

Bioinformatics up to Date

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

CONTENTS

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

Adviser:
Dr.P.G.Rao

Editors:
Dr. R.L.Bezbaruah
Miss Pompi Sharma
Mr Dhrubjyoti Gogoi

Bioinfo News

The E Coli EHEC genome sequenced by BGI and assembled by Nick Loman has been annotated with Era7 Bioinformatics BG7 bacterial genome annotation pipeline.

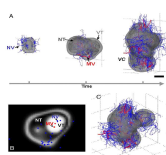
COVER STORY I

The Next Generation Sequencing

The current electrophoretic methodology for DNA Sequencing is a very expensive (\$5 million) to make genome sequencing ubiquitous. To overcome this problem a comprehensive, non-electrophoresis-based sequencing has been developed known as next-generation sequencing (NGS) that enable massive throughput in the gathering of genomic information. To date, several *de novo* assembly tools such as **SSAKE**, **VCAKE**, **Euler-sr**, **Edena**, **Velvet**, **ABYSS** and **SOAPdenovo** etc have been developed recently to assemble short sequencing reads generated by next-generation sequencing platforms. The data management in Next Generation Sequencing is a difficult task without proper training or workshop. In India the application of NGS is still under development.

COVER STORY II

Multiparameter Multiscale Computational Modeling of Tumor Growth and Drug Response



A novel predictive three-dimensional mathematical and computational model based on first-principle equations (conservation laws of physics) that describe mathematically the diffusion of cell substrates and other processes determining tumor mass growth and invasion. This model enables correlation of **glioma morphology to tumor growth** by quantifying interdependence of tumor mass on the microenvironment (e.g., hypoxia, tissue disruption) and on the cellular phenotypes (e.g., mitosis and apoptosis rates, cell adhesion strength).



Structure available
As of Tuesday Oct 25,
2011 at 5 PM
PDB: 76814

BIOINFO. QUIZ

- The first bioinformatics database was created by
 - Richard Durbin
 - Dayhoff
 - Michael j.Dunn
 - Pearson
- SWISSPROT protein sequence database began in
 - 1985
 - 1986
 - 1987
 - 1988
- An example of Homology & similarity tool?
 - PROSPECT
 - EMBOSS
 - RASMOL
 - BLAST
- The tool for identification of motifs?
 - COPIA
 - patternhunter
 - PROSPECT
 - BLAST
- First molecular biology server Expaty in the year?
 - 1991
 - 1992
 - 1993
 - 1994
- Deposition of cDNA into inert structure is
 - DNA finger printing
 - DNA polymerase
 - DNA probes
 - DNA microarrays

Answers on page 5

Computational Chemistry Chemgenome

Chemgenome is an **ab-intio gene prediction** software, which find genes in prokaryotic genomes in all six reading frames. The methodology follows a physico-chemical approach and has been validated on 372 prokaryotic genomes.

Method

DNA Space: The method takes complete or part of genome sequence of prokaryotic species in FASTA format as input file. It searches for genes based on physico-chemical properties of double-helical deoxyribonucleic acid (DNA).

Protein Space: The method takes the result generated from DNA space as input file and works as a filter based on stereochemical properties of protein sequences to reduce false positives.

Swissprot Space :The method takes the result generated from protein space as input file and calculates the standard deviation of a query nucleotide sequence (predicted gene sequence) with the swissprot proteins based on the frequency of occurrence of aminoacids.

Proteomics ALFIE

ALFIE predicts fragment sizes resulting from restriction endonuclease digestion and subsequent amplification in an AFLP reaction. All currently sequenced genomes are available for querying. A list of target sequences may also be supplied.

SilkSatDb: a microsatellite database of the silkworm, *Bombyx mori*



The SilkSatDb (silkworm microsatellite database) (<http://www.cdfd.org.in/silksatdb>) is a relational database of microsatellites extracted from the available expressed sequence tags and whole genome shotgun sequences of the silkworm, *Bombyx mori*.

