BIOINFORMATICS

ABSTRACT

Bioinformatics is the applications of computer technology to the management of biological information. Computers are used to gather, store, analyze and integrate biological and genetic information, which can then be applied to gene-based drug discovery, and development .The technology has advanced so much that we are in a position to detect and cure some genetical disorders.

In this paper, we have discussed the fundamentals of bioinformatics. We have discussed how large genomic data banks can be maintained and used. The latest trends such as the Micro-array technology have been discussed. We have developed algorithms comparing the DNA-protein encoding to computer program execution. We also depicted the cell as a state machine.

• INTRODUCTION

Bioinformatics is the field of science in which biology, computer science, and information technology merge to form a single discipline. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned. At the beginning of the "genomic revolution", a bioinformatics concern was the creation and maintenance of a database to store biological information, such as nucleotide and amino acid sequences.

The term that was coined to encompass computer applications in biological sciences is bioinformatics.

Bioinformatics= Biotechnology + Information technology

The term bioinformatics is now used to mean rather different things, from artificial intelligence and robotics to genome analysis. The term was originally applied to the computational manipulation and analysis of biological sequence data (DNA and/or protein), but now tends also to be used to embrace the manipulation and analysis of 3D structural data.

DEFINITION:

"The mathematical, statistical and computing methods that aim to solve biological problems using DNA and amino acid sequences and related information."

ORIGIN & HISTORY OF BIOINFORMATICS:

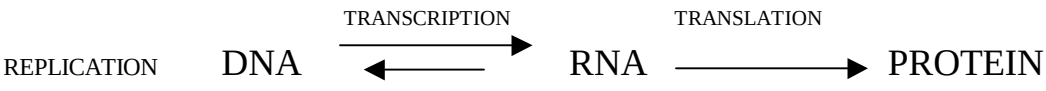
Over a century ago, bioinformatics history started with an Austrian monk named Gregor Mendel. He is known as the "Father of Genetics". He cross-fertilized different colors of the same species of flowers. Mendel illustrated that the inheritance of traits could be more easily explained if it was controlled by factors passed down from generation to generation.

Marvin Carruthers and Leory Hood made a huge leap in bioinformatics when they invented a method for automated DNA sequencing. In 1988, the Human Genome organization (HUGO) was founded. This is an international organization of scientists involved in Human Genome Project. In 1989, the first complete genome map was published of the bacteria Haemophilus influenza. Bioinformatics was fuelled by the need to create huge databases, such as GenBank and EMBL and DNA Database of Japan to store and compare the DNA sequence data erupting from the human genome and other genome sequencing projects

BASICS OF BIOLOGY

DNA IS…

DNA(Deoxyribonucleic acid)is the genetic material of cell,carring information in a coded form from cell to cell and from parent to off-spring.when a gene is active,or expressed,it's information is copped first into another nucleic acid, RNA(ribonucleic acid),which inturn directs the systhesis of the ultimate gene products, the specific proteins.RNA is also the genetic material of some viruses that contain RNA instead of DNA.these concepts, which constitute the central dogma of molecular biology, were summarized by Francis Crick in the following diagram.

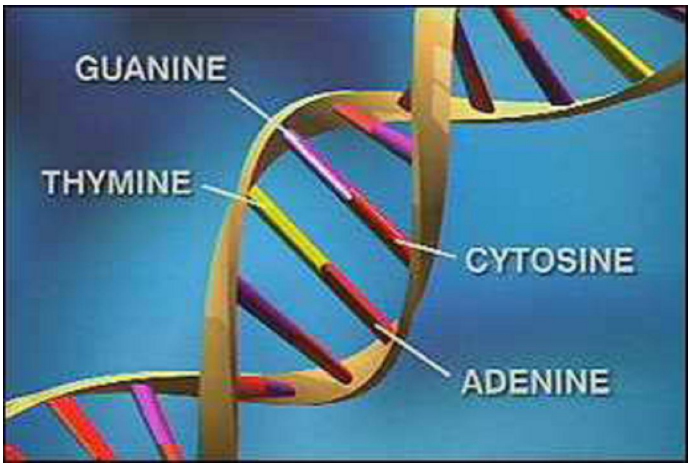


DNA is a double-stranded molecule twisted into a helix (think of a spiral staircase). Each spiraling strand, comprised of a sugar-phosphate backbone and attached bases, is connected to a complementary strand by non-covalent hydrogen bonding between paired bases. A and T are connected by two hydrogen bonds. G and C are connected by three hydrogen bonds.

