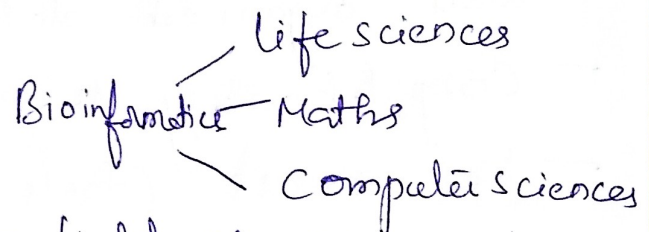


Bioinformatics

Introduction

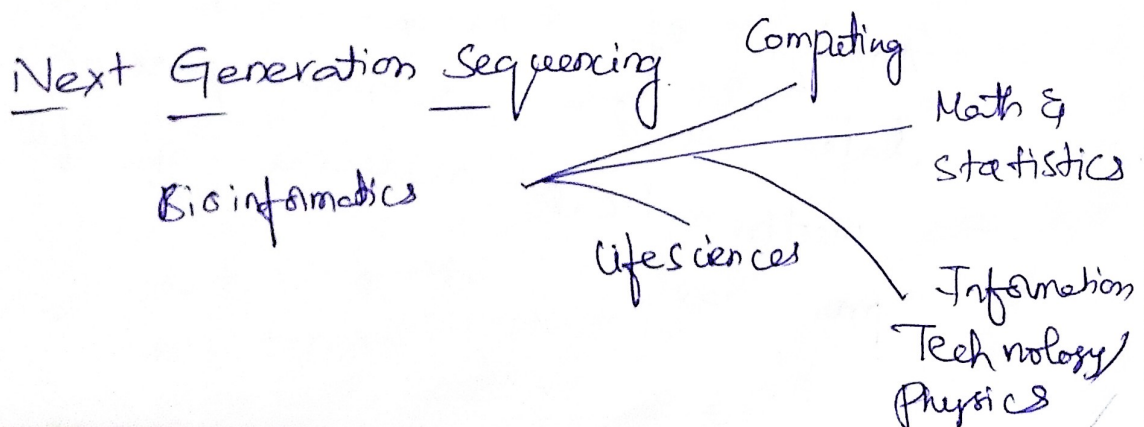
- ① Nucleic acids — DNA & RNA, Heredity
- ② proteins — Catalayst, transporters, structural protein - ~~Strength~~ strength, regulates, metabolism
- ③ lipids — fats, oils, cell signaling
- ④ carbohydrates — Human diet, Carbon sources

→ Databases, Implementation in Computer Sciences



→ Bioinformatics is a field of sciences in which biology, computer science, and information technology merge into single discipline to analyze biological information using computers & statistical techniques.

→ Term bioinformatics was coined by Paulien Hogeweg in 1979 for the study of informatics in biological systems



→ Biological expt can generate large amounts of data (macromolecular sequences, structures, expression.

→ Bioinformatics is about acquiring, managing, analyzing and understanding those data.

Major aspects

- + well organized databases
- Computationally derived hypothesis
- Web servers (tools / online applications)
- Virtual Screening of Compounds for drug development
- Big data (Next Generation Sequence) analysis.

Scope

- Genome and sequence analysis
- From sequence to 3D structural prediction
- Analysis of Genome-wide biomedical data and functional genomics.
- Mathematical modeling of life processes.

- Data base building and management

History:

- In 1978 - Paulien Hogeweg & Ben Hesper coined the term Bioinformatics

- 1987 - Human Genome Project
- 3 billion characters in the DNA Sequence - Human genome

- Protein databases - 1st prepared then nucleotide databases

translated into biologically meaningful information
↓
using computer
↓
Bioinformatics

- 1959 - VM Ingram
Compare sickle cell
haemoglobin & normal haemoglobin demonstrates

- 1965 - O. Dayhoff - protein sequences prepared - Atlas of

→ O. Dayhoff - Prepared Computer methods

- ① In comparing protein sequences
- ② detecting distantly related sequences & duplication within sequences
- ③ deducing the evolutionary histories from alignment - protein sequences

Protein
sequence &
structure

↓
1st published

National Biomedical Research
Foundation

- D. J. Lipman - Director of

National Center for Biotechnology - collaborative project produced BLAST

(Basic Local Alignment Search Tool)

- 1980 - DNA Sequence - EMBL

↓
European Molecular Biology Laboratory

- 1984 - NBRF - Protein Information Resource (PIR)

↓
National ~~Institute~~ Biomedical Research Foundation.

