BIOMIRROR

An Open Access Journal

Bioinformatics as an Emerging Tool for Biological and Medical Fields

[a] Reeti Chaudhary*,

[b] Umaid Singh,

[a] Pragati Sharma,

[a] Sonia Takshak

[a] Department of Biotechnology, Deenbandhu Chhotu Ram University of Science & Technology, Murthal, Haryana, INDIA.

[b] Retired Professor, Hisar , Haryana, INDIA.

ARTICLE INFO

ABSTRACT

In the present scenario, bioinformatics has become an approach of choice for forensic sciences, disease risk assessment and diagnosis and many other issues owing to the growing network of biological information databases covering human genomes, transcriptomics and proteomics. This will also help in gaining a better understanding of beneficial and harmful human genes and also of the genetics of responses to therapy and customized treatment including identification of drug targets and gene therapy. In this review, efforts have been made to discuss the above areas with a greater emphasis on the emerging areas of bioinformatics for the benefits of students, teachers and researchers engaged in various fields of biological sciences and human health care system.

Introduction:

In the developed countries, the technological innovations have progressed at a highly accelerated pace and this has made tremendous impact on almost every facet of our lives [1]. Biological data are being produced at phenomenal rate and storing, analyzing and interpreting such data in a meaningful way is assuming greater significance. Bioinformatics is an interdisciplinary subject which is primarily aimed at using information to understand biologic. The areas of medical fields and deliveries of health care services have received increasing emphasis in recent years. The health care systems capable of providing a wide range of effective diagnostic and therapeutic treatments are becoming increasingly important in present day world. Of particular importance in this evolutionary process has been understanding and treatment of human genetic disorders [2, 3].

Corresponding author: Reeti Chaudhary*

E-mail address: reeti.malik@gmail.com

Further, the applications of genomic technologies require an understanding of three key sets of concepts. How genetic information is stored, how is it processed and how is it transmitted from parents to offsprings [3].

The engineering professionals and medical care experts are intimately involved in many medical ventures. As a result of this, the discipline of biomedical engineering has emerged as an integrated medium for two dynamic professions, i.e. medicine and engineering [5]. This approach has played an important role in the struggle against diseases of numerous types by providing tools such as biosensors, bioinformatics, image processing and artificial intelligence which could be utilized for research, diagnosis and treatments by health care professionals. Main obstacles in the treatment of various diseases are the lack of diagnosis, therapies and prognosis. Bioinformatics is a critically important part of systems biology [6, 7]. In this context, networking and bioinformatics providing a wide range of effective diagnostic and preventive approaches for human health could play a very vital role. In view of these observations, efforts have been made to present and discuss the following relevant areas of bioinformatics in this paper.

Network Basics and Bioinformatics:

The networks are extremely important in collection and compilation of data of biological and medical significance. A network is an easy concept to grasp. Typically, it is the interconnection of several elements, known as nodes, to provide some sort of added benefit. Networks are categorized in many ways depending on the size of the network, its purpose, the underlying technology and its physical configuration. Each network is interconnect--ed to facilitate access to databases. Data networking is a relatively new technology which has been transforming the way we obtain information, organize our work, manage and plan future strategic approaches for accomplishing the desired results. For example, the internet is a huge global network of data for various aspects of bioinformatics. Now,

networking scientists are transforming/transmitting a billion bits per second through one strand of fiber by using light waves [8].

A network has been established in recent years to link various laboratories which used biocomputing and bioinformatics in molecular biology research [9]. The network, known as EMB net was developed to provide information, services and training to users in various laboratories engaged in research and development activities in biological fields.

Now the technologies are available which could be used to encode, transmit, and then decode the data on its way through the network. Further the networks are smart enough to convert the protocols as data passes from the edge of one network to another. Millions of people worldwide are going online every year and this phenomenon is growing incredibly fast. Networks give us tools we did not have before. Now, we can teleconference with a geographically dispersed group of people. With help of networks, we can collaborate and conduct research with people in different continents and could have access to information and support services that would be impossible otherwise. For example, there is a shortage of diagnosticians in the medical field and circumstances make them to share critical resources. Medical alliances can now be formed by interconnecting through the newest networking technologies. Retrieving and distributing information is critical to most health care systems. At present times, network and telecommunications infrastructure support a broad range of applications in this field. New technologies, such as the World Wide Web (www) allow the doctors and medical experts to reach and interact with patients and other resource persons in entirely new ways.

