Date
Page

Introduction to Bio informatics.

Bioinformatics is an interdisciplinary field mainly involving molecular biology and genetics, computer science, mathematics, and statics data intensive, large scale biological problems are addressed from a Computational point of view. The most common problem are modeling biological processes at the molecular level and making inferences from collected data. A Bioinformatics solutional usually involves the following steps.

- Collect statistics from biological data.
- Build a Computational model.
- Solve a Computational modeling problem.
- Test a evoluate a Computational algorithem.

Sequence analysis is the analysis of DNA and Protein sequence clues regarding function and includes subproblems such as identification of homologs, multiple sequence alignment, searching sequence patterns and evolutionary analysis.

Protein structure are three-dimensional data and the associated problems are structure prediction (secondary and tertiary), analysis of protein structure for clues regarding function and structural alignment.

Date .....
Page .....

Gene expression data is usually represented as matrices and analysis of micro array data mostly involves statics analysis, classification and clustring approaches, Biological networks such as gene regulatory networks, metabolic pathways and protein-protein interaction networks are usually modeled as graph and graph theoretic approaches are used to solve associated problems such as construction and analysis of large scale - networks.

History

The term "Bioinformatics" was invented by Paulien Hogeweg and Ben Hesper in 1970.

- The collection of amino acid sequence was complied in the Atlas of protein sequence and structure by the National bio medical foundation.
- This collection was edited by margaret O. Dayhoff O. Dayhoff from 1965 to 1978.
- Dayhoff and coworkers contributes to the consulsion of amino acid sequences by developing computer software for detecting distantly related sequence.

- The EMBL established their data library in 1980. to collect, organize and distribute nucleotide sequence data and related information.
- NCBI was established in U.S.A NCBI serves as primary information data bank and provides of information.
- The National Biomedical Research foundation established in PIR in 1984.

DNA Sequence

- The symbol used to represent DNA sequence Data.
- The four bases denoted by a single letter A (Adenin), C, (cytosine), T (Thymine) and ~~uracil~~. G (guanine)
- But often data sequence contain ambiguities in that it is not clear as to which of the four base present at several positions
- for example sequence data may indicate that the base present at a specific position may be either G or A, it is purine.
- Similarly, if a position may have either C or T it is pyrimidiene.
- The base sequence of the two complementary strand of a DNA molecules are represented by this system of a symbol.

Date .....
Page .....

Amino Acid sequence of protein :—

- The amino acid conventionally represented by three letter symbols .. eg - Ala for Alanine, Val for ~~Valani~~ Valine etc.
- But in Bioinformatics they are denoted by single letter eg A for alanine.. C - for Cytosine D as Aspartic Acid etc.
- But some positions in protein sequences have ambiguities this sequence is comparable to that for DNA sequences.
- for eg it may not be clear that a position has glutamine or glutamic acid the symbol is given the symbol Z.
- The protein synthesis begin at the N-terminus and proceeds to the C- terminus.
- The amino acid sequences in databases are listed from the N- terminus to the C-terminus of the polypeptide.

Date .....
Page .....

Types of Sequences in Nucleotide Sequence Databases.

- The databases on DNA sequences contain a different types cDNA sequences:-
- A cDNA molecule is obtained by reverse transcription of an RNA molecule.
- The cDNA sequences, therefore represent that part of the genome that is trancribed into RNA.
- If the cDNA is obtained from mRNA, it will represent only the exon sequences of the gene ~~sequence~~ expressed in the concerned cell / tissue / organisms.

Genomic DNA Sequence :-

- These sequence represent the complete genome of the organisms.
- When the genome sequences is completed, it will contain the sequences of the entire genome of the organisms.
- In case of prokaryotes genome consists of usually, a single chromosome while in case of eukaryotes it relates to the nuclear DNA.

#

Date .....
Page .....

Expressed Sequence Tag (EST) sequences:-

- The sequences are obtained by sequencing only a part of the cDNA molecules produced using mRNA.
- These sequence are dubbed as 'Tags' because they can be used as probe for the isolation of the concerned genes from the genomic DNA.
- This approach was used by J. Craig venter and his group for obtaining the sequence of expressed portion of human genome.
- The EST technique generated enormous sequence data that permitted the construction of a preliminary transcript map of the human genome.

