ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 7 Issue 6 November 2019

International Manuscript ID : 23482001V7I6112019-95

(Approved and Registered with Govt. of India)

Bioinformatics Data Analytics and Research Perspectives

Shahnawaz Ayoub

Research Scholar

Shri Venkateshwara University

Gajraula, U.P., India

Dr. Rakesh Kumar

Associate Professor

Shri Venkateshwara University

Gajraula, U.P., India

Abstract

Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data, in particular when the data sets are large and complex. As an interdisciplinary field of science, bioinformatics combines biology, computer science, information engineering, mathematics and statistics to analyze and interpret the biological data. Bioinformatics has been used for in silico analyses of biological queries using mathematical and statistical techniques. Bioinformatics includes biological studies that use computer programming as part of their methodology, as well as a specific analysis "pipelines" that are repeatedly used, particularly in the field of genomics. Common uses of bioinformatics include the identification of candidates genes and single nucleotide polymorphisms (SNPs). Often, such identification is made with the aim of better understanding the genetic basis of disease, unique adaptations, desirable properties (esp. in agricultural species), or differences between populations.

Keywords: Bioinformatics, Data Analytics, Data Science

ISSN (Online) : 2348 - 2001
International Refereed Journal of Reviews and Research
Volume 7 Issue 6 November 2019
International Manuscript ID : 23482001V7I6112019-95
(Approved and Registered with Govt. of India)

Introduction

Bioinformatics has become an important part of many areas of biology. In experimental molecular biology, bioinformatics techniques such as image and signal processing allow extraction of useful results from large amounts of raw data. In the field of genetics, it aids in sequencing and annotating genomes and their observed mutations. It plays a role in the text mining of biological literature and the development of biological and gene ontologies to organize and query biological data. It also plays a role in the analysis of gene and protein expression and regulation. Bioinformatics tools aid in comparing, analyzing and interpreting of genetic and genomic data and more generally in the understanding of evolutionary aspects of molecular biology. At a more integrative level, it helps analyze and catalogue the biological pathways and networks that are an important part of systems biology. In structural biology, it aids in the simulation and modeling of DNA,[2] RNA,[2][3] proteins[4] as well as biomolecular interactions.[5][6][7]

Historically, the term bioinformatics did not mean what it means today. Paulien Hogeweg and Ben Hesper coined it in 1970 to refer to the study of information processes [8] in biotic systems.[9] This definition placed bioinformatics as a field parallel to biochemistry (the study of chemical processes in biological systems).[9]

Sequences of genetic material are frequently used in bioinformatics and are easier to manage using computers than manually. Computers became essential in molecular biology when protein sequences became available after Frederick Sanger determined the sequence of insulin in the early 1950s. Comparing multiple sequences manually turned out to be impractical. A pioneer in the field was Margaret Oakley Dayhoff.[10] She compiled one of the first protein sequence databases, initially published as books and pioneered methods of sequence alignment and molecular evolution. Another early contributor to bioinformatics was Elvin A. Kabat, who pioneered biological sequence analysis in 1970 with his comprehensive volumes of antibody sequences released with Tai Te Wu between 1980 and 1991. In the 1970's, new techniques for sequencing DNA were applied to bacteriophage MS2 and øX174,

ISSN (Online) : 2348 - 2001
International Refereed Journal of Reviews and Research
Volume 7 Issue 6 November 2019
International Manuscript ID : 23482001V7I6112019-95
(Approved and Registered with Govt. of India)

and the extended nucleotide sequences were then parsed with informational and statistical algorithms. These studies illustrated that well known features, such as the coding segments and the triplet code, are revealed in straightforward statistical analyses and were thus proof of the concept that bioinformatics would be insightful.

To study how normal cellular activities are altered in different disease states, the biological data must be combined to form a comprehensive picture of these activities. Therefore, the field of bioinformatics has evolved such that the most pressing task now involves the analysis and interpretation of various types of data. This includes nucleotide and amino acid sequences, protein domains, and protein structures. The actual process of analyzing and interpreting data is referred to as computational biology. Important sub-disciplines within bioinformatics and computational biology include:

Development and implementation of computer programs that enable efficient access to, management and use of, various types of information. Development of new algorithms (mathematical formulas) and statistical measures that assess relationships among members of large data sets. For example, there are methods to locate a gene within a sequence, to predict protein structure and/or function, and to cluster protein sequences into families of related sequences.

The primary goal of bioinformatics is to increase the understanding of biological processes. What sets it apart from other approaches, however, is its focus on developing and applying computationally intensive techniques to achieve this goal. Examples include: pattern recognition, data mining, machine learning algorithms, and visualization. Major research efforts in the field include sequence alignment, gene finding, genome assembly, drug design, drug discovery, protein structure alignment, protein structure prediction, prediction of gene expression and protein–protein interactions, genome-wide association studies, the modeling of evolution and cell division/mitosis.

