



A brief study on role of current Bioinformatics in Biological data analysis

M Jyothi² and Suneetha V¹

VIT University, Vellore, Tamilnadu, India -632014.

Abstract: The exponential growth in computational speed and storage capabilities has opened a new era in biological analysis of data. Bioinformatics with its continually evolving tools have become indispensable in biological research. Bioinformatics changes large amounts of existing data into information that can be used to facilitate new discoveries in biology. At present research in bioinformatics can be classified into (1) genome studies, (2) proteomics, (3) cell visualization and simulation and (4) application to the development of drugs. Current important approaches in bioinformatics research are (1) analysis based on wet-lab data, (2) finding new results from mathematical modeling, (3) an integrated approach that integrates mathematical modeling with search techniques. The current bioinformatic tools have a potential to develop vaccine and drugs, to know the cause of diseases at molecular level, to develop low cost biological agents for bioremediation. Here are the current roles of bioinformatics, on the other hand bioinformatics software tools and resources are reviewed.

Keywords: Bioinformatics, Genome sequencing, Proteomics, Software tools, Bioremediation.

Introduction:

Bioinformatics is an interdisciplinary science that develops methods and software tools to study and process biological data. Bioinformatics gives low cost data analysis. This resulted in the increase of discoveries, the vaccine and drug design and the design of antimicrobial agents. Add on to this bioinformatics analysis has enhanced our understanding of genome structure and microbial restructuring. Applications of bioinformatics is useful in the areas of molecular biology, genomics, proteomics and in the identification of new molecular targets for drug discovery. In systems biology it helps to analyze biological pathways and networks. In structural biology it helps in the modeling of DNA, RNA and protein structures and their interactions.

Materials and Methods:

Genome sequencing and analysis:

Bioinformatics contribution in genome sequencing has been in the (1) integration of PCR (polymerase chain reaction) or BAC (bacterial artificial chromosome) based amplification with automated sequencing methods, (2) joining the smaller fragments of sequences together to form a complete genome sequence, (3) in protein coding region prediction. Number of organisms are taken and their DNA is decoded and this information is stored in the large genomic repositories like the database of the European Molecular Biology Laboratory, the national centre for biotechnology information database, Institute of bioinformatics India. These institutes help in the study and analysis of expressed sequence tags (ESTs) for gene expression and human cancer genomes are studied to identify disease linked mutations. Genome assembly algorithms are important in the analysis of data

in bioinformatics. Important aspect in sequence analysis of bioinformatics is annotation, this helps to find different genes by computational gene finding.

Automated identification of genes :

Annotation is automated process of marking genes and biological features in a DNA sequence, made possible because of the recognizable start and stop regions of the gene. Recognizing protein coding regions or ORFs (open reading frames) in the genomes are done by various techniques. The recognition of ORFs is done in following ways (1) using techniques such as GLIMMER and GeneMark (2) by searching gene databases like genbank (3) by using algorithms. After finding ORFs the next step is to annotate the gene with proper structure and function. By using sequence search and pair wise gene alignment techniques function of gene has been identified. The most popular algorithms used to annotate the function of gene are (1)BLAST (2) dynamic programming technique smith-waterman alignment and its variations (3) index based scheme FASTA and its variations(4) BLOCKS.

Evolutionary bioinformatics:

It deals with the study of origin and evolution of organisms by measuring DNA changes. Bioinformatics scientists compared extensively multiple genomes to study evolution. Research has established that overall evolution is a combination of point based mutation giving rise to speciation.

Genome comparison:

It helps to find homology between genes in different organisms. The study of comparison of genomes showed that genome restructuring occurs by combination of insertion, deletion, duplication, fusion of domains as well as genes.

Genetics of disease:

It helps to map genes of complex diseases like cancer, parkinsons. Whole genome studies are essential to pinpoint mutations for such diseases.

Analysis of mutations in cancer:

In cancer the genomes of affected cells are not arranged in predictable ways. point mutations are identified by genome sequencing techniques and analysis. Bioinformatics tools continually evolving producing specialized automated systems to manage sequenced raw-data.

