

Bioinformation up to Date

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

Contents	Cover Story			
Cover Story 1	NEIST, Jorhat Gain Access to DBT's Electronic Library Consortium (DeLCON)			
Genomics 2	North-East Institute of Science & Technology & Bioinformatics Infrastructure Facility, Biotechnology Division, North-East Institute of Science of Technology, Jorhat-785006, Assam are now a member of the "DBT e-Library Consortium (DeLCON)" form this year 2010.			
Software Mania 3	The selected e-resources for NEIST, Jorhat & BIF, NEIST, Jorhat are given below:			
Bio Server 3	Sl. No	Name of Society/Journals	Url/Weblink	No. of access Journals
Bioinfy Quiz 3	1	American Association for Cancer Research (AACR)	http://www.aacr.org	08
Proteomics 4	2	American Chemical Society (ACS)	http://pubs.acs.org	37
Molecule of the Month 4	3	American Society for Biochemistry and Molecular Biology (ASBMB)	http://www.jbc.org	1
Bioinfy Animator 4	4	Lippincott William and Wilkins (LWW) / Wolter and Kluwer / OVID	http://ovidsp.ovid.com	11
Answer to Quiz 4	5	Marry ANN Liebert	http://www.liebertonline.com	7
Contact Us 4	6	Nature Publications	http://www.nature.com	40
	7	Oxford University Press (OUP)	http://www.oxfordjournals.org	18
	8	Springer India	http://www.springerlink.com	34
	9	Wiley-Blackwell	http://www3.interscience.wiley.com/cgi-bin/home	86
	10	Elsevier Science (ScienceDirect)	http://www.sciencedirect.com	415
	11	American Society for Plant Biologist (ASPB)	http://www.aspb.org/	2
	12	American Society for Microbiology (ASM)	http://www.asm.org/	12
	Further the DeLCON Core Library members have proposed to include the following journals in the near future for "North-East Region NER-DeLCON".			
	Sl No	Name of Society/Journals	Url/Weblink	No. of unaccess Journals
	1	American Association for Advance-ment of Science (AAAS)	http://www.sciencemag.org	1
	2	Annual Reviews (AR)	http://www.annualreviews.org	23
	3	Informa Healthcare / Taylor and Francis	http://www.informaworld.com	7
	4	Cold Spring Harbor Laboratory Press Journals (CSHL)	http://www.cshl.edu	4
	5	Society for General Microbiology (SGM)	http://mic.sgmjournals.org	3
	6	Society for Hematology (Blood)	http://bloodjournals.hematologylibrary.org	1

Adviser:

Dr. P.G. Rao

Editors:

Salam Pradeep Singh

Dr. R.L. Bezbaruah

Upcoming Events

1. Training programme on "Structural Biology & Bioinformatics in Drug Design" @ Bioinformatics Center, Biotechnology Park, Lucknow from October 7th-9th, 2010
2. Training programme on "Computational Approaches in Data Mining" @ Bioinformatics Center, Indian Agricultural Research Institute, New Delhi from 4th - 8th October, 2010

Genomics



BioCyc Database Collection

The BioCyc database collection is a set of 673 Pathway/Genome Databases. Each database in the BioCyc collection describes the genome and metabolic pathways information for individual organism. BioCyc is maintained by SRI International, in Menlo Park, California. As of June 2010, BioCyc contained databases for 500 genomes. The databases (DBs) within the BioCyc collection are organized into tiers according to the amount of manual review and updating they have received.

1) Tier 1 PGDBs: have been created through intensive manual efforts, and receive continuous updating. The BioCyc Tier 1 DBs are EcoCyc (Escherichia coli K-12) and MetaCyc (experimentally elucidated enzymes and metabolic pathways from more than 1,500 organisms.)

A) EcoCyc: EcoCyc is a bioinformatics database that describes the genome and the biochemical machinery of E. coli K-12 MG1655. The long-term goal of the project is to describe the molecular catalog of the E. coli cell, as well as the functions of each of its molecular parts, to facilitate a system-level understanding of E. coli. EcoCyc is an electronic reference source for E. coli biologists, and for biologists who work with related microorganisms. EcoCyc contains the complete genome sequence of E. coli, and describes the nucleotide position and function of every E. coli gene. Users can retrieve the nucleotide sequence of a gene, and the amino-acid sequence of a gene product.