- 1988 - NIH (National Institute of Health)

↓
USA developed NCB (National Center for Biotechnology Information)
↓
As ~~the~~ division of National Library of Medicine (NLM)

↓
Information System of molecular biology

- 1988 - DDBJ, EMBL, Gene Bank, of ~~the~~ International Nucleotide Sequence Database Collaboration had a meeting and agreed to use a common format

- All the databases have collaboration with each other. They regularly exchange their data. Now sequence data are accumulating day by day.

Definition of Bioinformatics

- Fred; Tekala - The mathematical, statistical and computing methods that aim to solve biological problem using DNA & amino acid sequences & related information.

includes

- It ~~includes~~ Computational biology (population modeling & numerical simulations)

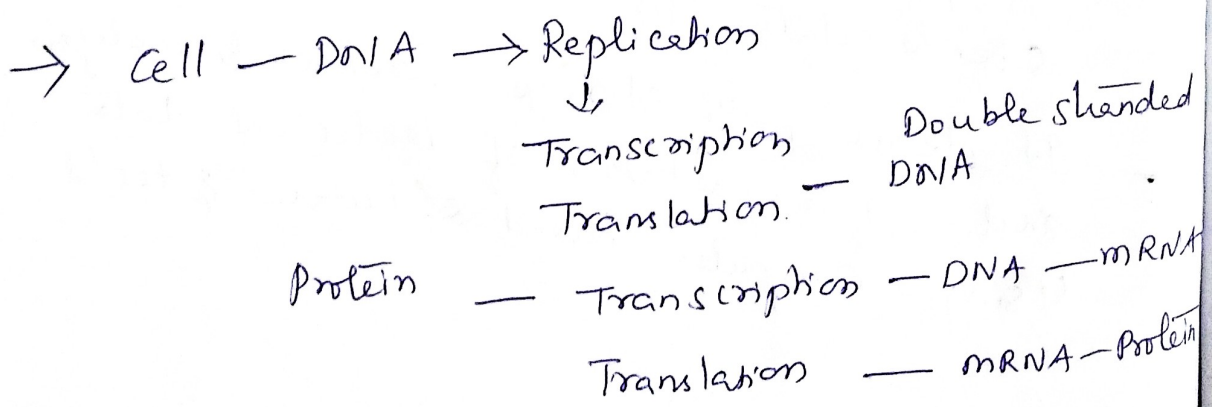
- Bioinformatics is a branch of science that integrates computer science, maths & statistics, chemistry & engineering for analysis, exploration, integration & exploitation of biological science data in Research & Development.

- It deals with storage, retrieval, analysis and interpretation of biological data using computer based software & tools.

objectives

- Development of new algorithms & statistics
 - assessing relationships among large sets of biological data
- Application : Tools for analysis & interpretation of various biological data
- Development of database : storage, access & management of the large body of various biological information.
- Approaches to improving drug discovery ~~and functions in these processes which~~

Central Dogma of molecular Biology



- Protein — sequence (chains of amino acids)
eg:- Lec — Meth — Arg — Try

- Nucleotide sequence

① DNA & RNA

→ Nitrogen bases — A [Adenine]
G [Guanine]
C [Cytosine]
T [Thymine]

- Base pairing — $A \equiv T$
 $G \equiv C$

→ RNA — Instead of T — replaced by Uracil.
Types of RNA — mRNA (messenger RNA)
tRNA (Transfer RNA)
rRNA (ribosomal RNA).

→ Genes — Segment of DNA representing nucleotides.

→ Genome — Total genetic material of an organism.

eg:- Human genome — 3 billion nucleotides

→ Chromatin — DNA + protein — chromatin
(Histone protein)
— Structural unit of chromatin → Nucleosome
— Histone protein & Non-Histone protein
— Nucleosome — solenoids
↓
100,000 bp of DNA.