The global networks of biological information and service providers have made the bioinformatics as a revolutionary science [10]. Internet, considered as a communication revolution, is a global network of computers, which link academic, research and education institutions. This allows computers to talk to each other in their own electronic languages. Biological information is stored on many different computers around the world. The easiest way to access this information is to join all those computers in a network. The concept of information sharing between remote locations and the ramification of rapid data dissemination and communication has found immediate applications in bioinformatics. In the present times, www the most advanced information system deployed on the internet and making a profound impact in the field of bioinformatics. Now, the computer programs are available such as bioinformatics sequence markup language (BSML) which graphically describes genetic sequences and methods for storing and transmitting encoded sequence and graphic information through networking. Biopolymer markup language (BIOML) is data type information for the annotation of molecular biopolymer sequence information and structure data.

Databases and Bioinformatics:

Generally, biological data are gathered and stored all over the world and bioinformatics play a big role in the interpretation of these data in a biologically meaningful way[1]. Databases and programs allow us to access the existing information and to compare these data to find similarities and differences. Data bases can be searched or cross referenced either on the internet or using downloaded versions on local computers or computer networks by multiple users. Bioinformatics helps in obtaining information available in these data bases which can be searched, compared, retrieved and analyzed. Most popular statistical methods used for gene finding are hidden Markov models (HMMs). Much work in genomic science and bioinformatics focuses on problems of identifying biologically important regions of very long DNA sequences such as chromosomes or genomes. Hidden Markov models (HMMs) are a class of mathematical tools that excel at identifying this type of regions [11]. A large amount of biological information is available all over the web but in a scattered form which is difficult to manage. Such data need to be organized in form of databases for easy usage and retrieval. Bioinformatics is helping researchers and scientist by providing efficient retrieval and analysis methods [12].

A database in itself, containing nucleic acid sequences or protein sequences, not only stores information but also provides measures for collecting information and analytical processing such as sequence retrieval from the database, comparison of two or more sequences, translating nucleic acid sequences into protein sequences or vice versa and even prediction of protein structures based on similar sequences that pre-exist and even, molecular graphics. This type of information is extremely important in understanding the direction and type of newer protein synthesis. Translation of DNA sequences to protein sequences is very vital in medical and biological fields.

There are different types of databases, depending both on the nature of the information being stored and on the manner of data storage [8]. From bioinformatics point of view, data bases are broadly classified into two types, namely, generalized databases and specialized databases. Examples of generalized databases are DNA, proteins, carbohydrates or similar databases. Examples of specialized databases are expressed sequence tags (EST), genome survey sequences (GSS), single nucleotide polymorphism (SNP), sequence tagged sites (STS), or similar databases. These mainly comprise of sequences and structures. While sequence databases contain primary sequences of nucleic acids or amino acids, structure databases are fed with records of biochemically solved structures of macromolecules (eg. protein three dimensional (3D) structure). Essentially, bioinformatics has three components : i) designing databases that allow storage and management of large biological data, ii) statistical analysis for determining relationship among large data sets, and iii) the use of tools for the analysis and interpretation of biological data including sequences, structures, gene expressions and metabolic pathways.

Genome, DNA Sequencing and Bioinformatics:

Genes are stored in the form of DNA molecules in a genome which provide the basic units for heredity, i.e. transmission of traits from one generation to subsequent generation. Primary function of genome sequencing is to generate high resolution genetic and physical maps of the genome to define segments of increasing resolution and to sequence the segments in an orderly manner [13]. In addition, an approach, which is referred on the direct shotgun approach, is also followed for genome sequencing [14]. This approach is aimed at breading up the genome into random, overlapping fragment and then to sequence the fragments and assemble the sequences using computer algorithms. Analysis

of genomic sequences reveals that each organism has an array of genes required for basic metabolic processes and genes whose products determine the specialized function of the organism [15]. Complete genome sequencing provides a knowledge base on which to build information about gene and protein expression [9], but is not always sufficient on its own to define the entire protein components of the organism.