Proteins are a very important constituent of the cell structure, they regulate cell processes by switching them on and off, sometimes they are the enzymes which catalyze the production of chemicals that manage the body, some proteins ferry small molecules within and outside the cell. Just as genes are made up of many molecules of DNA, proteins are macromolecules constituted by building blocks called amino acids of which only twenty are found in the living organisms. During synthesis of a protein, the 4-letter nucleic acid language of DNA is transcribed on to a RNA template, which in turn translates this into a 20-letter protein language, three letters in DNA code for one amino acid. This called triplet code.

GENOME AND GENOMICS:

GENOME is the total genetic information possessed by an individual organism. Each cell contains a complete copy of the genome. Sequencing and characterization of the genome and analysis of the relationship between gene activities and cell functions are called as GENOMICS



GENE SEQUENCING

To identify these sequences many projects have begun…

The goal of any sequencing project is to obtain a genomic sequence and identify a complete set of genes; the ultimate goal is to gain an understanding of when, where how a gene is turned on, a process commonly referred to as a gene expression. Once we begin to understand where and how a gene is expressed under normal circumstances, we can then study what happens in an altered state, such as hereditary diseases. To accomplish the later goal we must identify and study the protein, or proteins, coded for by a gene.

For example, other laboratories may have published data that established a link between a particular protein and their disease of interest. Researchers would then work to isolate that protein determine its function and locate the gene that coded for the protein. Once the chromosomal location was determined, scientists would use biochemical methods to isolate the gene and its corresponding proteins.

Now, however, the time required to locate and fully describe a gene has rapidly decreased, due to new technology used to generate what are called expressed sequence tags, or ESTs. Expressed sequence tags are small pieces of DNA sequence (Usually 200 to 500 nucleotides long) that are generated by sequencing either one or both ends of an expressed gene.

EST's are powerful tools in the hunt for known genes as they greatly reduce the time required to locate a gene. Scientists first use observable biological clues to identify ESTs that may correspond to diseased candidates having genetical disorders. Scientists have been examining the DNA of diseased persons for mutations.

DNA 'MICRO ARRAY' TECHNOLOGY

The emergence of DNA micro array technology facilitates the identification and classification of this DNA sequence information and the assignment of functions to these new genes. A micro array works by exploiting the ability of a given mRNA molecule to bind specifically to, or hybridize to, the DNA template from which it originated. By using an array containing many DNA samples, scientists can determine (in a single experiment) the expression levels of hundreds or thousands of genes within a cell by measuring the amount of mRNA bound to each site on the array.

With the aid of a computer, the amount of mRNA bound to the spots on the micro array is precisely measured, generating a profile of gene expression in the cell. A micro array is a tool for analyzing gene expression that consists of a small membrane or glass slide containing samples of many genes arranged in a regular pattern.

Micro arrays are a significant advance both because they may contain a very large number of genes and because of their small size. Micro arrays are therefore useful when one wants to survey a large number of genes quickly or when the sample studied is small.

Micro arrays may be used to assay gene expression within a single sample or to compare a gene expression in two different cell types or tissue samples, such as in healthy and diseased tissue. Because a micro array can be used to examine the expression of hundreds or thousands of genes at once, it promises to revolutionize the way scientists examine gene expression.

CANCER: HEREDITARY DISEASE

Cancer is a genetical disease… The chromosomes of cancer cells are arranged incorrectly, or have large pieces missing. For example, on chromosomes 13, there seems to be a hotspot for damage at position 13q14 that is observed in many types of leukemia. Studying the genes found in this region will give clues about how the breakage is affecting the cell at the molecular level and hence by knowing this information, the corresponding drug can be used to remove the hotspot. Hence avoiding cancer.

Proteins help in, fighting infections, control body and in general, keep our bodies functioning smoothly. Identifying a protein shape, or structure, is key to understanding its biological function and its role in health disease.

BIOINFORMATICS PROJECTS

BIOJAVA:

The BioJava Project is dedicated to providing Java tools for processing biological data which includes objects for manipulating sequences, dynamic programming, file parsers, simple **statistical routines, etc.**

BIOPERL:

The BioPerl project is an international association of developers of Perl tools for bioinformatics and provides an online resource for modules, scripts and web links for developers of Perl-based software.

