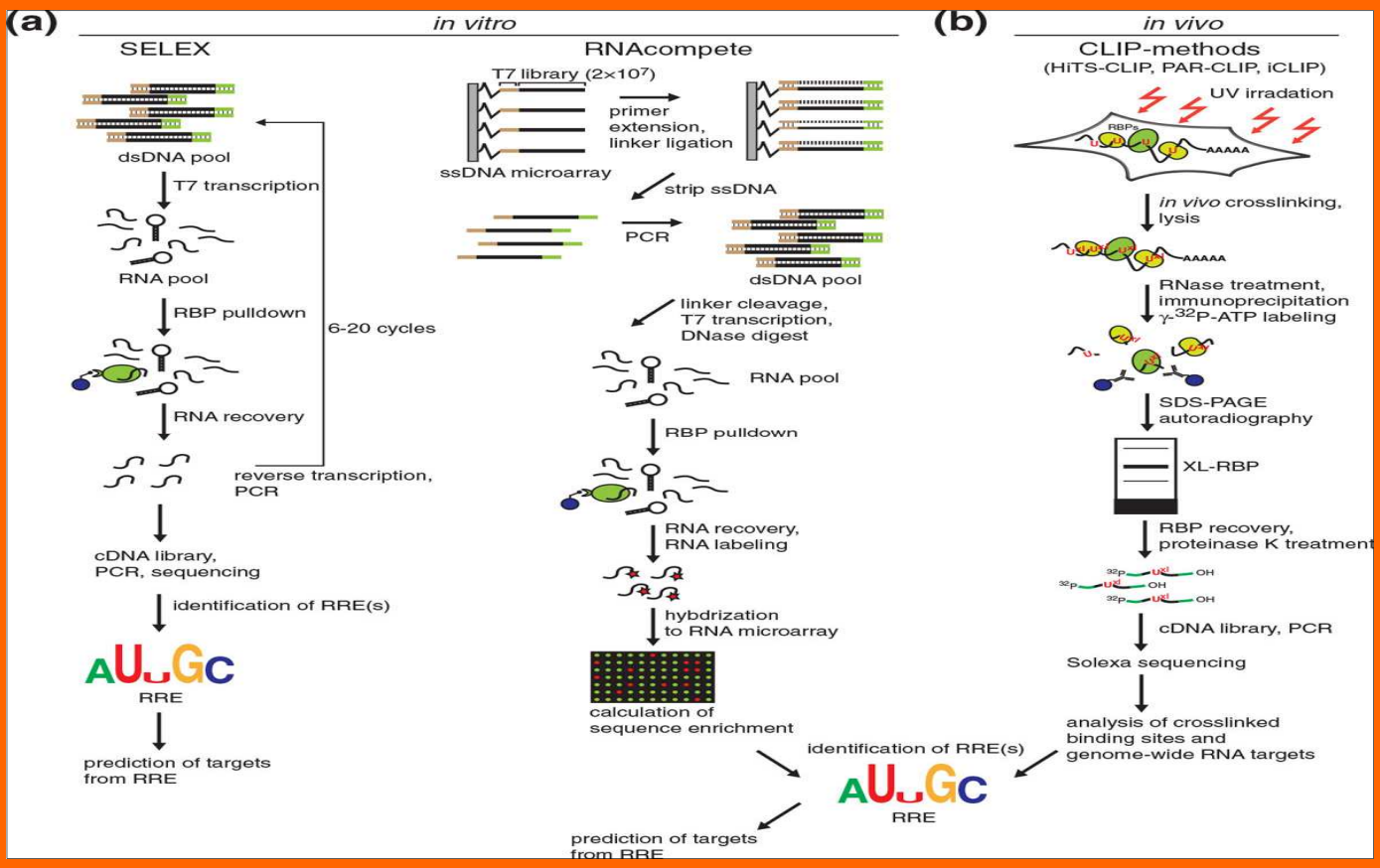
MIX: A Genomic tool

Mix is a tool to for genome finishing that combines multiple assemblies obtained from NGS reads. Its algorithm takes two assemblies and generates another one that mixes them in order to extend the length of resulting contigs. It builds an assembly graph in which all of the contigs are vertices and edges represent the best possible alignments between two contigs that have the potential of being used as basis for contig extension. The resulting output assembly corresponds to a certain path in this assembly graph that maximizes the cumulative contig length.



Mix runs in a command line environment. The package contains one main script called Mix.py that coordinates the execution of the whole process. The tool developed by the **Bordeaux Bioinformatics Center**(*CBIB).*

Bioinfo. Animation



Overview of *in vitro* and *in vivo* methods for RRE determination

Patent News

Modular bioinformatics platform

US 20030177143 A1

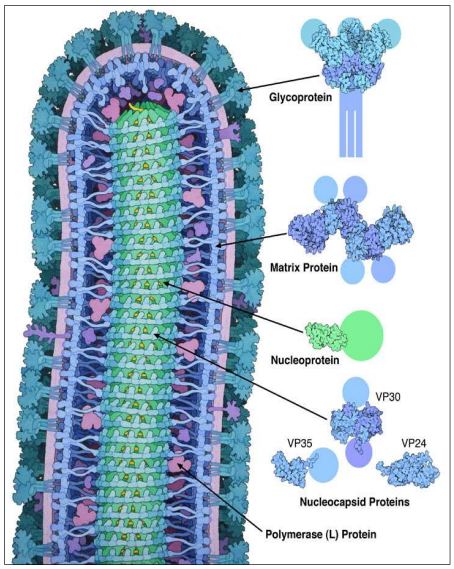
Inventors : Steve Gardner

Publication date: Sep 18, 2003

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.

Ebola Virus Proteins



The genome of ebola virus contains instructions for building seven proteins, which assemble with the genomic RNA to form one of the deadliest viruses. Ebola virus is surrounded by a membrane stolen from an infected cell, and studded with ebola glycoproteins. A layer of matrix proteins support the membrane on the inside, and hold a cylindrical nucleocapsid at the center, which stores and delivers the RNA genome.

The ebola glycoprotein has the task binding to receptors on a cell surface and getting the ebola genome inside. It shares many of the features of other viral fusion proteins, such as influenza hemagglutinin and HIV envelope glycoprotein. It extends from the surface of the virus, and is covered with carbohydrate chains that hide it from the immune system. It is also a highly dynamic protein that snaps into a different shape when it binds to a cell surface, dragging the virus and cell close enough to each other that the membranes fuse. The structure shown here (PDB entry 3csy) includes the receptor-binding portion of the glycoprotein and the machinery for membrane fusion. To allow crystallization, a small domain has been removed that includes much of the carbohydrate that normally covers the glycoprotein.

Upcoming Events

ICDBT 2014, Institute of Advanced Study in Science & Technology (IASST)

Paschim Boragaon, Garchuk, GUWAHATI -781035, ASSAM, INDIA

Web: http://www.iasst.gov.in/ICDBT/abouticdbt.html

2nd International Conference on Biotechnology and Bio-informatics (ICBB-2015)

Biotech & Bioinfo Conference, Pune, India, 6-8 February 2015

Web: http://icbb.in/

Kindly send us your feedback to

Robin Das
Project Fellow; BIF, Biotech Division.
CSIR-North East Institute of Science and Technology, Jorhat, Assam
E-mail: robindas460@gmail.com
Ph No-07399923578