Volume 7, Issue 1 January, 2014

Bioinformatics up to Date

(Bioinformatics Infrastructure Facility, Biotechnology Division)
North-East Institute of Science & Technology
Jorhat - 785 006, Assam
(http://www.rrljorhat.res.in/biotechnology.html)

Inside…..

Protein Data Bank

As of Tuesday January 28, 2014 at 4 PM PST there are 97362 molecule Structures

CLC-Bio and Korilog Release KLAST Software Plug-in for Fast Accurate Sequence Similarity Search in Next-Generation Sequencing

CLC bio, a QIAGEN company (NASDAQ:QGEN)(FWB:QIA) recently announced in collaboration with Korilog the release of the KLAST plug-in for CLC Genomics Workbench and CLC Main Workbench to accelerate the analysis of data from next-generation sequencing (NGS). KLAST is a new sequence similarity search tool which builds on a highly optimized implementation of the PLAST algorithm published in BMC Bioinformatics. KLAST is comparable to the current standard BLAST providing a similar accuracy, but can perform up to 25 times faster (with comparable search parameters and multicore threaded configuration). Another important KLAST benefit is that its algorithm is designed to compare two large sets of sequence databanks in a single run.

"The integration of the KLAST plug-in into CLC Genomics Workbench makes it possible to perform sequence similarity queries on very large data sizes, which adds significant utility to the very critical challenge in analyzing next-generation sequencing data," said Mikael Flensborg, Director of Global Partner Relations for CLC bio. "This solution will allow users to significantly accelerate their research by providing them with this fast, accurate, and NGS-scalable sequence similarity search software that integrates with our CLC Genomics Workbench, and offers a very user-friendly graphical interface."

"We are very enthusiastic to join forces with CLC bio's market-leading bioinformatics platform," said Patrick Durand, Founder and CEO of Korilog. " The logic of KLAST is different from BLAST, and can provide significant accelerations of seed-based heuristic comparison methods using multi-threading and SSE instructions that make the tool high-performing and more scalable than BLAST on multiple cores.

[http://online.wsj.com/article/PR-CO-20140110-904083.html]

Bioinfo. Carrier

1. **PhD student position - Bioinformatics/Plant Genomics : Department of Biology, The University of South Dakota, Vermillion, SD, United States http://www.usd.edu/graduate-school/international-admissions.cfm**
2. **Postdoctoral Fellow in Bioinformatics : Nashville, U.S; Department of Biomedical Informatics, Vanderbilt University Medical Center. http://bioinfo.mc.vanderbilt.edu/.**

Bio-It

Researchers Reveal Crowdsourced RNA Designs Outperform Computer Algorithms

Scientists at the University of Massachusetts Medical School (UMMS) have developed a new method for piecing together the short DNA reads produced by next-generation sequencing technologies that are the basis for building complete genome sequences. Dr. Job Dekker and colleagues have shown that entire genomes can be assembled faster and more accurately by measuring the frequency of interactions between DNA segments and by using their three-dimensional shape as a guide. Employing this technique, they have been able to place 65 previously unaccounted for DNA fragments in incomplete regions of the human genome. Details of the study appear online in Nature Biotechnology.

Next-generation sequencing techniques can easily read hundreds of millions DNA sequences at a time. However, these sequences are randomly broken into extremely short pieces and need to be assembled into larger pieces using computer algorithms that can match up overlapping pieces. The end result of this initial assembly is typically a set of as many as 100,000 DNA fragments which then need to be organized with respect to one another in the correct order to create a complete genome.

Dekker added, "This new approach to genome assembly can help produce higher-quality genome sequences faster and easier than current methods. It will be especially interesting to apply this method to identify chromosomal aberrations, which are a hallmark of cancer."

[http://scicasts.com/bio-it/7247-researchers-reveal-crowdsourced-rna-designs-outperform-computer-algorithms/]

Proteomics

Researchers Discover Potential Drug Targets for Early Onset Glaucoma

Using a novel high-throughput screening process, scientists have identified molecules with the potential to block the accumulation of a toxic eye protein that can lead to early onset of glaucoma. Researchers have implicated a mutant form of a protein called myocilin as a possible root cause of this. Mutant myocilin is toxic to the cells in the part of the eye that regulates eye pressure. These genetically inherited mutants of myocilin clump together in the front of the eye, preventing fluid flow out of the eye, which then raises eye pressure. This cascade of events can lead to early onset-glaucoma, which affects several million people from childhood to age 35.

To find molecules that bind to mutant myocilin and block its aggregation, researchers designed a simple, high-throughput assay and then screened a library of compounds and identified two molecules with potential. "These are really the first potential drug targets for glaucoma," said Raquel Lieberman, an associate professor in the School of Chemistry and Biochemistry at the Georgia Institute of Technology in Atlanta, whose lab led the research. Lieberman presented her findings on January 20 at the Society for Laboratory Automation and Screening conference in San Diego, Calif. The study was published on Nov. 26, 2013, in the journal ACS Chemical Biology. The National Institutes of Health and the Pew Scholar in Biomedical Sciences program provided support for the research. The work was a collaboration involving Georgia Tech, Emory University and the University of South Florida.