B) MetaCyc: MetaCyc is a database of non-redundant, experimentally elucidated metabolic pathways. It stores predominantly qualitative information rather than quantitative data, although we have begun capturing some quantitative data such as enzyme kinetics data. A unique property of MetaCyc is that it is curated from the scientific experimental literature according to an extensive process, such that:

- (i) More than 1900 different organisms are represented
- (ii) The majority of pathways occur in microorganisms and plants
- (iii) More than 1530 metabolic pathways are stored, with more than 8,600 enzymatic reactions and more than 24,000 associated literature citations.

MetaCyc stores all enzyme-catalyzed reactions that have been assigned EC numbers by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). MetaCyc also stores thousands of additional enzyme-catalyzed reactions that have not yet been assigned an EC number

2) Tier 2 PGDBs: were computationally generated by the PathoLogic program, and have undergone moderate amounts of review and updating.

PathoLogic: PathoLogic predicts the metabolic pathways of an organism from its genome. In that sense, the metabolic pathways in the PathoLogic-based databases were derived computationally, in contrast to the literature-derived pathways in the EcoCyc and MetaCyc databases. PathoLogic also includes an operon predictor a predictor of what genes fill reactions with missing enzymes in predicted metabolic pathways, and a program that predicts what transport reactions are facilitated by transport proteins based on their functional description. The input required by PathoLogic is an annotated genome for the organism, such as in the form of a Genbank entry. The output produced by PathoLogic is a new Pathway/Genome Database for the organism.

3) Tier 3 PGDBs: were computationally generated by the PathoLogic program, and have undergone no review and updating. There are 643 DBs in Tier 3.

BioCyc Tools

The BioCyc Web site contains many tools for navigating and analyzing these databases, and for analyzing omics data, including the following.

- Genome browser
- Display of individual metabolic pathways, and of full metabolic maps
- Visual analysis of user-supplied omics datasets by painting onto metabolic map, regulatory map, and genome map
- Comparative analysis tools

The downloadable version of BioCyc that includes the Pathway Tools software provides more speed and power than the BioCyc Web site. Multiple database configurations are available for installation with the software including multiple E. coli and Shigella genomes, multiple Bacillus genomes, multiple Mycobacterium genomes, and multiple mammalian genomes.

Software Mania



Gene Designer

Gene Designer is a free bioinformatics software package. It is used by Molecular Biologists from academia, government and the pharmaceutical, chemical, agricultural and biotechnology industries to design, clone and validate genetic sequences.

Features:

- Gene Designer enables Molecular Biologists to capture the entire gene design process in one application, using a range of design tools.
- Algorithms for in silico cloning, codon optimization, back translation and Primer Design
- Graphic Molecular View to display, annotate and edit constructs
- Customizable database to store, manage and track genetic elements, Genes and constructs
- Drag-and-Drop interface for moving sequence elements within or between constructs
- Search feature for Sequence motifs, Restriction sites and Open reading frames.
- Codon optimize for recombinant protein expression in any organism using multiple algorithms.
- Remove or add Restriction sites or other Sequence motifs
- Recode Open reading frames
- Check translation frames and fusion junctions
- Design Oligonucleotides for sequencing primers, includes a real time melting point calculator
- Cloning Tool with drag-and-drop ability to cut, combine and clone insert and vector.

Bio Servers

mVISTA

mVISTA is a set of programs for comparing DNA sequences from two or more species up to megabases long and visualize these alignments with annotation information. mVISTA has a clean output, allowing for easy identification of sequence similarities and differences, and is easily configurable, enabling the visualization of alignments of various lengths at different levels of resolution. It is implemented as an on-line server that provides access to global pairwise, multiple and glocal (global with rearrangements) alignment tools. Stand-alone components (AVID and visualization module mVISTA) can be downloaded from the mVISTA Web site and used on our own computer. Source code of the LAGAN toolkit is freely available under the GNU Public license (GPL) and available through <http://lagan.stanford.edu>.

Alignment Programs in mVISTA

AVID: is a program for globally aligning DNA sequences of arbitrary length. The key features of the algorithm are that it can align hundreds of kilobases quickly, its accuracy and ability to detect weak homologies, and its ability to handle one of the sequences in draft by ordering and orienting the contigs automatically. The program works by recursively finding strong anchors from the collection of maximal matches in the sequences.

