

# Bioinformation up to Date

(Bioinformatics Infrastructure Facility, Biotechnology Division)  
**North-East Institute of Science & Technology**  
 Jorhat - 785006, Assam

Contents	
Cover Story	1
Genomics	2
Software Mania	3
Bio Server	3
Bioinfy Quiz	3
Computational Chemistry	4
Molecule of the Month	4
Bioinfy Animator	4
Answer to Quiz	4
Contact Us	4

### Adviser:

Dr. P.G. Rao

### Editors:

Salam Pradeep Singh

Dr. R.L. Bezbaruah

### Upcoming Events

1. “International Conference Frontiers in Biological Sciences (InCoFIBS-2020)” @ Department of Life Science, National Institute of Technology, Rourkela from 1<sup>st</sup> - 3<sup>rd</sup> October 2010.
2. “National Symposium on Assessment and Conservation of Plant Genetic Resources under Changing Environment” @ Center for Advanced Studies in Botany, NEHU, Shillong from 8<sup>th</sup> – 10<sup>th</sup> Sept, 2010.

### Cover Story

#### European ribosomal RNA database

The European ribosomal RNA database compiles all complete or nearly complete SSU (small subunit) and LSU (large subunit) ribosomal RNA sequences. Sequences are provided in aligned format. The alignment takes into account the secondary structure information derived by comparative sequence analysis of thousands of sequences. Additional information such as literature references, taxonomy, secondary structure models and nucleotide variability maps, is also available.

The database was founded in the research group of Professor De Wachter at the university of Antwerp, Belgium in 1984. Since 2002, the database is maintained at the university of Gent, Belgium in the research group of Professor Van de Peer. The database is maintained by Jan Wuyts and Yves Van de Peer.

#### Important features:

##### 1. ssu rRNA:

A database on the structure of ssu ribosomal subunit RNA. This database contains sequences stored in a special distribution format. The database includes indications of secondary structure elements for each sequence, based on the adopted rRNA secondary structure, which in turn is corroborated by the observation of compensating substitutions in the alignment.

##### 2. lsu rRNA database:

A database on the structure of lsu ribosomal subunit RNA. This database contains sequences stored in a special distribution format. The database includes indications of secondary structure elements for each sequence, based on the adopted rRNA secondary structure, which in turn is corroborated by the observation of compensating substitutions in the alignment.

##### 3. rRNA BLAST:

A BLAST search on ssu rRNA and lsu rRNA database.

##### 4. Quick phylogeny search:

This service is an implementation of the bibi program by Guy Perrière and Gregory Devulder. It uses the programmes BLAST and ClustalW to make a quick approximate identification of a sequence. In a first step, the query sequence is searched against a ribosomal RNA database using BLAST. From the BLAST results the best matching sequences are extracted and they are subsequently aligned using ClustalW which also produces a phylogenetic tree. The final results are displayed on a webpage. The java applets ATV and Jalview (Jalview - a java multiple alignment editor) are included to easily inspect the phylogenetic tree and the alignment respectively.

##### 5. rRNA secondary structure models:

The database also includes secondary structure models SSU (small subunit) ribosomal RNA molecules. The sequences are ordered according to the first 3 levels of the NCBI taxonomy. The first level corresponds to the super kingdom but subsequent levels are not necessarily kingdom and phylum.

##### 6. rRNA variability maps:

The database also includes variability maps for the SSU and LSU rRNA molecules from prokaryotes and eukaryotes. For the moment, 6 2D variability maps are available as well as 3 3D variability maps.

## Genomics

### Recent advances in asthma genetics

It has been found that over 100 genes have been reported to be associated with asthma or related phenotypes. In a previous review by Hofjan et al. published in 2003, functional information was reported for 40 polymorphisms and listed another 22 genes which have such data. Some important genes such as filaggrin, interleukin-13, interleukin-17 and the cysteinyl leukotriene receptor-1 which not only were replicated by independent association but also have functional data.

Filaggrin (FLG) is a key protein of the epidermis and is therefore important in the formation of the protective skin barrier. FLG was initially described as a candidate gene for atopic dermatitis.

FLG was initially identified as an ichthyosis vulgaris candidate gene in a study of 15 families segregating this single gene disorder. There are three exons in FLG and two mutations (R501X and 2282del4) were found in the ichthyosis vulgaris patients. Both mutations are in exon 3 and stop protein translation within the first FLG repeat and result in complete loss of FLG peptide production. Therefore, these two functional mutations were considered as the likely causal mutations for ichthyosis vulgaris in these families.