Genome Sequence Tag (GST)

GST were developed for the the identifying the genes of plasmodium falciparum.

- It was observed that the enzyme mung bean nuclease (nmase) cuts P. falciparum genomic dna b/n genes.
- GSTs are developed by sequencing the DNA fragments on either side of the points of cuts generated by nmase.

Date .....
Page .....

Organellar DNA Sequence :—

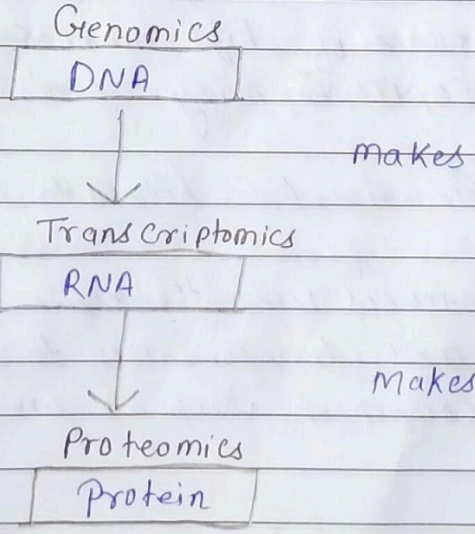
- Organeller DNA is the DNA found in mitochondria (mDNA) and chloroplast (cpDNA).
- The sequence of the data are complied in database.

Branches of Bioinformatics

A living cell is a system where cellular components such as genome, the gene transcript, and the proteins interact with each other, and these interactions determine the fact of the cell, e.g - whether a stem cell is going to become a liver cell or a cancer cell.

The Three Branches of Bioinformatics

(i) Genomics
(ii) Transcriptomics
(iii) Proteomics



Date .....
Page .....

Genomics

Genomics plays a significant role in modern biological research in which the nucleotide sequences of All the chromosome of an organisms are mapped and the location of different genes and their sequence is determined.

- This involves extensive analysis of the nucleic acids through molecular biology techniques before the data are ready for processing by computer.
- It is a science that attempts to describe a living organisms in terms of the sequence of its genome.
- It was not reliable to estimate the no of genes in an organism based on the no. of nucleotide base pair. because of the presence of high numbers of redudant copies of many genes.
- Genomics has helped to rectify this problem.
- Genomics uses technique of molecular biology and bioinformatics to identify Cellular Component such as proteins, rRNA, tRNA etc.

Date
Page

and analyse the sequence and non-coding sequence.

- The first automatic DNA sequencer was developed in 1986 by Leroy Hood.
- Haemophilus influenza was the first bacterium to be sequenced in 1995.
- Even if one can identify all the genes on a genome, the genes only indicate that at some point in time, it might be transcribed to produce cellular component.
- eg - A human genome contains about 30,000 to 60,000 protein coding genes, but only a subset of them is expressed in a particular cell type at a particular time.

#(ii) Transcriptomics

Transcriptomics is the study of the transptome which include the whole set of mRNA molecules in one of a population of biological cells.

- This study help us to depict the expression level of genes, often using techniques such as DNA micro arrays, that is capable of sampling ten thousands of different mRNAs at a time.

Date .....
Page .....

- This kind of new technique has helped biologist to routinely monitor the gene expression between the control cells and treatment cells.

- Transcriptomics has a few limitations.

- The relative abudance of transcripts as characterized by the sequential analysis of gene expression (SAGE) is missoadly experiments.

Proteomics / Προτεομικς

- proteomics represents the earliest to identify a major sub-class of a cellular component, the proteins and their interactions.

- Proteomics involves the sequencing of amino acid in a protein determining its 3D structure and relating it to the function of protein.

- Before computer processing comes into the picture, extensive data, particularly through crystallography and nuclear magnetic resonance. (NMR)

Date .....
Page .....

- With such data known as proteins, the structure and its relationship to the function of newly discovered protein.
- In such areas, bioinformatics has enormous analytical and predictive potential.
- Metabolic proteins such as haemoglobin and insulin have been subjected to intensive proteomic investigation.
- The term 'proteomics' was coined to make an analogy with genomics.
- Scientists feel that the bioinformatics of protein is crucial, to understands the cellular components and the interactions completely.