The database has been rendered with a simple and robust web-based search facility, developed using PHP. The SilkSatDb also stores information on primers developed and validated in the laboratory. Users can retrieve information on the microsatellite and the protocols used, along with informative figures and poly-

Bioserver

SPODOBASE

SPODOBASE database provides integrated access to expressed sequence tags (ESTs) from the lepidopteran insect *S. frugiperda*. It is a publicly available structured database with insect pest sequences which will allow identification of a number of genes and comprehensive cloning of gene families of interest for scientific community. SPODOBASE is available from URL: <http://bioweb.ensam.inra.fr/>. So far, the SPODOBASE contains 29,325 ESTs, which are cleaned and clustered into non-redundant sets (2294 clusters and 6103 singletons). The SPODOBASE is constructed in such a way that other ESTs from *S. frugiperda* or other species may be added. User can retrieve information using text searches, preformatted queries, query assistant or blast searches. Annotation is provided against NCBI, UNIPROT or *Bombyx mori* ESTs databases, and with GO-Slim vocabulary.

> All ESTs **sequences** (79148) come from **8** different **libraries** :

Sf1F : Fat body.

Sf1H : Hemocyte.

Sf1M : Midgut.

Sf1P : Pool of various tissues.

Sf2H : Immune Challenged hemocytes.

Sf2L : Sf21 Cell lines sequences from R. CLEM.

Sf2M : Xenobiotic Induced Midgut.

Sf9L : Sf9 cell lines sequences.

SPODOBASE ■■

An integrated database for the genomics of the Lepidoptera *Spodoptera frugiperda*



Sequences database
Blast
Publications
Contributors
Silkworm Database
Contact

A fruitful collaboration :

Integrative biology and virology of insects (BIVI)

UMR INRA / UM II
Université Montpellier II - Case courrier 101 - Place Eugène Bataillon
34095 Montpellier cedex 5

Insect Genome Laboratory

National Institute of Agrobiological Sciences
1-2 Owashi, Tsukuba, Ibaraki
305-8634, Japon

UMR Biotic Interactions and Plant Health

UMR INRA CNRS / Université de Nice
400 route des Chappes - BP 167
06903 Sophia Antipolis Cedex



Computers for Biologists

SynBiN**Synonymous Bases in Nucleotide sequences**

http://www.hpa-bioinformatics.org.uk/cgi-bin/synbin/synbin_form.cgi

SynBiN is a tool that allows the user to determine synonymous and non-synonymous mutations in protein coding nucleotide sequences. This is achieved by comparing the translations of query sequences to a the translation of a selected reference sequence. More than one query sequence may be uploaded at one time, and different queries can be assigned to different reference sequences.

Step 1: Upload sequences and specify translation table

Step 2: Assign reference sequences

Step 3: Results

Synonymous mutations are identified where a triplet in the query sequence is shown, but no amino acid is displayed for the query at that position.

Non-synonymous mutations are identified where a triplet in the query is shown, and an amino acid is displayed for the query at that position. The amino acid at that position in the query sequence will differ from the corresponding amino acid in the reference sequence... hence a non-synonymous mutation.

Genomics

dbVar

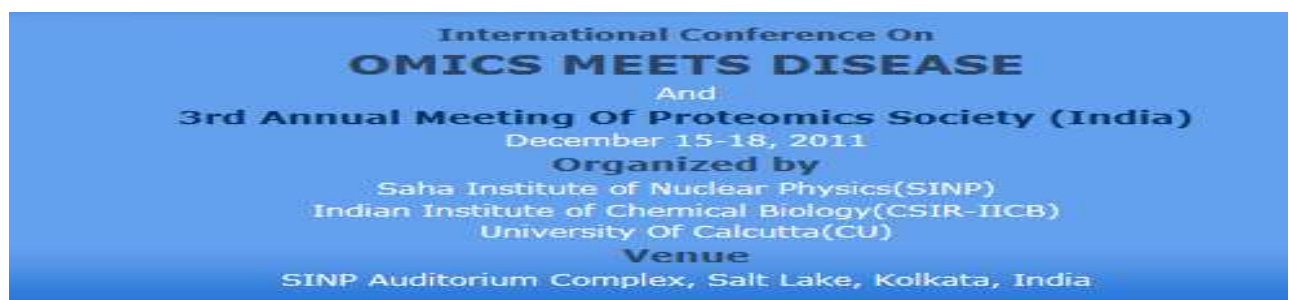
dbVar is a database of genomic structural variation that lets you search, view and download variant data from studies submitted on any organism or phenotype. It includes variant DNA ≥ 1 bp in size, provides access to the raw data (when available) and links to other NCBI and external resources. dbVar is a free resource that is developed and maintained by the National Center for Biotechnology Information (NCBI), at the U.S. National Library of Medicine (NLM), located at the National Institutes of Health (NIH).

