



Bioinformation up to Date

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

Contents

Cover Story	1
Readers Contribution	1
Readers Contribution	2
Proteomics	2
Genomics	2
Software Mania	3
Bio Server	3
Current Trends	3
Bioinfy Quiz	3
Computers for Biologist	4
Molecule of the Month	4
Bioinfy Animator	4

Adviser

Dr. P.G. Rao

Editors

Salam Pradeep Singh

Dr. R.L. Bezbaruah

Research Labs in North-East

- North East Institute of Science and Technology, Jorhat
- North East Regional Institute of Science and Technology, Itanagar.
- Institute of Bioresource and Sustainable Development, Imphal.

Cover Story

Bioinformatics Infrastructure Facility Activity

Since the Department of Biotechnology, Government of India, has sanctioned with a release of Rs. 20 lacs for establishment of Bioinformatics Infrastructure Facility (BIF) for the promotion of Biology Teaching Through Bioinformatics (BTBI), the activity has already began at the Biotechnology Division, North East Institute of Science and Technology, Jorhat, Assam. On 28th May, 2008 at around 3:00 pm there was a presentation on “Sequence Analysis and Alignment - Introduction to biological sequence databases”. There was also a demonstration on bioinformatics software called CLC Sequence Viewer.

On 27th June, 2008 another presentation was delivered on “Central Dogma of Molecular Biology” which includes animated videos. *.....Biotechnology Division, NEIST*

Readers Contribution

NEIST (CSIR) Jorhat has submitted a few partial sequence of 16SrRNA genes in NCBI (National Center for Biotechnology Information) GenBank, Bethesda, USA proving our contribution in the biotechnology research. The work is a part of exploratory work to contribute in basic microbial taxonomy from diversity rich Indo-Burma hot zone to enrich both the national and international gene bank.

Partial sequence of these 16SrRNA gene of the following functionally important indigenous strains were accepted and can be accessed via NCBI-GenBank database.

- DQ299944:** *Streptomyces rimosus* strain RRLJ AY-1 16S ribosomal RNA gene, partial sequence. gi|83317141|gb|DQ299944.1|83317141]
- DQ299945:** *Bacillus pumilus* strain RRLJ SMAD 16S ribosomal RNA gene, partial sequence. gi|83317142|gb|DQ299945.1|83317142]
- DQ299946:** *Paenibacillus lautus* strain RRL AY-2 16S ribosomal RNA gene, partial sequence. gi|83317143|gb|DQ299946.1|83317143]
- DQ166950:** *Pseudomonas beteli* strain RRLJ SMAR 16S ribosomal RNA gene partial sequence. Gi83317144|gb|DQ299947.1|[83317144]
- DQ166951:** *Streptomyces sp.* SM-1 RRLJ, 16S ribosomal RNA gene, partial sequence. gi|74058446|gb|DQ166950.1|[74058446]
- EU165705:** *Aspergillus elegans* strain SM01, 18S ribosomal RNA gene partial sequence. [gi:157429025|gf|EU165705.1]
- EU159585:** *Brevibacillus laterosporus* strain BPM3, 16S ribosomal RNA gene, partial sequence. [gi:157460225|gb|EU159585.1]
- EU165269:** *Hypocrea schweinitzii* strain DF-05, 18S ribosomal RNA, partial sequence. gi|157956148|gb|EU165269.1|[165269]
- EU513377:** *Streptomyces tanashiensis* A2D, 16S rRNA gene partial sequence. gi|169787630|gb|EU513377.1|[169787630]
- EU428823:** *Pseudomonas aeruginosa* strain RRLJH2, 16S rRNA gene partial sequence. gi|167887761|gb|EU167887761.1|[426623]
- EU278596:** *Streptomyces sannaniensis* strain 118, 16S rRNA gene partial sequence. gi|161377616|gb|EU278596.1|[278596]