Aims of Bioinformatics

- The various important ways in which bioinformat
- Can be used. Such as include data acquisition
- tool and database development, data analysi and data integration, and data management.

Date .....
Page .....

Data Acquisition

- Data Acquisition is primarily concerned with accessing and storing data generated directly from the biological experiments

- The data generated by various sequencing projects have to be retrieved in the appropriate format, and capable of being linked to all information related to the DNA samples.

- The data are organized in different databases so that the researchers can access existing information.

Tool and Database development

- Many laboratories generate large volume of data such as DNA sequence, gene expression information, 3D molecular structure and highly-throughput screening.

- Consequently, they must develop effective databases for storing and quickly accessing data. The other aim is to develop tools and resources that aid in the analysis of data.

Date
Page

Data Analysis

- The third aim is to use these tool to analyse the data and interpret the results in a biologically meaningful manner. Efficient analysis require an efficiently designed database.

- It must allow researchers to place their Query effectively and provide them with all the information they need to begin their data analysis.

- If queries cannot be performed, or if the performance is too slow, the whole system breaks down since scientists will not inclined to use the databases.

Data Integration

- Once information has been analysed, a researcher must often associate or integrate it with the related data from the other databases.

- for example Scientists may run a series of gene expression analysis experiments and observe that a particular set of 100 genes is more highly expressed in a cancerous lung tissue than in a normal lung tissue.

- The Scientist may wonder which of the genes is most likely to be truly related to the disease.

Date .....
Page .....

Bioinformatics Application

Molecular medicine:-

- The human genome will have profound effects on the fields of biomedical research and clinical medicine. Every disease has a genetic component.
- This may be inherited as a result of the body's response to an environmental stress which causes alterations in the genome (eg- Cancer, heart disease, diabetes).
- The completion of the human genome means that we can search for the genes directly associated with different diseases and begin to understand the molecular basis of these diseases more clearly.
- This new knowledge of the molecular mechanism of disease will enable better treatment, cures and even preventive tests to be developed.

Date
Page

Personalised Medicine

- Clinical medicine will become more personalised with the development of the field of Pharmacogenomics.
- This is the study of how an individual's genetic inheritance affects the body's response to drug.
- At present, some drugs fails to make it to the market because a small percentage of the clinical patient population show adverse to a drug due to sequence variants in their DNA.
- As a result, potentially life saving drug never make it to the market place. Today's doctor have to use trial and error to the best drug to treat a particular patients as those with the same clinical, symptomps can show a wide range of respons to the same treatment.

Drug development

- At present all drugs on the market target only about 500 proteins.
- With an improved Understanding of disease mechanisms and using computational tool to identify and validate new drug targets, more specific medicines that act on the cause, not merely the symptoms of the disease can be developed.

Date
Page

- These highly specific drugs promise to have fewer side effect that many of today's medicines.

Gene therapy

- In the not too distant future, the potential for using genes themselves to treat diseases may become a reality.
- Gene therapy is the approach used to treat, cure or even prevent disease by changing the expression of a persons genes.
- Currently, this fields is in it's infantile stage with clinical trials for many different types of cancer and other diseases ongoing.

The Reality of Bio weapon :—

- Scientists have recently built the virus poliomyelitis using entirely artificial means.
- They did this using genomic data available on the internet and materials from a mail-order chemical supply.
- The research was financed by the US Department of defence as part of biowarfare response program to provide the world the reality of biowarfare

Date .....
Page .....

The researches also hope their work will discourage officials from ever relaxing programs of immunisation. This project has been met with very mixed feelings.

Antibiotic Resistance

- Scientists have been examining the genome of enterococcus faecalis leading cause of bacterial infection among hospital patients.

- They have discovered a virulence region made up a number of antibiotic - resistant genes that may contribute to the bacterium's transformation from a harmless gut bacteria to a menacing invades.

- The discovery of the region, known as pathogenicity island, could provide useful markers for detecting pathogens strains and help to establish controls to prevent the spread of infection in wards.

The End