dbVar Organism page

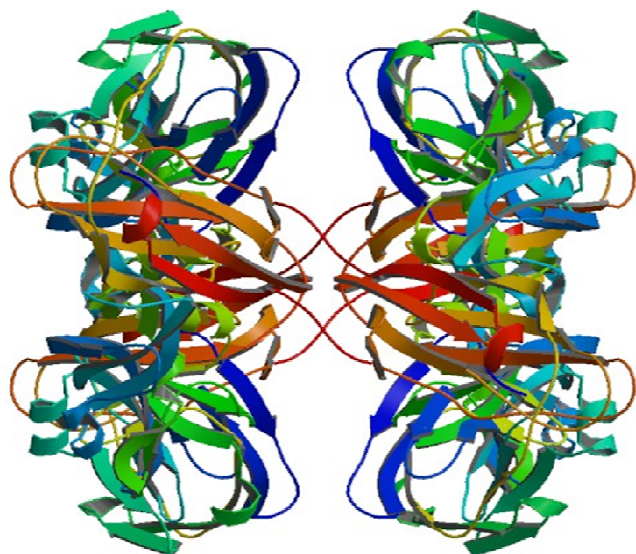
e.g., <http://www.ncbi.nlm.nih.gov/dbvar/organisms/9606>

The search by organism pages display a table of accessioned studies for the selected organism. The table contains the following information:

- Study ID — dbVar study accession, clicking the Study ID will take you to the Study page for that submission
- Submitter — last name of submitter
- Method — eg 'Oligo aCGH' or 'Paired-end mapping' (from a controlled vocabulary list)
- Number of Variants — total number of dbVar (#sv) variants in that study
- Number of Samples — total number of samples in that study
- Publication — link to the study publication(s) in PubMed

**Upcoming Events**

Bioinfo. Animator



Crystal Structure of the Japanese Encephalitis Virus Envelope Protein, strain SA-14-14-2.

Molecular Description

Classification: Viral Protein

Structure Weight: 43686.70

Molecule: envelope glycoprotein

Polymer: 1

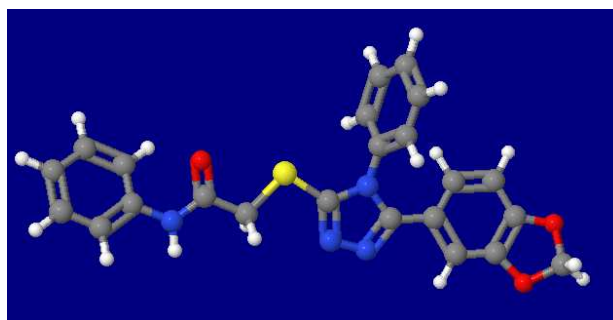
Type: protein

Length: 406

Chains: A

Fragment: JEV Envelope residues 295-700

Molecule of the month



2-[[5-(1,3-benzodioxol-5-yl)-4-phenyl-1,2,4-triazol-3-yl]sulfonyl]-N-phenylacetamide

Properties

Molecular Weight: 430.47902 [g/mol];n Molecular Formula: C₂₃H₁₈N₄O₃S; XLogP3-AA: 4.3; H-Bond Donor: 1; H-Bond Acceptor: 3

(<http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=2353020#x291>)

Answers of Bioinfo Quiz:) 1: c, 2:c, 3:d, 4:a, 5:c, 6:a.

Kindly send us your feedback to

Pompi Sharma and Dhrubajyoti Gogoi,

Joint Editor, Bioinformatics up to Date and Project Assistant (II),

Biotechnology Division, CSIR-North East Institute of Science and Technology, Jorhat, Assam

E-mail: pompi.sharma86@gmail.com /dhruba.bio.du@gmail.com

Mobile: 7896689444