In a separate study, Lieberman's lab characterized the toxic myocilin aggregates. The researchers are now focusing on mapping the structure of myocilin to learn more about what myocilin does and why it is in the eye in the first place.

[http://scicasts.com/proteomics/2043-protein-functions/7229-researchers-discover-potential-drug-targets-for-early-onset-glaucoma/]

TimeLogic

TimeLogic, a division of Active Motif, is a leader in hardware accelerated bioinformatics search tools, which accelerate genome annotation by combining optimized bioinformatics applications with powerful FPGA-based PCIe accelerator cards. This blend of specialized hardware and optimized software provides a perfect combination of performance, accuracy and value. Furthermore, these systems are simple to maintain and scale easily. Utilizing a TimeLogic system reduces pressure on over-used CPU-clusters by off-loading BLAST, Smith-Waterman (SW) and Hidden Markov Model (HMM) tasks to a highly time- and energy-efficient solution. From the earliest genome sequencing projects in the 1990s to the largest metagenomics projects undertaken to date, TimeLogic has provided the enabling technology to make this research possible. Active Motif/TimeLogic operates globally through its corporate headquarters in Carlsbad, CA, European headquarters in Rixensart, Belgium and Japanese headquarters in Tokyo, Japan. Active Motif/TimeLogic applies a multi-disciplinary approach to create new and modify existing technologies to meet the current and future needs of life science researchers.