.....Contributed by Dr. T.C. Bora, Scientist F, Biotechnology Division

Readers Contribution

Amino Acid Applications

The industrial application of amino acids dates back to the late 1950's- 60's. Of the 20 standard proteogenic amino acids 8(Ile, Leu, Lys Met, Phe, Thr, Try, Val) are essential amino acids since the body cannot synthesize them and so they must be obtained through food. Amino acids for feed now play very important roles in improving the efficiency of protein utilization in animal feeding. DL-methionine, produced by chemical synthesis began finding its way into poultry feed. Production of L-Lys by fermentation was started in Japan during 1960's. Amino acids like L-Lys HCl, L-Thr and L-Try are produced by the fermentation method. By cultivating a special microbial strain developed for the production of each amino acid, in medium containing glucose or sugar and other nutrients, an amino acid can be efficiently produced. The amino acid is extracted from the fermentation broth with ion exchange resin treatment, etc. The fermentation yield of the strain is a key factor in the productivity of amino acids, which has been steadily improving with the introduction of improved biotechnology tools. Amino acids, for e.g. L-Lys, have contributed a lot in increasing the efficiency of animal protein production by improved conversion of proteins from feed to meat. Nitrogen excretion due to animal farming is a threat to human health through ammonia or nitrate pollution in soil and water. Decreasing the excessive protein in feed by supplementation of amino acids is a cost effective way to solve the problem of nitrogen accumulation associated with animal feeding. Thus improving the efficiency of protein utilization in animal feeding with the application of amino acids for feed will become important in securing the protein supply and also protecting the environment.

Optically pure amino acids are important for many pharmaceutical industries and food products. L-amino acids for e.g. L-Met, L-Lys (feed additives), Aspartyl-phenylalanine-1-methyl ester (Aspartame, a low calorie artificial sweetener), L-homophenylalanine (Enentopril, a pharmaceutical), L-dihydroxyphenylalanine (treatment for Parkinsonism) and D-amino acids for e.g. D-phenylglycine, D-Ser (antibiotics Ampicillin and Amoxicillin), D-Ala (artificial sweetener, Alitame), D-Val (agrochemical, Fluvinate) have found important uses in various industries. An important aspect of production of optically pure amino acids is that chemical synthesis of amino acids yields a racemic mixture of DL-amino acids, which needs to be resolved to optically pure form before they can serve their purpose in pharmaceutical and other industries. The racemic DL-amino acids produced by chemical or fermentation methods can be resolved to pure form by use of the enzymes (biocatalytic resolutions). The enzymes important in this regard are proteinases, amino peptidases, aminoacylases, D-hydantoinases and most importantly amino acid oxidases (D-amino acid oxidase or L-amino acid oxidases)

.....Contributed by Susmita (CSIR-JRF), Biotechnology Division

Proteomics

Protein Information Resource



The Protein Information Resource (PIR), located at Georgetown University Medical Center, Washington, D.C. is an integrated public bioinformatics resource to support genomic and proteomic research, and scientific studies.

It was established in 1984 by the National Biomedical Research Foundation (NBRF) as a resource to assist researchers in the identification and interpretation of protein sequence information.

PIR has provided many protein databases and analysis tools freely accessible to the scientific community, including the Protein Sequence Database (PSD), the first international database, which grew out of Atlas of Protein Sequence and Structure.

In 2002, PIR along with its international partners, EBI (European Bioinformatics Institute) and SIB (Swiss Institute of Bioinformatics), were awarded a grant from NIH (National Institute of Health) to create UniProt, a single worldwide database of protein sequence and function, by unifying the PIR-PSD, Swiss-Prot, and TrEMBL databases.

PIR offers a wide variety of resources mainly oriented to assist the propagation and standardization of protein annotation:

PIRSF - Protein family classification system

iProClass - Integrated protein knowledgebase

iProLINK - Literature, information & knowledge

Courtesy: Protein Information Resource, USA

Genomics

HGP - Human Genome Project

The Human Genome Project (HGP) is an international scientific research project with a primary goal to determine the sequence of chemical base pairs, which make up DNA and to identify the approximately 25,000 genes of the human genome from both a physical and functional standpoint.

The project began in 1990 initially headed by James D. Watson at the U.S. National Institutes of Health. A working draft of the genome was released in 2000 and a complete one in 2003. A parallel project was conducted by the private firm Celera Genomics. Most of the sequencing was performed in universities and research centers from the United States, Canada and Britain. The mapping of human genes is an important step in the development of medicines and other aspects of health care.

While the objective of the Human Genome Project is to understand the genetic makeup of the human species, the project also has focused on several other nonhuman organisms such as E. coli, the fruit fly, and the laboratory mouse. It remains one of the largest single investigational projects in modern science

The HGP originally aimed to map the nucleotides contained in a haploid reference human genome. Several groups have announced efforts to extend this to diploid human genomes including the International HapMap Project, Applied Biosystems, Perlegen, Illumina, JCVI, Personal Genome Project, and Roche-454.owser.

Courtesy: Human Genome Project, USA

Software Mania

CLC Sequence Viewer

CLC Sequence Viewer is a software package for performing basic bioinformatics work. This software is publicly available. It is developed by CLC bio, Aarhus, Denmark. It is a platform independent software and works on Apple-Mac, MS-Windows and Linux and also data can be exported or imported between these different platforms. The latest version was released in march 2008. There are also commercial versions available namely CLC DNA Workbench, CLC Protein Workbench, CLC RNA Workbench etc. CLC Sequence Viewer can be download from the website:

<http://www.clcbio.com/>

Some of the important toolbox in CLC Sequence Viewer are:

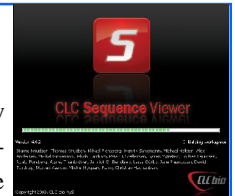
Alignments and Tress - Create multiple alignments of nucleotide and protein sequence for building phylogenetic trees.

