Computation on Bio-Informatics

Shubham Naga[1*], Sapna Choudhary[2]

[1,2]Department of Computer Science and Engineering, Shri Ram Group of Institutions, Jabalpur, India

Corresponding Author: subhamnaga@gmail.com

Abstract— Bioinformatics is an interdisciplinary field mainly involving molecular biology and genetics, computer science, mathematics, and statistics. Data intensive, large-scale biological problems are addressed from a computational point of view. The most common problems are modelling biological processes at the molecular level and making inferences from collected data. As an emerging discipline, it covers a lot of topics from the storage of DNA data and the mathematical modelling of biological sequences, to the analysis of possible mechanisms behind complex human diseases, to the understanding modelling of the evolutionary history of life, etc. It studies the DNA and genetic information.

Keywords—Bio-informatics, Prokaryotic, Eukaryotic, Peptides, genomics, pretomics, HGP,

I. INTRODUCTION

The term bioinformatics was coined by PaulienHogeweg in 1979 for the study of informatics processes in biotic systems.

It's objectives are :
- Enable the discovery of new biological insights.
- Create and maintain databases to store biological information,such as nucleotide and amino acid sequences.
- Provide in-silico capabilities.
- Allow for the integration of molecular biology/ genetics to advance the understanding of complex structures with respect to living cell and the entire organism.
- Major research goals in the field include sequence alignment, gene finding, genome assembly, drug discovery, protein structure alignment, protein structure prediction, prediction of gene expression and protein-protein interactions, genome-wide association studies and the modelling of evolution.
- It focuses on developing and applying computationally intensive techniques (e.g., pattern recognition, data mining, machine learning algorithm, and visualisation) to achieve this goal.
- This paper contains the II. Related work, III. Survey, IV. Conclusion

II. RELATED WORK

Bioinformatics studies the storage, manipulation, and interpretation of biological data, especially data of nucleic acids and amino acids, and studies molecular rules and systems that govern or affect the structure, function, and evolution of various forms of life from computational approaches. The word "computational" does not only mean "with computers," but it refers to data analysis with mathematical, statistical, and algorithmic methods, most of which need to be implemented with computer programs. As computational biology or bioinformatics studies biology with quantitative data, people also call it as quantitative biology.

A. Basic biology

Biology involved in this topic includes cells, nucleic acid, genes and the genomes.No matter what type of bioinformatics one is interested in, basic understanding of existing knowledge of biology especially molecular biology is a must.

a. *Cell*:-It is fundamental structural and functional unit of living organism. There are two types of cells :i) **prokaryotic cells**(represented by bacteria, blue-green algae, mycoplasma, PPLO) which are generally smaller and multiply rapidly.
ii) **eukaryotic cells**(includes all protists,plants,animals,and fungi) which have an extensive compartmentalization of cytoplasm and possess an organised nucleus with a nuclear envelope.

b. *DNA, RNA, and Chromosomes:-* **Deoxy-ribonucleic acid** is helically twisted double chain mixed polymer of deoxy-ribonucleotides which is not only the largest macromolecule but also represents genetic material of organisms. **Ribonucleic acid** was the first genetic material. There is now enough evidence that suggest essential life process revolves around RNA.

Chromosome is the DNA molecule with all the genome in it.

c. *Protein*:-They are **polypeptides**. It is a chain of **amino acid** linked with **peptide bonds**.

d. *Genes and Genome*:-**Genes** are the basic genetic elements that, together with interaction with environment, are decisive for the phenotypes. It is often defined as the segments of DNA by going through the knowledge of central dogma and genetic code. According to studies, we get that there are some genes which don't involve in protein product but play genetic roles
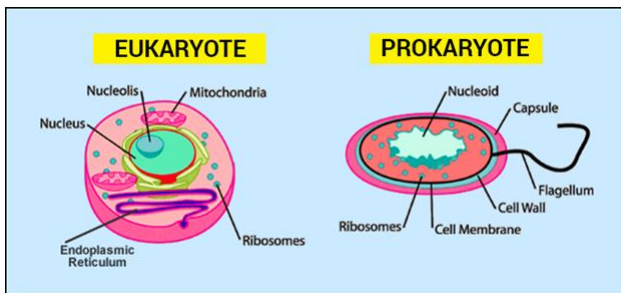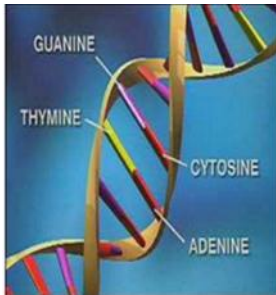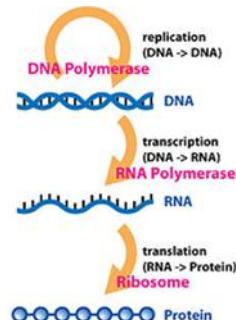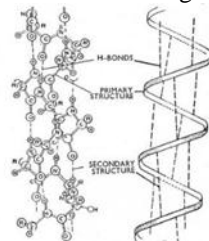
Fig 2. Diagram of Bacteria