Three studies have shown association of extrinsic AD(atopic dermatitis), asthma and rhinitis, asthma severity with the FLG variants although in some studies association was only seen with concomitant AD. All these data suggest that the two FLG mutant alleles are important risk factors for AD but only for asthma when it is found in the context of AD. An additional seven nonsense or frame shift FLG mutations were identified in the European population and two in the Asian population. There were three reasonably prevalent mutations (in addition to R501X and 2282del4) that showed association with childhood eczema

Interleukin-13 (IL13) is an asthma susceptibility gene because it is a cytokine produced by Th2 cells and because its genetic location on chromosome 5q31 has been linked to asthma and related phenotypes in multiple linkage studies. This cytokine is capable of promoting allergen-induced bronchial hyperresponsiveness, epithelial cell damage, goblet cell hyperplasia with mucus hyperproduction, and eosinophilia.

The IL13, IL17F was discovered recently. It is one member of the IL17 gene family and the coding sequence contains 7742 bp, including three exons. IL17F was investigated as an asthma candidate gene because of its function i.e. IL17F is one of the cytokines produced by activated mast cells, CD4+ T cells, and basophils and can upregulate IL6 and IL8 transcripts and protein expression in primary bronchial epithelial cells. IL17F is expressed in human liver, lung, and fetal liver tissue and increased expression was observed in the airways of allergic asthma patients. In addition, IL17F is located on chromosome 6p, which has been linked to asthma and asthma-related phenotypes in multiple genome scans.

Cysteinyl leukotrienes are bronchoconstrictors and proinflammatory mediators of the asthma response that act through two G protein-coupled receptors: cysteinyl leukotriene receptor-1 (CYSLTR1) and CYSLTR2. Both receptors are the targets of anti-asthmatic drugs. CYSLTR2 maps to chromosome 13q14, approximately 300 kb from D13S153, which was reported linked to asthma in two studies. The association of CYSLTR2 and asthma was replicated in two subsequent studies and two functional studies showed potency of leukotriene D4 on the M202V variant was lower compared with the wild-type receptor. CYSLTR1 is located on chromosome Xq21.

ORMDL3 is a poorly characterized gene and the underlying mechanism for the association with asthma is unclear. It is a member of a family of endoplasmic reticulum membrane proteins and has a ubiquitous pattern of expression in humans as well as in *Drosophila*. It encodes transmembrane proteins anchored in the endoplasmic reticulum.

The pathogenesis of asthma, a complex disease, involves gene-gene interactions as well as gene-environment interactions. Multiple modest risk factors work synergistically to influence asthma disease susceptibility. The number of asthma-susceptibility genes identified by genetic studies is still increasing. However, most studies lack information on the mechanism by which the SNPs lead to asthma. The functional consequences of SNPs is essential to confirm association results and to understand how the SNPs combine to influence susceptibility. Eventually, results from this important field are expected to improve preventive strategies and to aid in the development of diagnostic tools and therapies.

.....*Asha Deori, Dibrugarh University*

## Software Mania

### YASARA View

YASARA, Yet Another Scientific Artificial Reality Application, is a molecular visualisation, modelling, and dynamics program that can be used for a series of scientific applications as is expressed by the large number of scientific articles mentioning this software. The free version of YASARA (YASARA-View) is well suited for bioinformatics education.

YASARA View is available for free and contains all the functions you need to explore a macromolecular structure interactively. As a bonus, you get YASARA's innovative 3D engine, which is up to 35 times faster than what you usually know from OpenGL (see benchmarks), you can load multiple structures at the same time, create publication-quality ray-traced images including labels, and program your own macros and Python plugins. Included is YASARA Movie, a player for multimedia presentations and tutorials created with any of the higher stages, as well as the YASARA Python module to simply 'import yasara' in your Python scripts.

#### Key Features:

- Photo-realistic molecular graphics.
- Parallel (orthographic) and perspective projection.
- Support for over 70 molecular fileformats (based on OpenBabel).
- Optimal depth perception thanks to real-time shadows and ambient lighting, even on older hardware.
- Common graphics styles: CPK, balls & sticks, sticks, traces, tubes, ribbons, cartoons, hydrogen bonds, labels, arrows.
- Superpose structures and structure ensembles, calculate proper RMSDs.
- Align multiple proteins based on their structure or sequence.
- Measure distances, angles, dihedrals.
- Analyze and change atomic B-factors and occupancies, rename and renumber atoms and residues.

