



Bioinformatics up to Date

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

CONTENTS

Cover Story	1
Bioinfy Quiz	1
Computational Chemistry	2
Proteomics	2
Bioserver	3
Computers for Biologists	3
Genomics	4
Upcoming Events	4
Bioinfy Animator	5
Molecule of the month	5
Contact Us	5
Answers to quiz	5

Adviser:
 Dr.P.G.Rao

Editors:
 Dr. R.L.Bezbaruah
 Pompei Sharma
 Dhrubjyoti Gogoi

COVER STORY

Computer identifies viruses with close genetic ties to ‘swine’ flu

Scientists have used new mathematical and computational techniques to identify six influenza A viruses that have particularly close genetic relationships to the H1N1 ‘swine’ flu virus. The researchers found these strains by using powerful computers to analyse the relationships between the genomes of more than 5,000 strains of influenza A that have been isolated over several decades and recently sequenced. Janies jointly with Ohio State co-authors Shahid Bokhari, a research professor of biomedical informatics, and Laura Pomeroy, a postdoctoral researcher in veterinary preventive medicine, obtained data on all fully sequenced influenza A viruses available in a National Institutes of Health database.

They then used supercomputers to efficiently track how all of these influenza A viruses are related to each other and which paths through the network led to pandemic H1N1 of 2009.

The research was published online in the journal IEEE Transactions on Computational Biology and Bioinformatics.

“It’s not unlike a social network, except that it’s tracking an exchange of genetic material rather than gossip,” Daniel Janies, associate professor of biomedical informatics at Ohio State University and a co-author of the study, said.

“This network gives us an explicit historical and molecular map of how influenza A viruses evolved from several ancestors to modern-day viruses,” he said.

BIOINFY QUIZ

- Who developed UPGMA method and in which year
 A) Michener, 1957
 B) Sokal, 1957
 C) Michener and Sokal, 1957
- Who developed Neighbor-Joining method and in which year?
 A) Saitou and Nei, 1987
 B) Saitou and Nei, 1988
 C) Saitou, 1988
- Who developed Maximum Parsimony method and in which year?
 A) Michener, 1957
 B) Sokal, 1957
 C) Fitch, 1977
- Who gave the main principle of Maximum Parsimony method?
 A) Fitch
 B) Sokal
 C) Occam’s Razor
- Which of the following is the distance based method?
 A) UPGMA method
 B) Maximum Parsimony method
 C) Maximum Likelihood method

Answers on page 5

COMPUTATIONAL CHEMISTRY

CCCBDB

The NIST Computational Chemistry Comparison and Benchmark Database is a collection of experimental and ab initio thermochemical properties for a selected set of molecules.

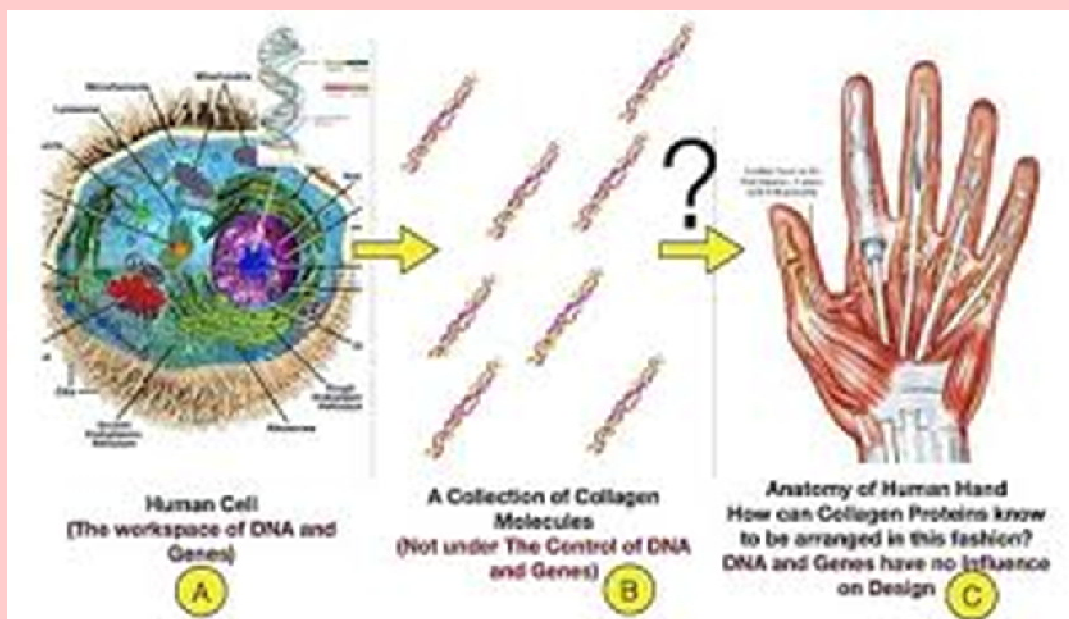
The goals are:

- 1) Provide a benchmark set of molecules for the evaluation of ab initio computational methods.
- 2) Allow the comparison between different ab initio computational methods for the prediction of thermochemical properties.

The thermochemical values with error bars included in the CCCBDB are:

1. Enthalpies of formation.
2. Entropies, heat corrections (integrated heat capacity).
3. Data needed to compute thermochemical properties, such as geometries, rotational constants, vibrational frequencies, barriers to internal rotation, and electronic energy levels.
4. Additional computed properties such as atomic charges, electric dipole moments, quadrupole moments, polarizabilities, and HOMO-LUMO gaps.

PROTEOMICS



BIOSERVER

GS²

GS² provides the fastest set similarity measure for GO annotations. It provides an online tool that calculates the functional similarity of large (1000+) amounts of genes in under a second. The measure quantifies the similarity of a gene set with GO annotations by averaging the contribution of each gene's GO terms and their ancestor terms with respect to the GO vocabulary graph.

They retrieve the GO annotations directly from the [Ensembl biomart](#) so that the annotations are up to date. For large gene sets (1000+) it may take around a minute to retrieve the annotation data, but the similarity calculation will take only a second!

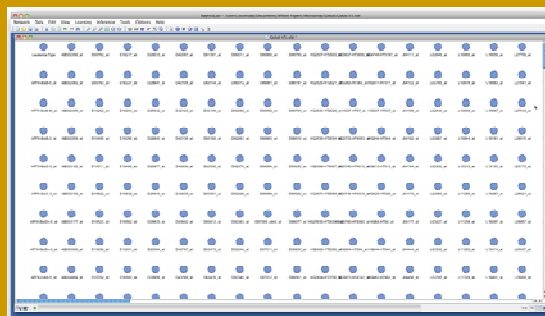
They also provide a Python implementation of our method. This library only requires [Python](#) to be installed on the machine, making this code platform independent. pyGS2 is distributed under the [LPGL](#).

COMPUTERS FOR BIOLOGISTS

BayesiaLab 5.0: analytics, Data Mining, MoDeLing & siMuLation

BayesiaLab is a powerful desktop application (Windows/Mac/Unix) for knowledge discovery, data mining, analytics, predictive modeling and simulation - all based on the paradigm of Bayesian networks. Bayesian networks have become a very powerful tool for deep understanding of very complex, high-dimensional problem domains, ranging from bioinformatics to marketing science.

BayesiaLab is the world's only comprehensive software package for learning, editing and analyzing Bayesian networks. It provides perhaps the easiest way to practically apply artificial intelligence tools, thus transforming and, more importantly, massively accelerating research workflows. Microarray Analysis with Bayesian Networks is one of the major applications of BayesiaLab.