General Sequence Analyses - Create sequence statistics, shuffle and join a nucleotide or protein sequence.

Nucleotide Analyses - Convert DNA to RNA & RNA to DNA, translate to proteins and find open reading frames.

Restriction Site Analyses - Find restriction site and annotate them on a sequence.

Database Search - Search sequences at NCBI-GenBank (National Center for Biotechnology Information)



Courtesy: CLC bio, Aarhus, Denmark

Bio Servers

Q-Site Finder

Q-SiteFinder is an energy-based method for the prediction of protein-ligand binding sites. It uses the interaction energy between the protein and a simple van der Waals probe to locate energetically favourable binding sites. Energetically favourable probe sites are clustered according to their spatial proximity and clusters are then ranked according to the sum of interaction energies for sites within each cluster.

Q-SiteFinder is effective in generating predicted sites that map accurately onto ligand coordinates. It also generates predicted sites with the lowest average volumes of the methods examined. Unlike pocket detection, the volumes of the predicted sites appear to show relatively low dependence on protein volume and are similar in volume to the ligands they contain. Restricting the size of the pocket is important for reducing the search space required for docking and de novo drug design or site comparison. The method can be applied in structural genomics studies where protein binding sites remain uncharacterized. Q-SiteFinder can be accessed freely online at

<http://www.bioinformatics.leeds.ac.uk/qsitefinder>

Courtesy: University of Leeds, U.K.

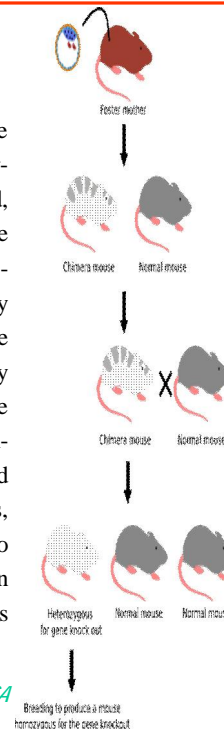
Bioinfy Quiz - 002

1. How many bases are there in an Okazaki fragment.
 - a) 1000
 - b) 10000
 - c) 100
2. Who first formulated the concept of messenger RNA.?
 - a) Sol Spiegelman
 - b) Jacob & Monod
 - c) Arthur Kornberg
3. Down syndrome is due to an extra copy of which human chromosome?
 - a) Chromosome 21
 - b) Chromosome X
 - c) Chromosome 13
4. Pasteurization is a liquid sterilization process which consists in:
 - a) Heating at 45°C for 10 mins
 - b) Heating 62°C for 30 mins
 - c) Steaming at 400°C for 2 min
5. Which amino acid when exposed to ninhydrin produces a brown color?
 - a) Asparagine
 - b) Alanine
 - c) Proline

Current Trends

Knockout Mouse

A knockout mouse is a genetically engineered mouse in which one or more genes have been turned off through a gene knockout. Knockout mice are important animal models for studying the role of genes which have been sequenced, but have unknown functions. By causing a specific gene to be inactive in the mouse, and observing any differences from normal behaviour or condition, researchers can infer its probable function. Mice are currently the most closely related laboratory animal species to humans, for which the knockout technique can easily be applied. They are widely used in knockout experiments, especially those investigating genetic questions that relate to human physiology. Gene knockout in rats is much harder and has only been possible since 2003. Examples of research in which knockout mice have been useful include studying and modeling different kinds of cancer, obesity, heart disease, diabetes, arthritis, substance abuse, anxiety, aging and Parkinson's disease. Knockout mice also offer a biological and scientific context in which drugs and other therapies can be developed and tested. Millions of knockout mice are used in experiments each year.



Courtesy: National Human Genome Research Institute, USA

BioPython

Python is a very high-level programming language. Its design philosophy emphasizes programmer productivity and code readability. Python's core syntax and semantics are minimalist, while the standard library is large and comprehensive. It supports multiple programming paradigms and features a fully dynamic type system and automatic memory management; similar to Perl, Ruby, Scheme, and TCL.

Python was first released by Guido van Rossum in 1991. The language has an open, community-based development model managed by the non-profit Python Software Foundation and Biopython Project is one of such kind .