Gene and protein expression:

Bioinformatics is very much involved in making sense of microarrays and high throughput mass spectrometry data. Such studies are often used to determine the genes involved in disorder.

3D structure modeling and docking :

One of the important applications in bioinformatics is protein structure prediction. In bioinformatics homology studies is used to predict gene functions. In structural bioinformatics homology is used to determine which part of protein are important in structure formation and interaction with other proteins. There are two major approaches for 3D modeling of structure of proteins(1)sequence homology based prediction (2) ab initio method. Docking is to identify best matches between 3D structures of two molecules

Network and systems biology:

Network analyses helps to understand relationships between biological networks. Network biology often tries to integrate many data types.

Systems biology involves the use of computer simulations of cellular subsystems like enzymes and signal transduction pathways to both analyze and visualize the complex connections of cellular subsystems.

Molecular interaction networks:

Interactions between proteins are analyzed and visualized using networks. Movement of atoms about rotatable bonds is fundamental principle behind computational algorithms for studying molecular interactions

Databases:

For bioinformatics research databases are essential. All kinds of data can be stored in databases .some of the databases are breast cancer database, human protein reference database etc

Software tools:

Biology software from internet falls into four groups: (1) resource sites (archives, bioinformatics service organisations), (2) web lists and catalogues, (3) discussions groups and (4) publications. There are many open source software tools to conduct research work in silico which has greatly contributed to the field of bioinformatics. The open source software tools used in bioinformatics are BioJava, Galaxy, biopython, Anduril,bioperl, bioconductor, AMPHORA, bioruby,BioPHP, bioJS, bioclipse, BioMOBY, Apache Taverna, UGENE, .NET Bio, EMBOSS.

Conclusion:

Despite being a young field bioinformatics has immensely helped biological research through the development of algorithms, techniques and tools. bioinformatics has aided in genome sequencing, in gene location, in finding transcription factor binding sites of the genes. Bioinformatics is used in new drug discovery, microbialde halogenation of pollutants , relation of microRNAs with human diseases. Human allergenicity to novel proteins, bioinformatics will provide details of the molecular basis of human health, the impact of bioinformatics on the foods is going to be at a profound level, bioinformatics helps to predict molecular specificities of the immune system, bioinformatics can give new insights into molecular mechanisms , personalized treatments for cancers, bioinformatics studies help to find regulatory regions in genomic DNA, bioinformatics aids in the translational studies of molecular medicine. The major contribution of the bioinformatics has been in automation of genome sequencing, development of integrated databases and analysis of genome to understand gene and genome function. Thus bioinformatics have myriad of applications in biology.

References:

1. Venter JC, *et al.*: The sequence of the human genome. *Science* 2001, 291:1304-1351. PubMed Abstract | Publisher Full Text
2. Zagursky RJ, Olmsted SB, Russell DP, Wooters JL: Bioinformatics: how it is being used to identify bacterial vaccine candidates. *Vaccines* 2003, 2(3):417-436.
3. Pearson WR, Lipman DJ: Improved tools for biological sequence comparison. *Proceedings National Academy of Science U S A* 1988, 85(8):2444-2448.
4. Don Gilbert :Bioinformatics software resources Biology Department, Indiana UniversityBloomington, Indiana 47405 USA Tel: +1 812 855 0587 E-mail: gilbertd@indiana.edu
5. Arvind K Bansalarvind@cs.kent.edu Bioinformatics in microbial biotechnology, *Microbial Cell Factories* 2005, 4:19 doi:10.1186/1475-2859-4-19
6. Bansal AK, Meyer TE: Evolutionary analysis by whole genome comparisons. *Journal of Bacteriology* 2002, 184(8):2260-2272.
7. Butcher EC, Berg EL, Kunkel EJ: Systems biology in drug discovery. *Nature Biotechnology* 2004, 22(10):1253-1259.