Bioinformatics essentially is the compilation of mathematical, statistical and computational analysis of genomic and proteomic data at one destination. As the technology for collecting genomic data has improved, so has the need for using novel methods for management and analysis of the massive amount of accumulated data[1]. According to this report, the topics of work in bioinformatics range from database design to systems engineering to artificial intelligence to applied mathematics and statistics, all with an underlying focus on genome science relating to human health and diseases. A gene sequence, whether predicted or experimentally determined must have its function predicted and many bioinformatics tools are available for this purpose. For example polymorphism data are used to search for correlations with the presence or absence of a disease in family pedigrees. Such approaches are all of fundamental importance and draw attention of scientists and biomedical engineers. Therefore, bioinformatics is becoming a highly multidisciplinary field for collaborations among researchers in many fields around the globe. This would lay a greater emphasis on the importance of networking and its methodologies. The term "clinical bioinformatics" combines clinical informatics, bioinformatics and medical informatics along with information technology, mathematics and omics [6, 16].

Bioinformatics model of gene expression are used as a blue print to produce a specific protein. Pattern in which a gene is expressed provide clues to its biological role. Gene expression controls all the functions of a living system. Comparative analysis of differential gene expression in healthy and diseased tissue can help identify which protein among the normal proteome is responsible for causing abnormalities relating the disease. This information will, in turn, help in the development of new diagnostic test for various diseases as well as new drugs to alter the activity of the affected genes or protein.

Recent advances in the field of biology and chemistry have facilitated an increase in the speed of sequencing genes and proteins. With the help of bioinformatics and models developed by its manifestation, it has become easier to insert foreign DNA sequences into many systems [10]. Rapid mass production of particular DNA sequences has also become possible through such advancement. Further, biocomputing of the information on DNA and protein sequencing has found its application in many areas [17].

Proteomics and Bioinformatics:

Proteomics is the cataloging and analysis of proteins to determine when a protein is expressed, how much is made and with what other proteins it can interact. The term proteomics indicates proteins expressed by a genome. It is the systematic analysis of protein profile of the tissue. Proteomics includes expression proteomics, i.e. the study of overall changes in protein expression and cell-map

proteomics i.e. the study of protein – protein interactions through the isolation of protein complexes. In a way, it identifies and determines the sequence of each protein and entering the data into database. Bioinformatics has specific role in such data bases. The vast majority of genes encoded for proteins [18]. For example, every human cell has a gene that encodes the protein insulin. Bioinformatics may be helpful in maintaining data for such proteins of biological significance.

Apart from providing the theoretical background and practical tools for scientist to explore proteins and DNA, it also helps in many other ways. For example pattern recognition techniques are used to direct similarity between sequences and hence to infer related structures and functions [19]. Also the prediction methods or models are used to deduce 3 D structure of protein and ultimately to infer their functions directly from the linear sequence of amino acids [20]. The direct prediction of protein three dimensional structures from the linear amino acid sequences / analogs is of the important objectives of bioinformatics.

One of the driving forces behind bioinformatics is the search of similarities between different biomolecules [21]. Apart from enabling systematic organization of data, identification of protein homologs has some direct practical uses. The oretila models of proteins are usually based on experimentally solved structures of close homologs. Wherever biochemical or structural data are lacking, study could be carried out in lower organisms like yeast and with the help of bioinformatics the results can be applied to homologs in higher organisms such as humans. It also simplifies the problems of understanding complex genomes by analyzing simple organism first and then applying the same principles to more complicated ones.

The prediction methods are becoming important tools of bioinformatics in protein sequencing [22]. These methods use parameters derived form 3D structures of protein and are also based on physiochemical criteria e.g. fold compactness, hydrophobicity, change of protein molecules, hydrogen bonding potential etc.

Health Benefits of Bioinformatics:

The biological function of the genes has a dramatic impact on human health. The complex of the problems involves the study of the biological systems to understand or characterize how various biological mechanisms function to maintain life in normal human and diseased conditions. The entire human genome sequence is now available and the locations of tens of thousands of human genes are known [12]. Once a genome is sequenced and its genes are found or predicted, the next step in the bioinformatics pipline is to determine the biological function of the genes. Comparative genomic approaches to assigning function to particular genes rely on bioinformatics data bases. Simple database searches are the most straightforward comparative genomic approach employed by the bioinformatics experts in recent years [12]. By employing whole genome sequencing of affected individuals, all genetics variants linked to disease can be identified [14]. This may lead to a broader understanding of disease mechanisms, development of new diagnostic tests and the discovery of new drug targets and design of drugs. With the help of bioinformatics researchers, it is possible to find variations in DNA of a large number of population of a particular group. Once matched with DNA of general

population, it may be easy to know if the changes in DNA of that particular group are associated with a hereditary disease of that group of population. This type of finding may further lead to developing advanced diagnostic tools and novel drugs to combat diseases among the fast depleting communities in the world [23]. Efforts have been made to clarify all known proteins into functional groups using comparative genomics. Bioinformatics play a key role in developing a model of cellular processes which control the metabolic activities of human biological systems.

Metabolic pathways within the cell provide the ability to convert energy within various molecules such as glucose, ketoacids and certain aminoacids into energy needed by the cell to perform various maintenance functions, all dimension or the functions required by the tissue to maintain the organism as a whole. The energy within those molecules is stored in chemical bonds between specific atoms within the molecules of interest. Bioinformatics is helpful in proposing a model to simply understand the processes and also to help identify key control points within a process. For example, for a drug delivery/discovery process, bioinformatics model may prove to be very helpful. The adoption of a bioinformatics based approach to drug discovery and delivery provides an important advantage. With the help of bioinformatics, genotypes or species associated with patho – physiological conditions could be defined which might lead to the identification of potential molecular targets. By using the sequence research techniques, it is possible to model the structure of the specific proteins on experimentally characterized structures.

Several applications in medical sciences have centered on gene expression analysis. Compiling expression data for cells affected by different diseases and comparing the measurements against normal expression levels is the major focus in this context. Identification of gens that are expressed differently in affected cells provides a basis for explaining the causes of illness and highlights potential drug targets. Bioinformatics with greater emphasis on human genome studies, will benefit forensic sciences, pharma industries, discovery of beneficial and harmful genes, and help contribute to a better understanding of human evolution, diagnosis of disease and disease risks, genetics of response to therapy and customized treatment, identification of drug targets and gene therapy.

The health care systems which seek new solutions for the difficult problems encountered dealing with modern societies, particularly for their desire to live longer and healthier life. At present, some genome projects are underway and these projects aim to determine the genetic basis of longevity and age related disorders by highlighting the basic role of bioinformatics. Many of the problems confronting health professionals involve the fundamental aspects of devices, system analyses, design and practical application and all these lie at the heart of bioengineering practices. Various concepts of engineering technologies, information technology and bioinformatics need to be applied together to solve health related issues. Hence, health care related fields such as biomedical engineering, bioengineering, bioinformatics and clinical engineering are getting closely meshed up to resolve basic research and service oriented problems and for science and knowledge development. Bioinformatics may have its origin from bioengineering

which often related to biotechnology and genetic engineering. The bioinformatics may specifically include the pursuits of specific areas of the following topics:

- Basic principle of molecular biology regarding genome and health science.
- Major types of data involved in genome projects including technologies for collections of such data sets.
- Practical applications and uses of genomic data
- Understanding of the major topics in the field of genetic material (DNA) sequence analysis.
- Use of key databases and web resources.

The understanding and treatment of human genetic disorders is the landmark of modern human health science. Obtaining information on the human genome which is associated with numerous human genetic diseases is the major pursuit of bioinformatics. Continued advances in technology are necessary to accelerate the pace together systematic data set to make breakthroughs in basic studies on human genome related to specific genetic diseases [15]. Bioinformatics play a critical role in developing era of personalized medicine [24, 25]. Improved data acquisitions, computational and statistical methods for interpretation of data form the integral part of bioinformatics. The creation of specialized computing hardware and software systems to circumvent massive computational barriers experienced by genome scientist is needed. This is possible by advances in bioinformatics research and development activities.