Fig 3. D.N.A

Fig 4. Central Dogma

Fig 5. Protein

e. *HGP*:-The **Human Genome Project** was an international scientific research project with the goal of determining the sequence of **nucleotide base** pairs that make up human DNA, and of identifying and mapping all of the genes of the human genome from both a physical and a functional standpoint

B. *Basic Statistics for Bioinformatics*

Many research topics in computational biology and bioinformatics heavily rely on the application of probabilistic models and statistical methods.

a. *Probabilities:-* The conclusion about a population of object is drawn by conducting experiments.

b. *Random Variables*

c. *Distributions*

d. *Random Sampling*

e. *Point Estimation*

f. *Interval Estimation*

C. *Computational Genomics*

This section involves following :

a. ***Genome Informatics*:-**Genome informatics is also data driven; many computational tools developed can soon be obsolete when new technologies and data types become available.

b. ***Finding Protein-Coding Genes:-*** This process is performed by keeping Central dogma in view.It involves: a) to identify a Coding Exon; b)Identifying Promoters; etc.

c. Computational Analysis of ***Transcriptional Genomics*** Data

d. Elements ***modelling and Representation***.

e. Genomic Arrays and ***aCGH/CNP Analysis***

f. Modelling ***Motif Enrichment*** in Sequences

D. *Data Mining in Bioinformatics*

Data mining approaches are ideally suited for bioinformatics. The ongoing influx of these data, the presence of biological signals despite high data noises, and the gap between data collection and knowledge extraction have collectively created new and exciting opportunities for data mining researchers in this field. The extensive availability of open-access biological databases has created both challenges and opportunities for developing novel knowledge discovery and data mining methods specific to molecular biology.

E. *Soft computing in Bioinformatics*

Within **genomics**, soft computing applications in gene expression classification and time series modelling are described. Within **proteomics**, clinical proteomics and protein identification are discussed. All of these areas have seen activity in the application of soft computing techniques to improve the process of performing, and analysing the data from, experiments.

F. *Artificial Intelligence*

Artificial Intelligence deals with creating robots with intelligence just like human brains. In this work they have created **artificial neurons** which are the replica of human neuron. Neuron is the single unit of nervous system, which is the most important system of human body. As we can see that this subject also requires Bioinformatics.

III.   SURVEY

Bioinformatics keeps the biological data. It studies DNA, protein and hereditary information of a family. The new generations inherits the property of previous generations. These **traits** involve *height, colour*, etc. *Colour of human skin, hairs and eye lenses* depends upon the amount of melanin in body.

Blood groups are also determined by the parent's blood group.

Parental blood groups	Child's blood group
O and O	O
O and A	O or A
O and B	O or B
O and AB	A or B
A and A	A or O
A and B	O or A or B or AB
A and AB	A or B or AB
B and B	B or O
B and AB	B or A or AB
AB and AB	A or B or AB

Fig.10 Table of ABO blood group in humans

As all these information of parent is inherited by children, there comes a issue of inheriting diseases and disorders. These are called Genetic diseases and disorders, as they are inherited through genes. Genes are of two types i.e., protein gene and non-protein gene. Gene is fragment of DNA. These genes carry the information from parent's body to children's body. These genes are also responsible for carrying diseases and disorders also from parents to children. There are many such problems nowadays are being noticed. Genetic disorders are caused by one or more changes, or mutations, in the instruction code of a particular gene, preventing the genes from functioning properly.
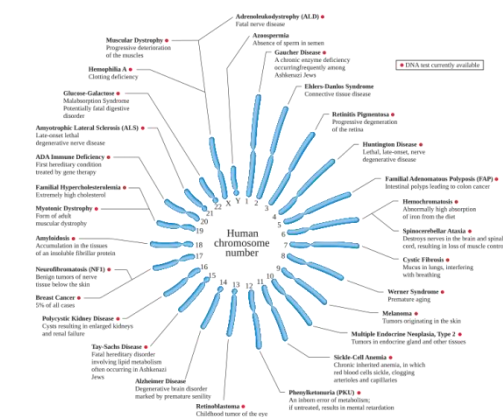

Fig 11. Genetic disorders.

