International Journal of Advanced Scientific Research
ISSN: 2456-0421
Impact Factor: RJIF 5.32
www.allscientificjournal.com
Volume 3; Issue 6; November 2018; Page No. 29-33

Bioinformatics challenges: A review

Manisha Mathur
Research fellow, Advanced Milk Testing Research Laboratory, Post Graduate Institute of Veterinary Education & Research, Rajasthan University of Veterinary and Animal Science, Jaipur, Rajasthan, India

Abstract
Recent endeavours in multidisciplinary studies untangle a human search for integrity and totality field that is Bioinformatics leading to deep understanding of the biological processes at the molecular level. Bioinformatics are developing tools to solve the age - old mysteries of biology like embryogenesis, morphogenesis, development, nerve function, behaviour are determined by a simple gene sequence or genes are still unresolved problems to Bioinformaticians. Research in Bioinformatics and computational biology can include anything from abstraction of the properties of a biological system into a mathematical or physical model, to implementation of new algorithms for data analysis to the development of databases and web tools to access them. Bioinformaticians needs to create novel and enhanced algorithms for data mining, analysis, comparisons, etc.

Keywords: bioinformatics, genomic sequences, data mining, drug discovery

Introduction

Scientists have always been on a journey to unfold the mysteries of the science, out ward and inward. It is an avid thirst and the journey has been ever science the development of human beings on this planet. From the invention of microscope to the latest developments in Life science like cloning, stem cell research, artificial intelligence and the human genome project-all these have been different attempts to understand the dimensions of Life science. Recent endeavours in multidisciplinary studies untangle a human search for integrity and totality field that is Bioinformatics leading to deep understanding of the biological processes at the molecular level. Bioinformatics are developing tools to solve the age - old mysteries of biology like embryogenesis, morphogenesis, development, nerve function, behaviour are determined by a simple gene sequence or genes are still unresolved problems to Bioinformaticians. Companies in the business of developing drugs, agricultural chemicals and hybrid plants, are developing Bioinformatics divisions and looking to bioinformatics to provide new target thus to help replace limited natural resources. Scientist working in the area of Bioinformatics has developed new techniques to analyze DNA sequences on an industrial scale in a new area of science known as Genomics. This shift from gene biology has resulted in the development of strategies from lab techniques to computer programs to analyze whole batch of genes at once. The gene expression level analysis has provided a challenge for Bioinformatics to develop new analytical tools for better understanding from gene to expression level and to overcome the existing problems as stated below:

- To solve the folding pathway of a protein given its amino acids sequence.
- To deduce biochemical pathway given collection of RNA expression profiles.
- Protein structure prediction.
- Homology searches.
- Multiple alignment and phylogeny construction.
- Genomic sequence analysis and gene finding.

Bioinformaticians are the tools- builders and it's critical that one understands the biological problem as well as computational solutions in order to produce useful tools. Research in Bioinformatics and computational biology can include anything from abstraction of the properties of a biological system into a mathematical or physical model, to implementation of new algorithms for data analysis to the development of databases and web tools to access them.

Challenges in Bioinformatics

The implicit goals of Bioinformatics is to read the entire genomes of living things, to identify every gene to match each gene with the protein it encodes and to determine the structure and function of each protein with the help of softwares and techniques. Detailed knowledge of gene sequence, protein structure and function and gene expression pattern to understand how life works at the highest possible resolution. Therefore, Bioinformaticians needs to create novel and enhanced algorithms for data mining, analysis, comparisons, etc. People with math and programming skills are highly required to bring fresh approaches and knowledge. A whole lot of coding should be done to mangle all the data but the growth in data along with its increasing complexity has thrown quit a few challenges which be divided into number of areas as discussed below:

Data Management and Organization

In the early handling of Bioinformatics data used to be carried out by individual research groups, but with advent of new automated experimental methods, generating enormous amount of data, for the storage and organization of these

complex data responsibility as led to new of national and international initiatives. Initially these databases were simple flat files, however increasingly relational databases are being used for better performance. Web interface has become an integral part of such databases for retrieval and analysis of the content of the databases. The challenges in the field of data management and organization are grouped as mentioned below:

- How to incorporate data and annotation emerging from different sequencing centres into a searchable resource, providing a user friendly interface, with extensive links to relevant resources and databases under development elsewhere.
- How to integrate genomic data with existing and future experimental result related to genome mapping, gene expression, protein function, protein-protein interactions, metabolic pathways etc. in both graphical and text based formats.
- How to establish a database that can incorporate sequence information from other related species.
- Providing web based access, email-based access in order to ensure full access to the sequence and related resources to all regardless of geographic location.