LAGAN: is a program for global pairwise and multiple sequence alignment of finished sequences or ordered and oriented draft merged in one contig. LAGAN performs progressive pairwise alignments, guided by a phylogenetic tree. Alignments are aligned to other alignments using the sum-of-pairs metric.

Inquiry

Your email address: *

<input type="button" value="Choose File"/> No file chosen	OR The GENBANK identifier(s): <input type="text" value="ACZF00000001"/>
<input type="button" value="Choose File"/> No file chosen	OR The GENBANK identifier(s): <input type="text" value="ACZF00000002"/>
<input type="button" value="Choose File"/> No file chosen	OR The GENBANK identifier(s): <input type="text" value="ACZF00000003"/>
<input type="button" value="Choose File"/> No file chosen	OR The GENBANK identifier(s): <input type="text" value="ACZF00000004"/>
<input type="button" value="Choose File"/> No file chosen	OR The GENBANK identifier(s): <input type="text" value="ACZF00000005"/>

Required fields are marked with *

Bioinfy Quiz - 025

1. Which of these enzymes contains Selenium (Se)?
 A) Alpha-amylase
 B) Glutathione peroxidase
 C) Nitrogenase component 1
2. Gas gangrene is due to a bacteria which cleaves collagen at:
 A) X | Gly-Pro-Y
 B) X-Gly | Pro-Y
 C) X-Gly-Pro | Y
3. What is the name of the database on 7-transmembrane G-linked receptors?
 A) 7TM
 B) GCRDb
 C) GLR7
4. Who developed the first analytical centrifuge?
 A) Buchner in 1897
 B) Svedberg in 1926
 C) Behrens in 1938
5. What is the hexamer that forms the telomeric repeat of human chromosomes?
 A) AAGGGT
 B) TTAGGG
 C) TTAGTA

Answers on Page 4

Proteomics:

PROTOMAP

PROTOMAP stands for PRotein TOpography and Migration Analysis Platform and was invented and developed by Ben Cravatt and colleagues at The Scripps Research Institute.

PROTOMAP is a recently developed proteomic technology for identifying changes to proteins that manifest in altered migration by one-dimensional SDS-PAGE. It is similar, conceptually, to two-dimensional gel electrophoresis and difference gel electrophoresis in that it enables global identification of proteins that undergo altered electrophoretic migration resulting from, for example, proteolysis or post-translational modification. However, it is unique in that all proteins are sequenced using mass spectrometry which provides information on the sequence coverage detected in each isoform of each protein thereby facilitating interpretation of proteolytic events.

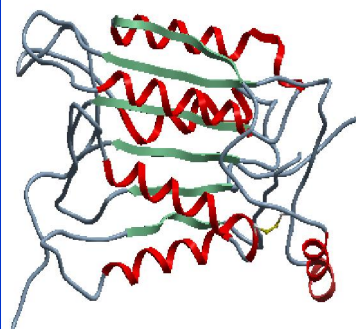
PROTOMAP is performed by resolving control and experimental samples in separate lanes of a 1D SDS-PAGE gel. Each lane is cut into evenly spaced bands (usually 15-30 bands) and proteins in these bands are sequenced using shotgun proteomics. Sequence information from all of these bands are bioinformatically integrated into a visual format called a peptograph which plots gel-migration in the vertical dimension (high- to low-molecular weight, top to bottom) and sequence coverage in the horizontal dimension (N- to C-terminus, left to right). A peptograph is generated for each protein the sample (thousands of peptographs are generated from a single experiment) and this data format enables rapid identification of proteins undergoing proteolytic cleavage by making evident changes in gel-migration that are accompanied by altered topography.