Applications of Bioinformatics

— Molecular medicine

- ① more drug targets
- ② Personalized medicine
- ③ Preventative medicine
- ④ Gene therapy

— Drug Targets

— Information Search & Retrieval

— Genome Related Applications

— Sequence Comparison

— Linkage Analysis

— Phylogenetic Analysis

— Genomics

— Microarrays

— Sequence Assembly

— Genome Annotation

— Proteomics

— Pharmacogenomics

— Drug-Discovery - Computer-aided Drug Design

— Systems Biology

Applications of Bioinformatics are studied mainly two levels

- ① Basic level, — Bioinformatics is used to organize biological data help researchers access information, add new information arising

out of experiments & modify existing information.

→ 3 datasets $\begin{cases} \text{genome sequences} \\ \text{macromolecular structures} \\ \text{functional genomics expt.} \end{cases}$

→ It is applied in phylogenetic tree analysis, metabolic pathways, medical information, text of scientific papers.

eg:- Protein Data Bank (<http://www.rcsb.org/pdb/>)

(Fast ~~Adaptive~~ Adaptive shrinkage Threshold Algorithm.)
② - Fasta (<http://www.ebi.ac.uk/Fasta3>)

↳ Provide sequence similarity & homology searching against nucleotide & protein databases.

- BLAST (<http://www.ncbi.nlm.nih.gov/BLAST>)
(Basic Local Alignment Search Tool)

— useful to identify distant sequence relationships, statistical interpretation. — Protein & DNA — uses a heuristic algorithm

global alignment

↓
sequences
- Relationships

③ Bioinformatics applications is to use these tools to analyze the data and interpret the result in a Biologically meaningful manner.

① Information Search & Retrieval

— eg:- Pub Med (<http://www.ncbi.nlm.nih.gov/pubmed/>)

↓
National Library of Medicine

↓
provide access to
over 12 million
Medline citations
back to 1960's
+ life science
journals.

ETBLAST (<http://chaos.swmed.edu/etblast/index.shtml>)

↓
① used to compare query set
sentences with database.

② identify the text database
that is most similar to
the query.

Genetics Related Applications

— 3 types of computational problems in
genetics:

① Analysis of single sequence to assess
similarity with known genes

② Identification of typical features
such as binding sites & derive
evolutionary relationships through

phylogenetic trees.

- ③ Complete genome analysis to identify members of gene families, determination the chromosomal location of gene etc

② Sequence Comparison.

- Sequence comparison ~~or~~ similarity search

eg:- Similarity search - ~~the~~ BLAST & FASTA.

- pairwise comparison of sequences

③ Linkage Analysis

- It identify - chromosomal location of genes, disease identification.

eg:- <http://linkage-rockefeller.edu/>
↓
site for search Linkage Analysis.

④ Phylogenetic Analysis

- It identify evolutionary information to phylogenetic trees.

- Information about phylogeny & Biodiversity

eg:- ① (<http://evolution.genetics.washington.edu/phylip.html>)

② (<http://tolweb.org/tree/phylogeny.html>)

⑤ Sequence Assembly

- Human genome project - popularized sequence - Assembly tools

- used to efficiently & accurately assemble large numbers of individual sequence reads.

eg:-1. <http://www-phrap.org/>

Used to assembly & finishing of shotgun sequencing information.

⑦ Microarrays

⑥ Genomics

→ Term coined by Thomas Roderick in 1986.
- It refers to mapping, sequencing and analysis of genomes.

Structural genomics :- Techniques & tools

1. Linkage analysis
2. Molecular cytogenetics
3. Physical mapping
4. EST sequencing (Expressed sequence tags)
5. Genome sequencing
6. Genome organization

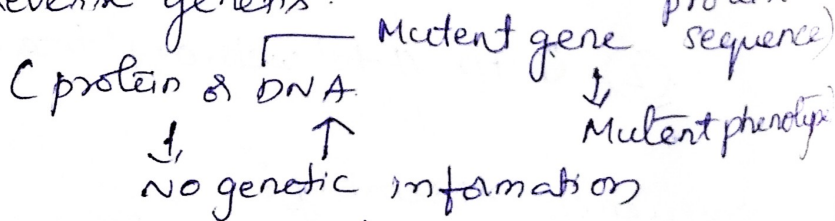
Functional genomics :- To study gene function

Tools :-

① Gene expression

② Forward genetics (gene → DNA

③ Reverse genetics.



④ Comparative genomics

⑤ proteomics

⑥ Metabolomics.