Data mining

Data mining techniques are used to discover new, previously unknown and hidden patterns in large data set, to represent and interpret those patterns in a human intelligible way. The common data mining tools are clustering, classification, dependency modelling, and regression. However Data mining in Bioinformatics is hampered by many facets of biological database including their size, number and diversity due to the lack of standard ontology to assist the querying of them as well as heterogeneous data of quality and provenance information they contain. Another problem is the range of levels and expertise present amongst potential users so it becomes difficult by the database curators to provide access mechanism appropriate to all. The integration of biological database is also lacking so it is very difficult to query more than one database at a time. Therefore, it is important to examine what are the important research issues in Bioinformatics and develop new Data mining methods for scalable and effective analysis which means to design new algorithms for software development (A challenge for the professionals working in this area).

Software and Algorithm Development

Bioinformatics generally prefer to develop software tools required for their very specific are of research. Algorithms used in such programs are often found to be too simplistic and not optimized. In most cases modification are required in some way of other to fit into the structures and requirements of problems of other researchers. Yet another challenge in development of Bioinformatics software tools is algorithm development. A special problem in the area of Bioinformatics is that programs are computationally very intense, so development of better algorithms in terms of resource requirements is a great challenge. Areas where better algorithms are a pressing necessity are sequence alignment, protein structure prediction, protein function prediction, protein-protein interactions etc.

Challenges and Opportunities

There are hundreds of technical hurdles to overcome before advances such as designer drugs and cures for genetic diseases can become affordable and as commonplace as over the counter drugs. For example, virtually all the advances in the analysis of genetic diseases require new computer-enabled technologies. Similarly, most molecular biologists concede that sequencing that human genome was relatively trivial task when compared with the challenges of understanding the human proteome.

Looking beyond the computational hurdles that will inevitably be overcome by the computer science community, there are broader issues and implications related to ethics, morality, religion, privacy, and economics. For example, the high-stakes economics game of biotechnology pits two groups, against each other. The first group consists of proponents of custom medicines, genetically modified foods, and cross-species cloning for species conservation and organ creation for transplantation. The second group consists of those who question the bioethics of embryonic stem cell research, the wisdom of creating "Franken foods" that may permanently after the ecology of the planet, and the morality of creating clones of pets of even people.

There are legal issues as well. For example, much of the bioinformatics R&D community is at least aware of bitter patent wars, with the realization that whoever the control of the key patents has in the field will enjoy a stream of revenues that will likely dwarf those at the height of the dotcom era. Rights to genetic codes (the sequences of base pairs found on strands of DNA) have the potential to both impede academic R&D and to guarantee the commercial success of drug development companies. The resolution of these and related issues depends on public politics and international laws that will define the rights of those who work in the field.

Despite these challenges, Bioinformatics has made significant inroads into the technical and social fabric of many nations. For example, China now feeds a significant percentage of its population with genetically modified foods. DNA vaccines promise cures for not only genetic diseases, but also for acquired diseases such as AIDS. In addition, as a result of sequencing the human genome and the genomes of other organisms, scientist have a better understanding of diseases form cystic fibrosis to malaria employing this create new better therapeutics.

As the history of computers and networks has demonstrated, the rate of change in computer-enabled technological innovation is accelerating with practical application of computing language. In this regard, Bioinformatics should be considered by programmers, systems architects, and other computer technology professionals as an opportunity to take a proactive role in defining and shaping not only their future, but the future of humanity as well.

Future Challenges for Life Sciences

Education

It is widely predicted that the application of high-throughput technologies to the quantification and identification of biological molecules will cause a paradigm shift in the life

sciences. However, if the biosciences have to evolve from a predominantly descriptive discipline to an information science, practitioners will require enhanced skills in mathematics, computing, and statistical analysis. Universities have responded to the widely perceived skills gap primarily by developing masters programs in Bioinformatics, resulting in a rapid expansion in the provision of postgraduate Bioinformatics education. There is, however, a clear need to improve the quantitative and analytical skills of life science undergraduates.

Bioinformatics Systems: Challenges and Current Workflow Definition Approaches

The number of web services, applications, data sources and scripts available to biomedical researchers is increased day by day. Complex analysis, annotation, and data integration could potentially be orchestrated from these services, if only one could be seamlessly connected. Biomedical researchers often do not have the time of expertise required to download, install, and adapt theses tools, now would the input and output formats required make this adoption easy. Some of the age-old mysteries of biology like embryogenesis, Morphogenesis, development, nerve function, behaviour and aging and how they are determined by genes are still unresolved problems to Bioinformaticians.

Current Challenges to Bioinformatics

There has developed a core set of "truths" in the field which focus effort and provide a set of paradigms upon which individual investigators work. The information about diabetes and the glucokinase gene reveals at once the great array of tools and databases that have become available, but also highlights significant work that remains-in particular the integration of Bioinformatics tools with clinical informatics tools concerned with the delivery of care.