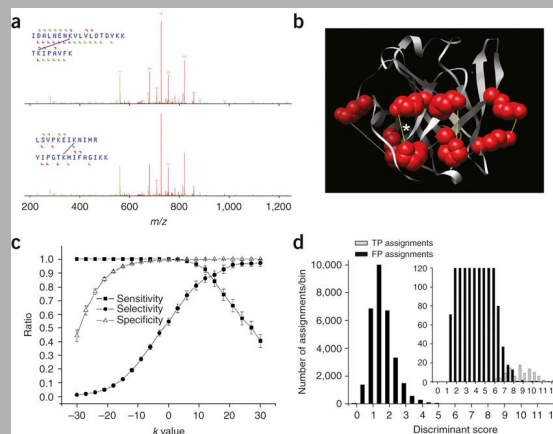
Microarray Analysis with Bayesian Networks and BayesiaLab

GENOMICS

xQuest

xQuest is a novel search engine for cross-linked peptides from complex samples. It was presented at the ASMS 2007, and was published in Nature Methods. xQuest works with small and large protein databases and features flexible fragment ion assignment, advanced scoring, and interactive evaluation tools.

xQuest searches MS² spectra of isotopically labeled cross-links. Samples containing cross-linked peptides are separated on a reverse phase column coupled to a mass spectrometer (liquid chromatography–mass spectrometry; LC-MS). The peptide masses are screened for isotopic pairs based on the presence of a characteristic isotopic shift. MS² spectra from these pairs are analyzed according to the absence or presence of an isotopic shift between peaks in the fragment ion spectrum. Accordingly, peaks are separated into pre-cross-link (common) and post-cross-link (cross-link) peaks. This peak sorting improves specificity by matching only a subset of all peaks against a fraction of all theoretical fragment ions .



Linear discriminant analysis separates true positive MS² assignments to standard protein cross-links from false positive hits in a large peptide-interaction