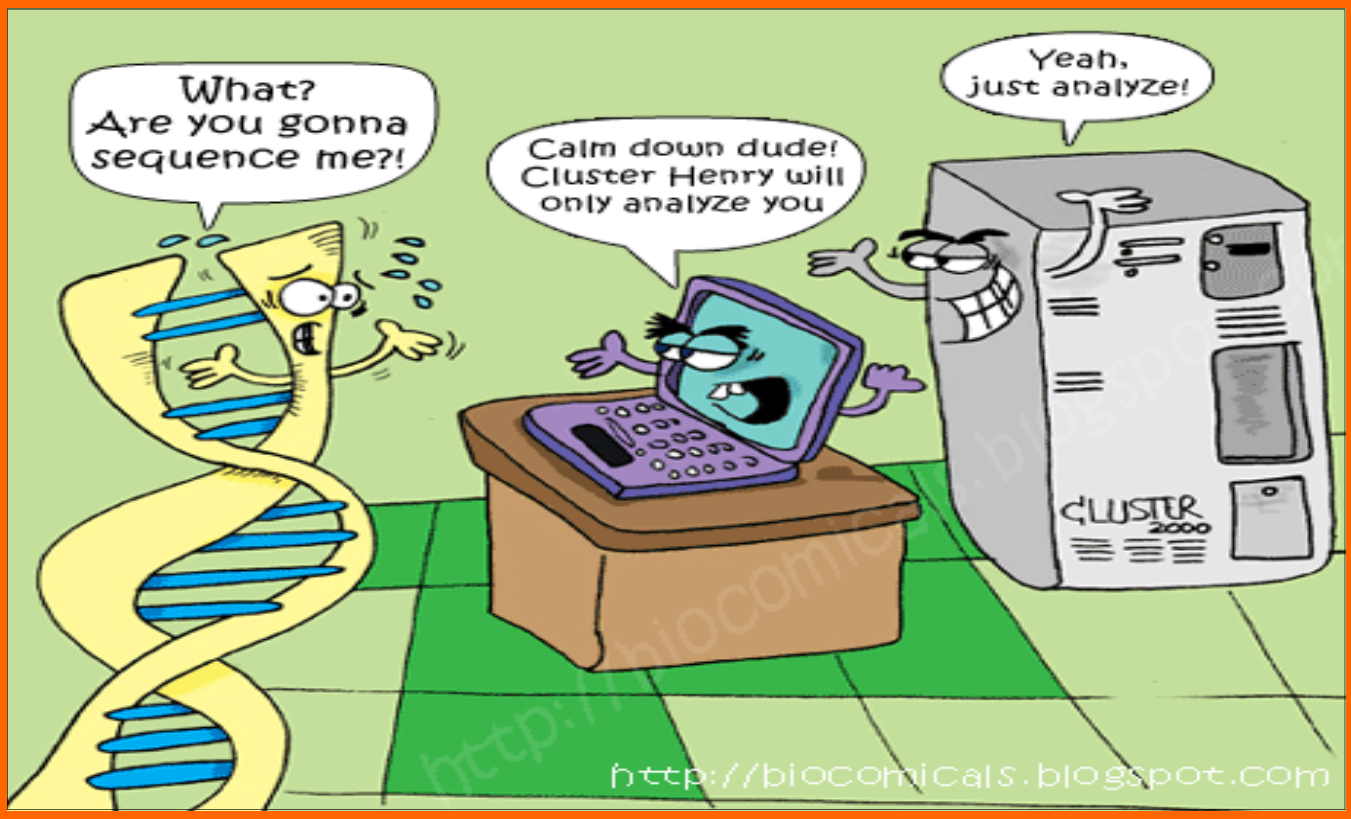
NEW BIOINFORMATICS TOOL FOR COMPLEX GENOME ANALYSIS

Scientists from The Genome Analysis Centre (TGAC) in Norwich UK have developed a new bioinformatics tool to boost complex genome analysis. The software, NextClip, generates a comprehensive quality report and extracts high class trimmed and de-duplicated data. The tool supports Illumina's recently released Nextera LMP kit, which enables the production of jumping libraries of up to 12kb. These Long Mate Pair libraries are an invaluable resource for analysing large areas of the genome, carrying out complex assemblies and other downstream bioinformatics analytics.

Richard Leggett, at TGAC, said: "Regulating laboratory protocols and selection of sequenced data for downstream analysis are vital in making effective use of mate pair libraries. However, quality control of the libraries can require significant bioinformatics analysis. Further processing is also required to extract true mate pair reads, remove fragment junction adaptors and clip reads". For this reason we developed NextClip, a tool for comprehensive quality analysis of Nextera LMP libraries and preparation of reads for scaffolding. Sequence reads from Long Mate Pair libraries are an important tool in the construction of complex genome assemblies because they connect large repeat regions. Grouping the data generated from mate pair library sequencing with shorter insert paired-end reads provide a powerful combination, allowing the joining together of longer DNA sequences, with higher certainty.

TGAC offers state of the art DNA sequencing facility, unique by its operation of multiple complementary technologies for data generation. The Institute is a UK hub for innovative Bioinformatics through research, analysis and interpretation of multiple, complex data sets. It hosts one of the largest computing hardware facilities dedicated to life science research in Europe.

Bioinfo. Animation



Patent News

Computer-implemented cellular modeling having parallel pathways

(U.S. Patent 8,521,438)

Inventors: Edson; Patrick (West Newton, MA), Paxson; Ricardo (Boston, MA)

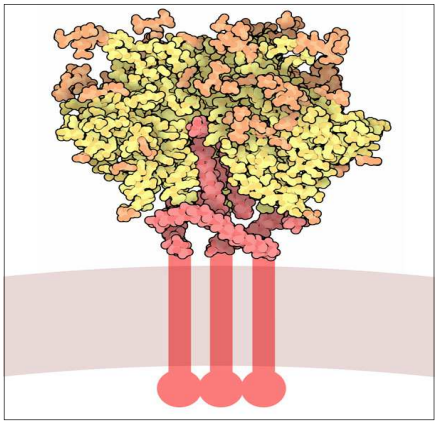
August 27, 2013

Abstract

A computer-implemented method for cellular modeling is provided. The computer-implemented method for cellular modeling may include generating a cellular model, the model comprising a plurality of compartments, each compartment comprising at least one reaction, at least one species, or a combination of at least one reaction and at least one species; receiving a selection of at least one compartment for simulation to obtain at least one selected compartment; simulating the at least one selected compartment to obtain a result; and providing an output of the result from simulating the compartment; wherein at least two of the compartments comprise parallel pathways through the cellular model.

Molecule of month

HIV Envelope Glycoprotein



The envelope glycoprotein (Env) of HIV performs the many complex steps needed for membrane fusion. First, it attaches itself to proteins on the surface of the cell. Then, it acts like a spring-loaded mousetrap and snaps into a new conformation that drags the virus and cell close enough that the membranes fuse. Finally, the HIV genome is released into the cell, where it quickly gets to work building new viruses.

The Env glycoprotein is found on the surface of HIV, forming little knobs composed of three identical subunits. It is covered with carbohydrate, with about 81 glycosylated sites on each trimer. It is encoded in the HIV genome as one long protein, which is called gp160 (gp is short for "glycoprotein"). Then, as it is transported to the infected cell's surface, the many carbohydrates are added and it is cleaved into two pieces, called gp120 and gp41. Then, it is incorporated into the virus as it buds from the cell surface.

The structure shown here, from PDB entry 4nco, is one of the most complete structures solved so far. It includes gp120 (yellow) with lots of carbohydrate (orange), and the outer portion of gp41 (red). The same molecule has also been studied by electron microscopy in PDB entries3j5m and 4cc8).

Upcoming Events

National Conference

http://jjtu.ac.in

Deadline : 2014-02-14

Organized By: Shri JJT University

Venue: Jhunjhunu , Rajasthan , India

5-Days Hands-on Workshop on Molecular Biotechnology and Bioinformatics Workshop

24th to 28th February 2014

Pune, Maharashtra, India

Website: http://www.icsccb.org/workshops/

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