BIOLOGICAL DATABASES

GENBANK:

GenBank (Genetic Sequence Databank) is one of the fastest growing repositories of known genetic sequences. It has a flat file structure, which is an ASCII text file, readable by both humans and computers. In addition to sequence data, GenBank files contain information like accession numbers and gene names, phylogenetic classification and references to published literature

SWISSPROT:

This is a protein sequence database that provides a high level of integration with other databases and also has a very low level of redundancy.

GDB:

The GDB Human Genome Data Base supports biomedical research, clinical medicine, and professional and scientific education by providing for the storage and dissemination of data about genes and other DNA markers, map location, genetic disease and locus information, and bibliographic information.

EMBL:

The EMBL Nucleotide Sequence Database is a comprehensive database of DNA and RNA sequences collected from the scientific literature and patent applications and

directly submitted from researchers and sequencing groups. Data collection is done in collaboration with GenBank (USA) and the DNA Database of Japan (DDBJ).

SOFT WARES USED FOR DRUG DETECTION

FTDOCK 1.0: The growing number of individual structures in the crystallographic databases and the relatively small number of solved complexes makes predictive docking an important theoretical method. FTDOCK is a UNIX program for bimolecular docking using shape complementarity and electrostatics.

SEQPUP 0.8: SeqPup is a biological sequence editor and analysis program. It includes links to network services and external analysis programs. It is usable on common computer systems that support the Java 1.1 runtime environment, including Macintosh, MS-Windows and X-Windows.

W2H 2.1: A new release of the WWW interface to the GCG Sequence Analysis Software Package. W2H tries to cover as much functionality as possible, and to do it very user friendly. It gives you the opportunity to access more than hundred programs from any platform where Netscape

APPLICATION AREAS FOR BIOINFORMATICS

APPLICATIONS IN ENGINEERING FIELD:

In the bioinformatics area primary focus is on use of the information, as well as on acquisition, preparation, and storage. Which all needs to be considered within the framework of information for Acquisition of information and documents, including creation of meta-data, submission of electronic media, and communication interfaces

- Preparation of information and documents, including taxonomy classification, categorization, workflow processes, transformation, conversion, and indexing for search and retrieval.
- Storing archival of knowledge, data and file formats for acquisition, dissemination, reliability, maintainability, availability, and disaster recovery; quality of service; networking, security, and accessibility.

- Information retrieval and preservation, including: precision, recall, efficiency, and effectiveness of retrieval; search input methods, including multi-media index and search techniques; presentation and visualization of results; browse technologies.

REAL WORLD APPLICATIONS OF BIOINFORMATICS:

MOLECULAR MEDICINE:

Human genome will have profound effects on the fields of biomedical research and clinical medicine. Every disease has a genetic component and inherited to body's response to an environmental stress which causes alterations in the genome (e.g. cancers, heart disease, and diabetes.)

PERSONALIZED MEDICINE: This is the study of how an individual's genetic inheritance affects the body's response to drugs. At present, some drugs fail to make it to the market because a small percentage of the clinical patient population show adverse affects to a drug due to sequence variants in their DNA.

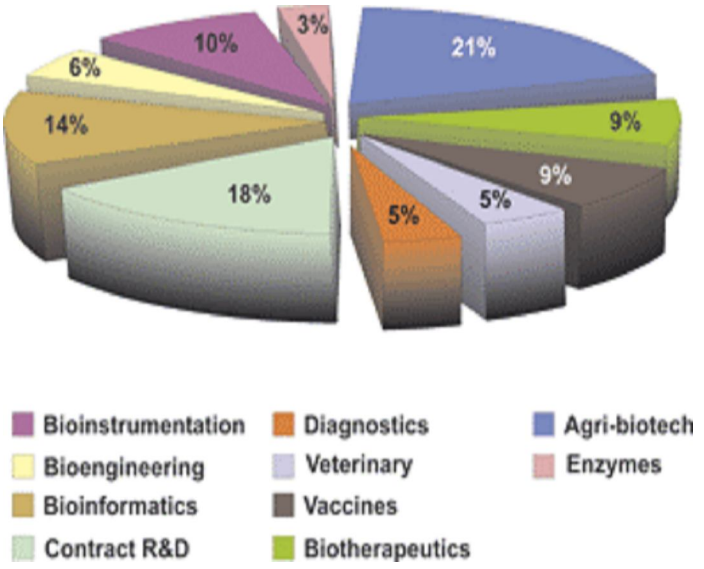
GENE THERAPY: In the not too distant future, the potential for using genes themselves to treat disease may become a reality. Gene therapy is the approach used to treat, cure or even prevent disease by changing the expression of a person's gene.