UPCOMING EVENTS

Pondicherry University

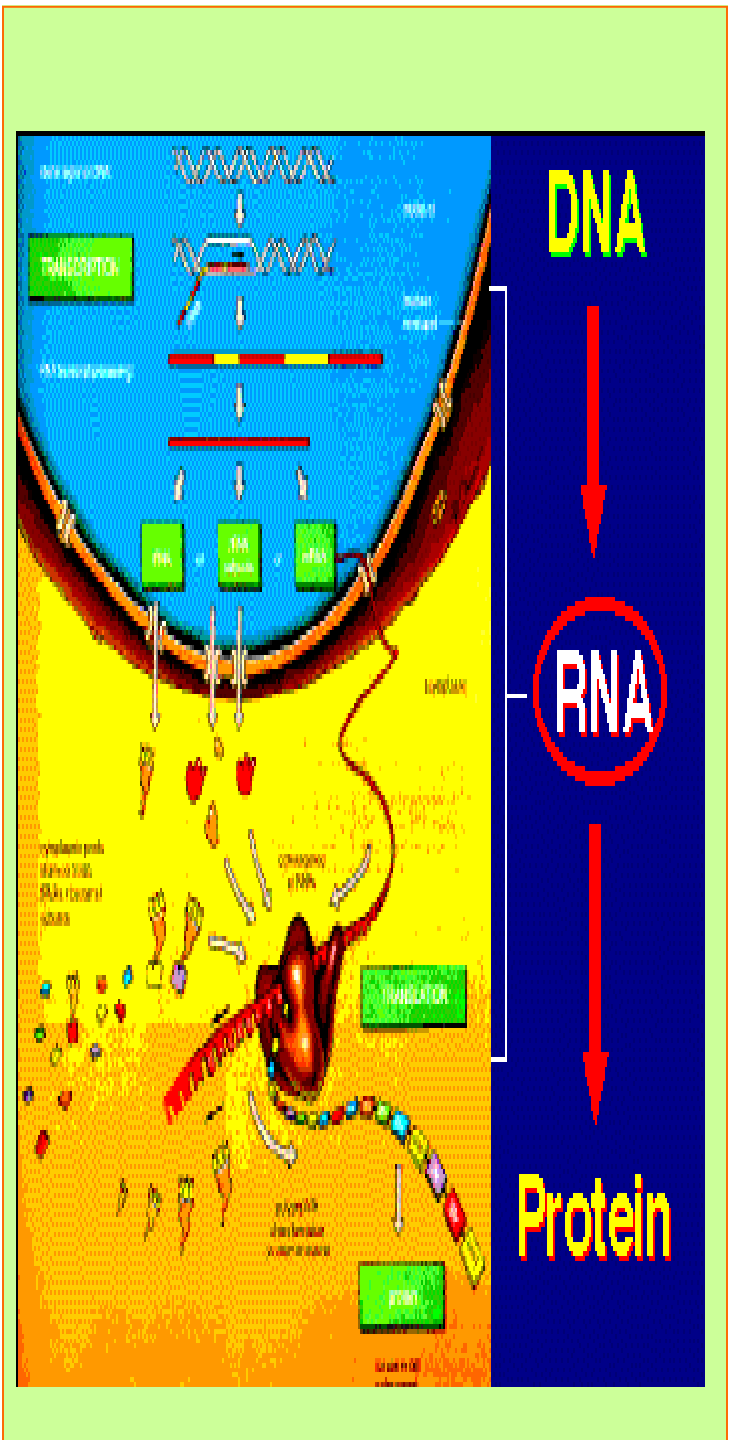
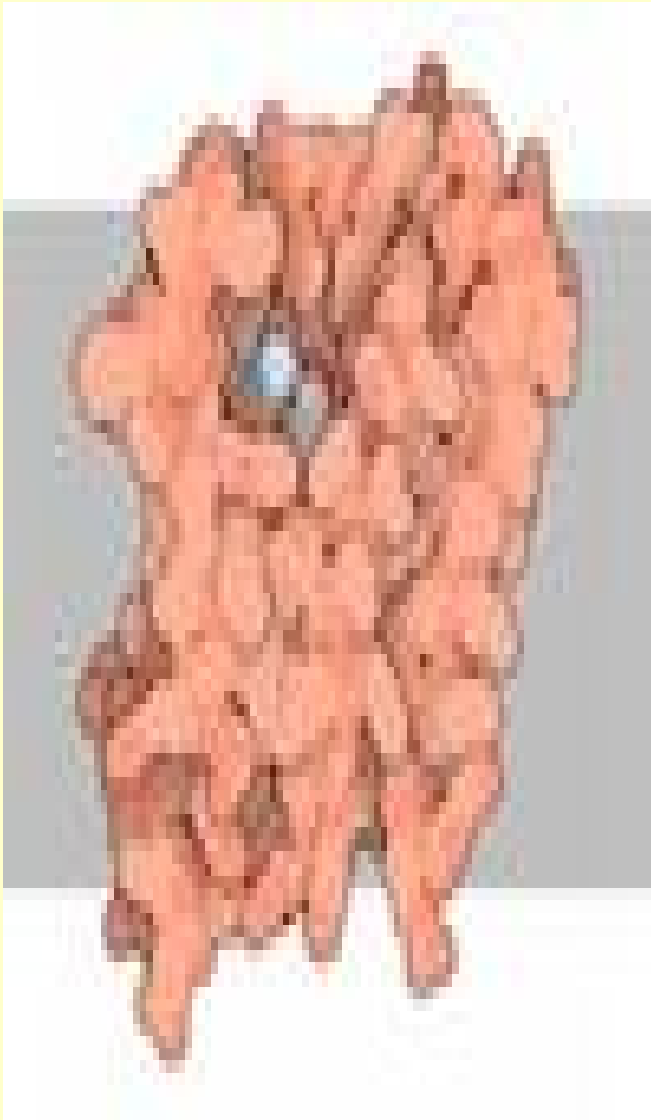
Centre of Bioinformatics

National Seminar on

"Current Trends in Genomics and Proteomics"

September 21-23, 2011

MOLECULE OF THE MONTH
Rhomboid PRotease GI PG



CONTACT US:

Pompi Sharma, Project Assistant (Level – II)

E-mail: pompi.sharma86@gmail.com

Dhrubajyoti Gogoi, Project Assistant (Level – II)

E-mail: dhruba.bio.du@gmail.com

Answers of Bioinfy Quiz

1) C 2) A 3) C 4) C 5) A