Molecule of the Month

Caspases

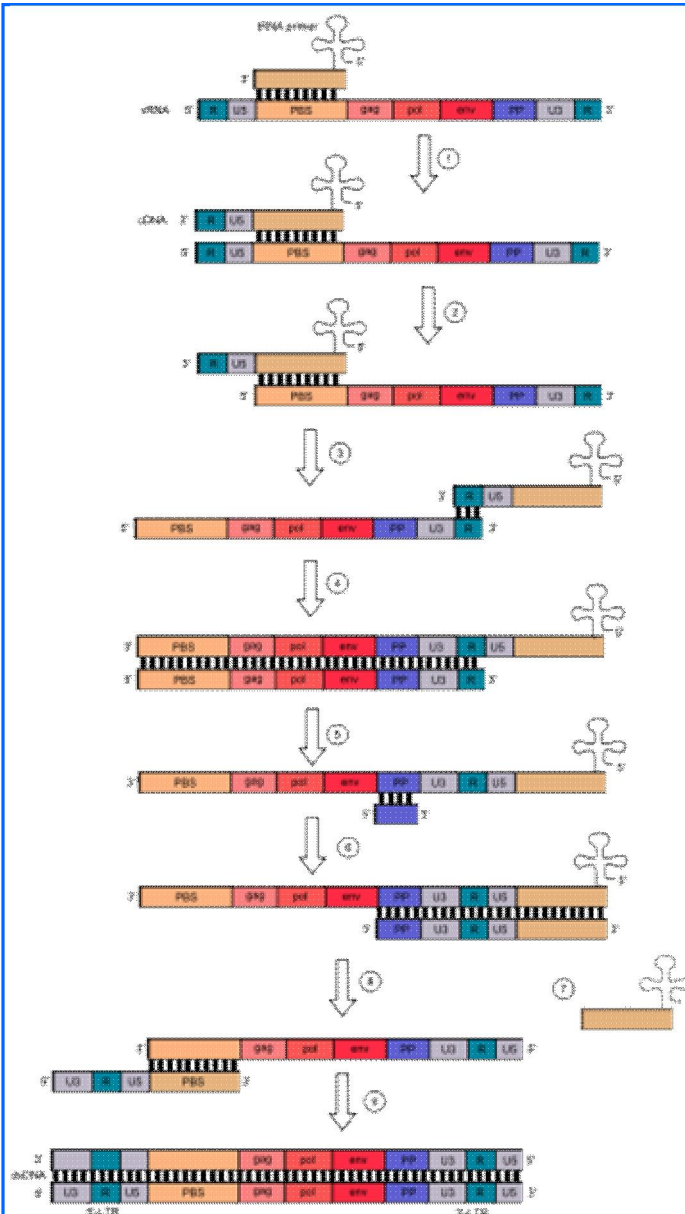
Caspases are the executioners of apoptosis. They are protein-cutting enzymes that chop up strategic proteins in the cell. The name refers to two properties of these enzymes. First, they are cysteine proteases that use the sulfur atom in cysteine to perform the cleavage reaction. Second, they cut proteins 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. Caspase-1 also known as interleukin-1beta-converting enzyme shown below was the first one discovered. It is not involved directly in apoptosis, but instead processes a cell signaling molecule in WBC.



Molecular Data

PDB ID	: 1ICE
Amino acids	: 255
Exp. Method	: X-Ray Diff
Chains	: 2 (A & B)



Bioify Animator:- Reverse Transcription

Mechanism of reverse transcription in class VI virus ssRNA-RT, human immunodeficiency virus (HIV).

Reverse transcription occurs in the cytoplasm of host cell. In this process, viral ssRNA is transcribed by the viral reverse transcriptase (RT) into double stranded DNA. Reverse transcription takes place in 3' 5' direction. tRNA ("cloverleaf") hybridizes to PBS and provides -OH group for initiation of reverse transcription. 1) Strong stop complementary DNA (cDNA) is formed. 2) Template in RNA:DNA hybrid is degraded by RNase H domain of reverse transcriptase 3) DNA:tRNA is transferred to the 3'-end of the template (synthesis "jumps"). 4) First strand synthesis takes place. 5) The rest of viral ssRNA is degraded by RNase H, except for PP site. 6) Synthesis of second strand of ssDNA is initiated from the 3'-end of the template. tRNA is necessary to synthesis of complementary PBS 7) tRNA is degraded 8) After another "jump", PBS from the second strand hybridizes with the complementary PBS on the first strand. 9) Synthesis of both strands is completed by the DNAP function of reverse transcriptase. Both dsDNA ends have U3-R-U5 sequences, so called long terminal repeat sequences (3'LTR and 5'LTR, respectively). LTRs mediate integration of the retroviral DNA into another region of the host genome.

Key: U3 - promoter region, U5 - recognition site for viral integrase; PBS - primer binding site; PP - polypurine section (polypurine tract); gag, pol, env - see HIV genome organisation). Colors mark complementary sequences. This diagram isn't drawn to scale.

For suggestions & contributions contact:

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

Bioify Quiz

025

Answers

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