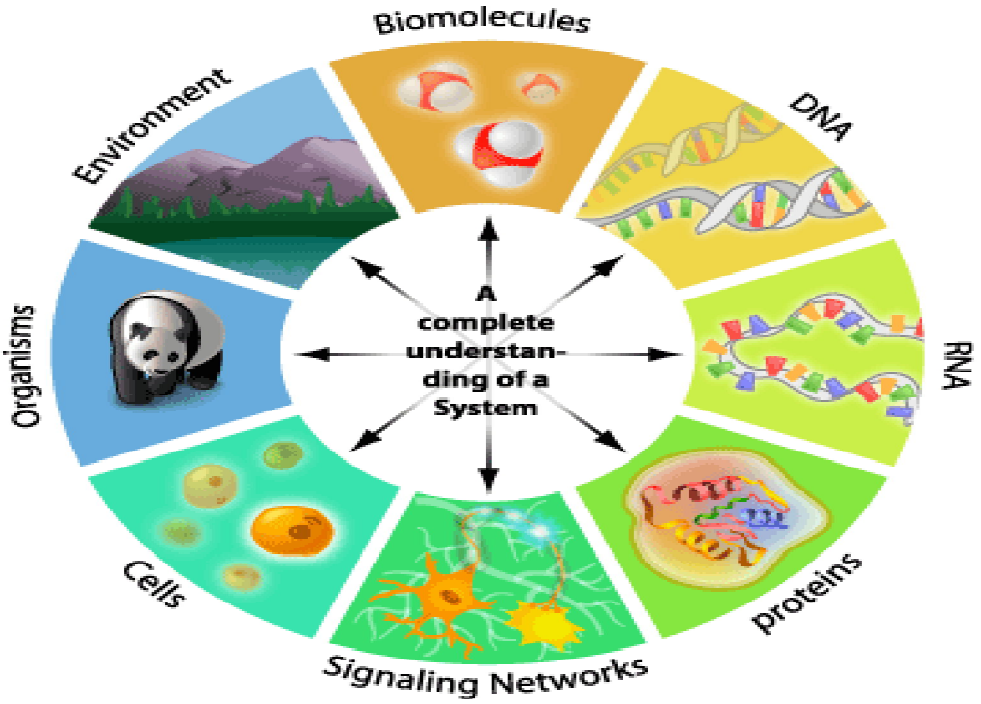
AGRICULTURE: The sequencing of the genomes of plants and animals should have enormous benefits for the agricultural community and can be used to search for the genes within these genomes and their functions, making them healthier, more disease resistant and more productive.

COMPARATIVE STUDIES: Analyzing and comparing the genetic material of different species is an important method for studying functions of genes, mechanisms of inherited diseases and species evolution.

COMMERCIAL APPLICATION OF BIOINFORMATICS:

- Definition of Bioinformatics Company.
    - Genomics in medicine, diseases monitoring, profiles for therapeutic molecular targeting.
- Diagnostics drug discovery, Pharmacogenomics and its applications.
- Proteomics in medicine and therapeutic target identification.





- **. The wheel of bioinformatics in biological environment**

FUTURE OF BIOINFORMATICS

Genomics is also being integrated into mainstream activities, and as this trend progresses, front-runners are starting to look beyond basic data integration to knowledge management, a fuzzy concept for scientists.

Bioinformatics aims genomics visualization but holding on to its dream of developing a platform.

In next-generation bioinformatics tools — going beyond things like self organizing maps and K-means

The most important new trend is the steady improvement in microarray data handling. Proteomics, metabolomics

CONCLUSION

It advances so much that we are in a position to detect and cure some genetical disorders. BLAST is an example of a Bioinformatics tool.

- DNA micro array technology facilitates the identification and classification of the DNA sequence information and assignment of functions to new genes.
- 21st century has seen a great boost in biotechnology like Human Cloning.

Thus, Bioinformatics is application of computer technology to the management of biological information

Under the guidance of…

K.RAMACHANDRA RAO (MCA DEPT)

BIBILOGRAPHY

- Bioinformatics Basics: Applications in Biological Science and Medicine By Hooman Rashidi and Lukas K. Buehler

- www.bio-itworld.com

- www.ncbi.nlm.nih.gov