## Bio Servers

### GO- based similarity of Gene Sets (GS2)

GO- based similarity of Gene Sets (GS2) is a novel measure to provide an efficient similarity measure for GO annotations. It provides the fastest set similarity measure for GO (Gene Ontology) annotations where thousands of genes can be handled in a fraction of 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. It manages to preserve semantic relationships without the use of weighting factors.

The growing availability of genome-scale data sets has attracted increasing attention to the development of computational methods for automated inference of functional similarities among genes and their products. One class of such methods measures the functional similarity of genes based on their distance in the Gene Ontology (GO). To measure the functional relatedness of a gene set, these measures are applied to every pair of genes in the set, and the average of all pairwise distances is calculated. However, as more data becomes available and gene sets become larger, such pair-based calculation becomes prohibitive.

This online tool calculates the functional similarity of a set of genes using their GO annotations. The GO annotations are directly retrieved from the Ensembl biomart for up to date annotations. It takes an input Ensembl gene identifiers. To input a set of genes– the appropriate organism dataset is selected first. And then, gene identifiers are entered using one line per gene as shown in the above picture.

.....Sanchaita, Dibrugarh University

### Bioinfy Quiz - 023

1. From which organism is S1 nuclease isolated?
  - A) Aspergillus oryzae
  - B) Citobacter intermedus
  - C) Xenopus laevis
2. Which was the first protein to be completely sequenced?
  - A) Cytochrome C
  - B) Insulin
  - C) Lysozyme
3. Which of these proteins is involved in Alzheimer's disease?
  - A) A4
  - B) Amylin
  - C) Fibrin
4. Which amino acids has the highest pKa value?
  - A) Arginine
  - B) Lysine
  - C) Tyrosine
5. How many genes are coded in the genome of the MS-2 phage?
  - A) 3
  - B) 12
  - C) 54

Answers on Page 4

## Computational Chemistry:

### CNDO/2

CNDO is the abbreviation for Complete Neglect of Differential Overlap. CNDO is based on quantum chemistry, it is more specifically one of the first semi-empirical quantum chemistry methods. It uses two approximations:

- core approximation - only the outer valence electrons are explicitly included.
- zero-differential overlap

CNDO/2 is the main version of CNDO. The method was first introduced by John Pople and coworkers

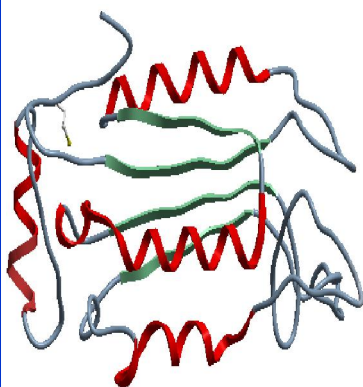
CNDO/1 and CNDO/2 were developed from Extended Huckel method (which explicitly ignores electron-electron repulsion terms) by explicitly including the electron-electron repulsion terms, but neglecting many of them, approximating some of them and fitting others to experimental data from spectroscopy. Quantum mechanics provides equations based on the Hartree-Fock method and the Roothaan equations that CNDO uses to model atoms and their locations. These equations are solved iteratively to the point where the results do not vary significantly between two iterations. It is noteworthy that CNDO does not involve knowledge about chemical bonds but instead uses knowledge about quantum wave functions.

CNDO can be used for both closed shell molecules, where the electrons are fully paired in molecular orbitals and open shell molecules, which are radicals with unpaired electrons. CNDO is considered to yield good results about partial atomic charges and dipole moment. Total energy and binding energy are calculated. Eigenvalues for calculating the highest occupied molecular orbital and lowest unoccupied molecular orbital are reported from the closed shell approach.