BioPython Project is an international association of developers of freely available Python tools for computational molecular biology and bioinformatics. It is a distributed collaborative effort to develop Python libraries and applications which address the needs of current and future work in bioinformatics. The source code is made available under the Biopython License, which is extremely liberal and compatible with almost every license in the world. It works along with the Open Bioinformatics Foundation, who generously provide web and CVS space for the project. The latest release is Biopython 1.46, released on 29 June 2008.

Courtesy: www.biopython.org/

Molecule of the Month

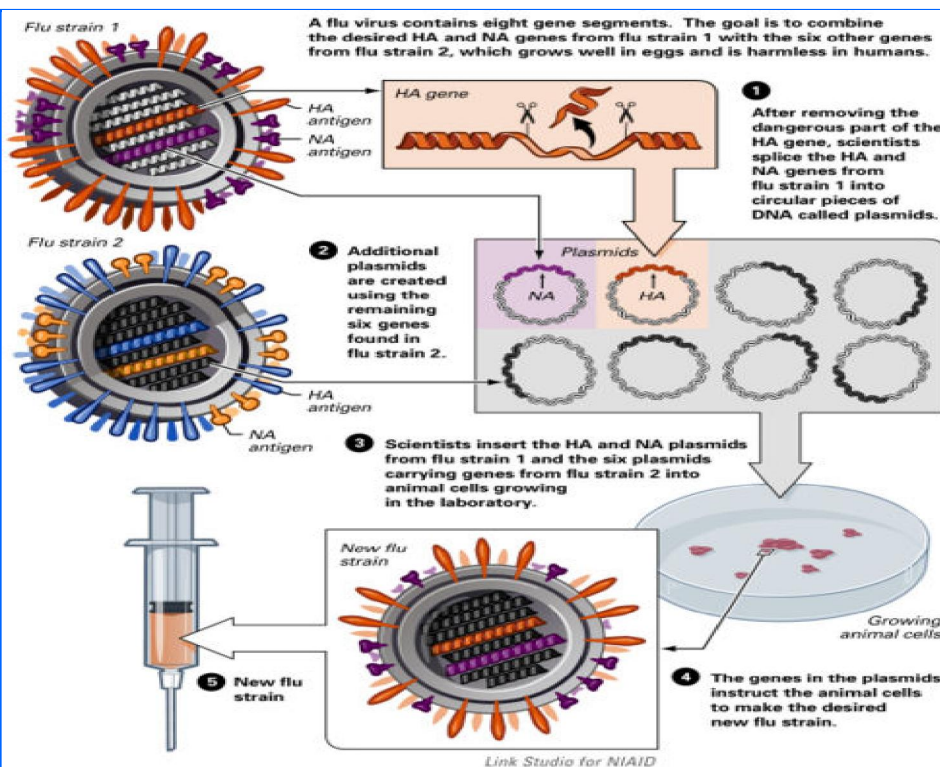
Neuraminidase (N9)

Neuraminidase is a glycoside hydrolase enzyme. It is frequently found as an antigenic glycoprotein on the surface of Influenza virus. It exist as a mushroom shape projection on the surface of the influenza virus. It has a head consisting of 4 co-planar & roughly spherical sub-units, and a hydrophobic region that is embedded within the interior of the virus membrane. It has functions that aid in the efficiency of virus release from cells. It cleaves terminal sialic acid residues from carbohydrate moieties on the surfaces of infected cells. Most of the influenza antiviral are neuraminidase inhibitors.



Molecular Data

PDB ID	: 7NN9
Amino acids	: 388
Atoms	: 3068
Exp. Method.	: X-Ray Diffraction
Chains	: A (1)



Bioinfy Animator - Avian Flu vaccine developed by Reverse genetics techniques.

Courtesy: National Institute of Allergy & Infectious Diseases, Maryland

Biotechnology Division

Equipped with all the modern laboratory facilities and technology, Biotechnology Division of NEIST, Jorhat is one of the CSIR research laboratories working in diverse areas of biotechnology research. There are ten scientists, four technical officials and fewer than a twenty research scholars & project assistants dedicating their body and soul both physically and mentally. The scientist in this division are working in research areas including microbial diversity & bioactive molecules, biotransformation, plant tissue culture, molecular biology & biochemistry, soil biotechnology etc. The division has completed various in-house projects as well as other projects in collaboration with firms like ONGC, Oil India Corporation etc. The division has also provided various training programs to post graduate students in the area of biotechnology. Salam Pradeep

Please contribute to this bulletin, please contact:

Salam Pradeep Singh ; Email: salampradeep@gmail.com
 C/o Dr. R.L. Bezbaruah ; Biotechnology Division, NEIST, Jorhat

Answers to Bioinfy Quiz

1-a ; 2-b ; 3-a ; 4-b; 5-a