Conclusion:

The contributions from the field of biological and medical sciences have facilitated a tremendous increase in the data on various aspects as highlighted above in the text. The links between various databases of biological and medical significance are important and bioinformatics plays a very vital role in this direction. The networks and bioinformatics have helped the students, teachers, scientists and researcher worldwide for developing new information technologies for better understanding of the molecular and genetic processes that underline health and diseases. Their specific aims include the creation of automated systems for storing and analyzing biological information, the development of advanced methods of computer-based information processing. In particular, networking in the areas of bioinformatics will help facilitation of user access to databases and software and the coordination of efforts to gather valuable information worldwide. Increasing coordination between networks and bioinformatics will be beneficial particularly for the developing countries which are often constrained with lack of information and limited resources. The results of genomic research will bring a revolution into the field of medicine. Diagnostic procedures will change rapidly as a result of genomic status which may focus on relationship between genotypes and complex phenotypes with the help of bioinformatics. The next generation of biologists will need to be as comfortable at a computer work station as they are at the laboratory benches and this will be a step forward in the field of bioinformatics.

References:

1) Ignacimuthu S (2005) Basic Bioinformatics, Narosa Publ. Pvt. Ltd., New Delhi.

2) Scherf U (2000) A gene expression database for the molecular pharmacology of cancer. Nat Genet 24: 236-244.

3) Teufel A, Krupp M, Weinmann A and Galle PR (2006) Current bioinformatics tools in genomic biomedical research. Int J Mol Med 17: 967-973.

4) Mount DW (2001) Bioinformatic : Sequence and Genome Analysis. Cold Spring Harbor Lab Press Cold Spring Harbor, New York, USA.

5) Enderle J, Blanchard S and Bronzino J (2005) An introduction to Biomedical Engineering. 2nd edition, Academic press, San Diego, USA.

6) Wu D, Rice CM and Wang X (2012) Cancer bioinformatics: A new approach to systems clinical medicine. BMC Bioinformatics 13: 71.

7) Hood L and Tian Q (2012) Systems Approaches to Biology and Disease Enable Translational Systems Medicine. Genomics Proteomics Bioinformatics 10: 181–185.

8) Patten K (2002) Data networking made easy. Firewall Media, New Delhi.

9) Kerr MK, Martin M and Churchill G (2000) Analysis of variance for gene expression in microarray data. J Computer Biology 7: 819-837.

10) Wheeler DL (2000) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 28: 10-14.

11) Sonnhammer E, Eddy S, Birney E, Bateman A and Durbun R (1998) Multiple sequence alignments and HMM Profiles of protein domains. Nucleic Acids Res 26: 320 – 322.

12) Spencer M (2005) Genomics and bioinformatics. Introduction to Biomedical Engg, Academic Press serve (ed John Enderle), San Diego, California, USA.

13) Gibson G and Muse SV (2002) A Primer of Genome Science, Sinauer, Associates, Sunderland, MA, USA.

14) Venter JC (2001) The sequence of human genome. Science 291: 1304 – 1351.

15) IHGSC (2001) International Human Genome Sequencing Consortium, Initial genome sequencing and analysis of the human genome. Nature 409: 860-921.

16) Buriani A, Garcia-Bermejo ML, Bosisio E, Xu Q, Li H, Dong X, Simmonds MS, Carrara M, Tejedor N, Lucio-Cazana J and Hylands PJ (2012) Omic techniques in systems biology approaches to traditional Chinese medicine research: Present and future. J Ethnopharma 140: 535– 544.

17) Waterman MS (1995) Introduction to computational biology, Chapman and Hall, London.

18) Atschul S, Gish W, Miller W, Myers E and Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.

19) Taft RJ and Mattick JS (2003) Increasing biological complexity is positively correlated with relative genome-wide expansion of non protein coding DNA sequences. Genome Biol 5: 1-13.

20) Needleman S and Wunch C (1970) A general method applicable to the search for similarities in the amino acid sequences of two proteins. J Mol Biol 48: 443 – 453.

21) Smith T and Waterman M (1981) Identification of common molecular subsequences. J Mol Biol 147: 195-197.

22) Bernardi G (2000) Isochores and the evolutionary genomics of vertebrates. Gene 241: 3-17.

23) Zhang L (1997) Gene expression profiles in normal and cancer cells. Science 276: 1268-1273.

24) Hong H, Zhang W, Shen J, Su Z, Ning B, Han T, Perkins R, Shi L and Tong W (2013) Critical role of bioinformatics in translating huge amounts of next-generation sequencing data into personalized medicine. Sci China Life Sci 56: 110–118.

25) Yan Q (2013) Translational Bioinformatics Support for Personalized and Systems Medicine: Tasks and Challenges. Transl Med 3:2.