There are following types of genetic disorders:
- X-linked dominant
- X-linked recessive
- Y-linked disorder
- Mitochondrial
- Autosomal dominant
- Autosomal recessive

Use of Bioinformatics in Biotechnology

Biotechnology deals with techniques of using live organisms or enzymes from organisms to produce products and processes useful to humans.

Principles of biotech involve genetic engineering and chemical engineering, maintenance of sterile conditions. Chemistry of DNA and RNA is altered to be introduced into host organism to change phenotype of the host organism.Three basic steps in genetically modifying an organism —

Identification of DNA with desirable genes
- Introduction of the identified DNA into the host
- Maintenance of introduced DNA in the host and transfer of the DNA to its progeny.

Tools of recombinant DNA are Restriction enzymes (often called molecular scissors),cloning vector, competent host, bioreactors.

Applications of Bioinformatics involve genomics, proteomics, cheminformatics and drug design, phylogenetic, crop improvement, biodefence etc.

IV.   RESULTS AND DISCUSSION

As we have gone through the survey of Genetic disorders, we can see that it is required to have some precautions for the future generations of a family.Thus it is necessary to keep a record of all the problems of family. This can give the idea of the possibilities of disorders in future generations of that family. There are some problems such as diabetes and kidney problems which can be prevented.

If we go through record of our family we can take the preventive measures and take the suitable diet and lifestyle. Thus it can help in some cases.

V.   CONCLUSION AND FUTURE SCOPE

By this study we can conclude that Bioinformatics is very important and has a huge scope. It is combination of mathematics, sciences and engineering. It studies the biological information. It deals with molecular biology and genetics also. It is generally said bioinformatics is like keeping computational record of DNA. There are many applications of bioinformatics in biotechnology.

There is also scope of Bioinformatics in Artificial intelligence and soft computing.

Using Bioinformatics and Biotechnology together it is possible to take some measures to control transfer of disorders through genes.

REFERENCES

[1] Y. Lin () Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China e-mail: ylin@math.tsinghua.edu.cn R. Jiang MOE Key Laboratory of Bioinformatics and Bioinformatics Division, TNLIST/Department of Automation, Tsinghua University, Beijing 100084,

[2] M.Q. Zhang () Department of Molecular and Cell Biology, The University of Texas at Dallas, 800 West Campbell Rd, RL11, Richardson, TX, USA 75080 Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China e-mail: mzhang@cshl.edu A.D. Smith Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

[3] Unsupervised learning of multiple motifs in biopolymers using expectation maximization. Mach Learn 21(1–2):51–80 4. Bairoch A (1992) PROSITE: a dictionary of site and patterns in proteins. Nucl Acids Res 20:2013–2018

[4] R. Jiang et al. (eds.), Basics of Bioinformatics: Lecture Notes of the Graduate Summer School on Bioinformatics of China, DOI 10.1007/978-3-642-38951-1 3, © Tsinghua University Press, Beijing and Springer-Verlag Berlin Heidelberg 2013

[5] Zvelebil M, Baum J (2007) Understanding bioinformatics. Garland Science, New York, NY. ISBN 978-0815340249

[6] J. Chen and S. Lonardi, Biological Data Mining. Chapman & Hall/ CRC, 2010

[7] Mitton, Jeffery, B (2002). "Heterozygous Advantage". eLS. doi:10.10.38/npg.els.0001760.

[8] Keane MG; Pyeritz RE (May 2008). "Medical management of Marfan syndrome". Circulation. 117 (21): 2802–13. doi:10.1161/CIRCULATIONAHA.107.693523. PMID 18506019.

[9] Walker FO (2007). "Huntington's disease". Lancet. 369 (9557): 218–28 [221]. doi:10.1016/S0140-6736(07)60111-1. PMID 17240289.

[10] Williams T. N.; Obaro S. K. (2011). "Sickle cell disease and malaria morbidity: a tale with two tails". Trends in Parasitology. 27 (7): 315–320. doi:10.1016/j.pt.2011.02.004.

[11] Kumar A[*] and Chordia N, School of Biotechnology, Devi Ahilya University, Khandwa Rd, Indore, Indore 452001,India,

Tel: +21674241888; Fax: 07312527532; E-mail:ak_sbt@yahoo.com