ISSN (Online) : 2348 - 2001
International Refereed Journal of Reviews and Research
Volume 7 Issue 6 November 2019
International Manuscript ID : 23482001V7I6112019-95
(Approved and Registered with Govt. of India)

Bioinformatics now entails the creation and advancement of databases, algorithms, computational and statistical techniques, and theory to solve formal and practical problems arising from the management and analysis of biological data.

Over the past few decades, rapid developments in genomic and other molecular research technologies and developments in information technologies have combined to produce a tremendous amount of information related to molecular biology. Bioinformatics is the name given to these mathematical and computing approaches used to glean understanding of biological processes.

Common activities in bioinformatics include mapping and analyzing DNA and protein sequences, aligning DNA and protein sequences to compare them, and creating and viewing 3-D models of protein structures.

Bioinformatics is a science field that is similar to but distinct from biological computation, while it is often considered synonymous to computational biology. Biological computation uses bioengineering and biology to build biological computers, whereas bioinformatics uses computation to better understand biology. Bioinformatics and computational biology involve the analysis of biological data, particularly DNA, RNA, and protein sequences. The field of bioinformatics experienced explosive growth starting in the mid-1990s, driven largely by the Human Genome Project and by rapid advances in DNA sequencing technology.

Analyzing biological data to produce meaningful information involves writing and running software programs that use algorithms from graph theory, artificial intelligence, soft computing, data mining, image processing, and computer simulation. The algorithms in turn depend on theoretical foundations such as discrete mathematics, control theory, system theory, information theory, and statistics.

ISSN (Online) : 2348 - 2001
International Refereed Journal of Reviews and Research
Volume 7 Issue 6 November 2019
International Manuscript ID : 23482001V7I6112019-95
(Approved and Registered with Govt. of India)

Since the Phage Φ-X174 was sequenced in 1977,[19] the DNA sequences of thousands of organisms have been decoded and stored in databases. This sequence information is analyzed to determine genes that encode proteins, RNA genes, regulatory sequences, structural motifs, and repetitive sequences. A comparison of genes within a species or between different species can show similarities between protein functions, or relations between species (the use of molecular systematics to construct phylogenetic trees). With the growing amount of data, it long ago became impractical to analyze DNA sequences manually. Computer programs such as BLAST are used routinely to search sequences—as of 2008, from more than 260,000 organisms, containing over 190 billion nucleotides.

Conclusion

Computational technologies are used to accelerate or fully automate the processing, quantification and analysis of large amounts of high-information-content biomedical imagery. Modern image analysis systems augment an observer's ability to make measurements from a large or complex set of images, by improving accuracy, objectivity, or speed. A fully developed analysis system may completely replace the observer. Although these systems are not unique to biomedical imagery, biomedical imaging is becoming more important for both diagnostics and research

References

[1] Henry Marshall Leicester; Herbert S. Klickstein (1951). A Source Book in Chemistry, 1400-1900. Harvard University Press. p. 309.

[2] Kiefer, D. M. (1993). "Organic Chemicals' Mauve Beginning". Chem. Eng. News. 71 (32): 22–23. doi:10.1021/cen-v071n032.p022.

[3] "August Kekulé and Archibald Scott Couper". Science History Institute. June 2016. Retrieved 20 March 2018.

[4] Streitwieser, Andrew; Heathcock, Clayton H.; Kosower, Edward M. (2017). Introduction to Organic Chemistry. New Delhipages=3–4: Medtech (Scientific

International, reprint of revised 4th edition, Macmillan, 1998). ISBN 978-93-85998-89-8.

[5] Roberts, Laura (7 December 2010) History of Aspirin. The Telegraph

[6] Bosch F & Rosich L (2008). "The contributions of Paul Ehrlich to pharmacology: A tribute on the occasion of the centenary of his Nobel Prize". Pharmacology. 82 (3): 171–9. doi:10.1159/000149583. PMC 2790789. PMID 18679046.

[7] "Paul Ehrlich, the Rockefeller Institute, and the first targeted chemotherapy". Rockefeller University. Retrieved 3 Aug 2012.

[8] "Paul Ehrlich". Science History Institute. June 2016. Retrieved 20 March 2018.

[9] Torker, Sebastian; MüLler, Andre; Sigrist, Raphael; Chen, Peter (2010). "Tuning the Steric Properties of a Metathesis Catalyst for Copolymerization of Norbornene and Cyclooctene toward Complete Alternation". Organometallics. 29 (12): 2735–2751. doi:10.1021/om100185g.

[10] Steingruber, Elmar (2004) "Indigo and Indigo Colorants" in Ullmann's Encyclopedia of Industrial Chemistry, Wiley-VCH, Weinheim. doi: 10.1002/14356007.a14_149.pub2