## Molecule of the Month

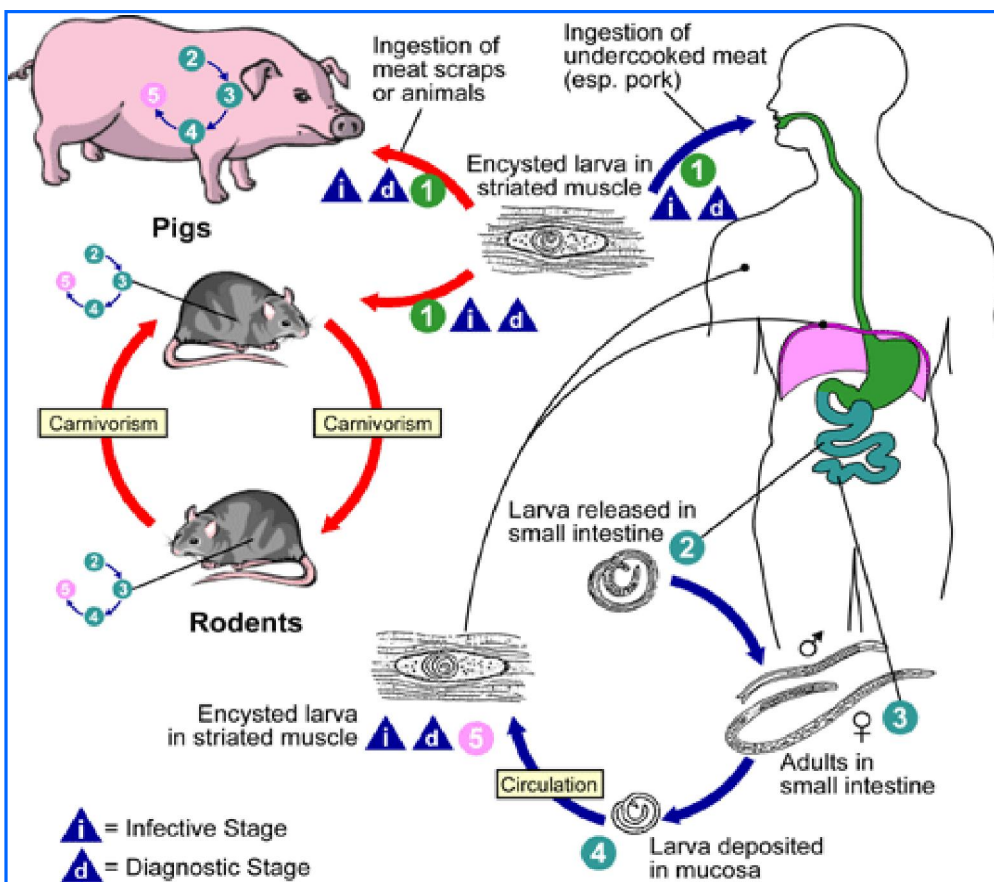
### Caspases

P-Caspases are the executioners of apoptosis. They are protein-cutting enzymes that chop up strategic proteins in the cell. The name refers to two properties of these enzymes. First, they are cysteine proteases that use the sulfur atom in cysteine to perform the cleavage reaction. Second, they cut proteins next to aspartate amino acids in their chains. They do not cut indiscriminately--instead, they are designed to make exactly the right cuts needed to disassemble the cell in an orderly manner. Almost a dozen caspases have been discovered in human cells, each with a slightly different task. Structures of many of them are available in the PDB.



### Molecular Data

PDB ID	: 1ICE
Amino acids	: 167
Exp. Method	: X-Ray Diff
Chains	: A & B (2)



### Bioinfy Animator:- Trichinellosis

Trichinellosis (trichinosis) is caused by nematodes (roundworms) of the genus *Trichinella*.

Trichinellosis is acquired by ingesting meat containing cysts (encysted larvae) [1] of *Trichinella*. After exposure to gastric acid and pepsin, the larvae [2] are released from the cysts and invade the small bowel mucosa where they develop into adult worms [3] (female 2.2 mm in length, males 1.2 mm; life span in the small bowel: 4 weeks). After 1 week, the females release larvae [4] that migrate to the striated muscles where they encyst [5]. Rats and rodents are primarily responsible for maintaining the endemicity of this infection. Carnivorous/omnivorous animals, such as pigs or bears, feed on infected rodents or meat from other animals. Different animal hosts are implicated in the life cycle of the different species of *Trichinella*. Humans are accidentally infected when eating improperly processed meat of these carnivorous animals.

### For suggestions & contributions contact:

Salam Pradeep; Project Assistant Level-II  
Email: salampradeep@gmail.com.

### Bioinfy Quiz

023

Answers

1 - A ; 2 - B ; 3 - A ; 4 - A ; 5 - A