The linkage of clinical medical information to molecular information represents one of the primary challenges for Bioinformatics in next century. As the genome sequencing projects mature and complete, we will have the genetic DNA sequences of both humans and a host of human pathogens, thus informatics tools will be necessary to deliver this information in appropriate ways to medical decision makers. The information will have direct impact on decisions about diagnosis, prognosis, treatment and epidemiology. Genomic information is already playing a leading role in the generation of new therapeutics of medicine. The ability to associate particular genes with particular organs, and the ability to associate defects in these genes with disease now allow drug targets to be identified primarily by computational analysis. Instead of the old paradigm of expensive, repetitive screening of candidate drug compounds against targets of interest one can now imagine a scenario in which DNA sequence information is selectively collected in a patient (or group of related patients) to determine the set of proteins involved in a pathological process. These proteins are then analyzed computationally to understand their functional properties and to look for places where their function can be augmented, diminished or modified (depending on the nature of the disease process). Other computational techniques are then used to design small compounds that interact with these proteins, based on principles of structural interactions. Finally, the compounds are analyzed and compared with known medications to assess possibilities for drug-drug interactions and toxicities.

The entire drug discovery process, under this (currently not possible) scenario is done computationally up to the point when the medication is actually synthesized and tested in animals. The expensive, large scale "shotgun" screening of today is avoided, and can be automated with computational technologies. The promise of integrating molecular biological information with the processes of delivering improved patient care and accelerating the discovery of useful new therapeutics requires significant progress in a number of areas, and these constitute the primary challenges to Bioinformatics.

Improved Support for Biomedical Investigation in a Data-Intensive Era

Bioinformatics does not only study the flow of information from genes to organisms and populations, but also studies the ways in which biomedical investigators use information in their cycles of hypothesis generation and testing. The same data about which we are so excited threatens to confuse and frustrate investigators because of its volume, it therefore becomes critical to develop methods to assist investigators (both clinical and basic science) In the robust analysis of data. These methods include the development of useful paradigms to support biomedical collaboration at a distance. How should scientists interact with resources that store data and computational methods for manipulating this data? How scientific results should be published, made available for others, indexed and effectively communicated. For example, the graphical display of biomedical information can be used to summarize information effectively. Certain domains of biology have developed standard conventions for the display of data Computer technologies are required to capture these convention and use them for the automated creation of graphics to display the contents of databases of the results of new algorithms. Similarly, as the developers of bio computing tools make them available to the scientific community, there is a danger that tools will be misapplied and data misinterpreted. Computational methods are required to bring relevant information to the attention of investigators, so that rapid progress can be made, and redundant work can be minimized. In addition to the incremental improvement of existing algorithms and data repositories, the development of simulation capabilities, the highly linked storage of data, and the development of tools to support the use of these capabilities represent the three major areas into which most current Bioinformatics efforts can be classified.

1. By providing algorithms,
2. Databases,
3. User interfaces,

And statistical tools, Bioinformatics makes it possible to do exciting things such as compare DNA sequences and generate results that are potentially significant and industrially valuable.

1. Data translation – translating data form one format to another

Biological data exists all over the world as various Web services, which provide biologists with much useful information. However, heterogeneous data formats present a technical hurdle for biologists to fully take advantage of the information. It needs some power tools to handle this issue. The grid technology could helps biology tools with high performance and high throughput. The process of information integration of heterogeneous biological data is complex and difficult. This demands the implementation of platforms like Bio-Java into our system for translating data into XML format so that the heterogeneous biological data sources can be integrated.

2. Protein folding

Prediction of protein folding rates from amino acid sequences is one of the most important challenges in computational and molecular biology. Since the biological macromolecules exist in highly states. So understanding of their behaviour requires the knowledge of all the conformations that they populate. It conducting by generating the 3D structure of the macromolecules. The structure of large complexes can be determines, but initially the resolution is quite low. Several investigations have been carried out to understand/predict the folding rates of proteins from protein 3D Structures.

3. Drug design and discovery

Bioinformatics was seen as an emerging field with the potential to significantly improve how drugs are found, brought to clinical trials and eventually released to the marketplace. Structural biology and Bioinformatics have assisted in lead optimization and target identification for lead discovery, exploiting high- throughput methods of structure determination which provide powerful approaches to screening of fragment binding. Computer- Aided Drug Design (CADD) is a specialized discipline that uses computational methods to simulated drug-receptor interactions. CADD methods are heavily dependent on bioinformatics tools, applications and databases. Bioinformatics has deciphered many key targets for drug discovery. These system pose significant challenges not only for characterization using structural techniques but also making the design of small molecule antagonists a difficult task as inter-surface protein are flat and similar. These challenges underline the importance of new approaches and the key roles of both academia and industry in advancing the growth of Bioinformatics in relation to drug discovery.

Thus it can be said that unlike genomics/ proteomics data, most drug, drug metabolism, ADR and ADME data is still in books or journals- not in electronic from this limits development of tools, databases and predictive software. As more data is made electronic, would lead to increased use of simulation and modelling software to predict ADME, ADR and toxicology.

The future…

- Expression of experimental data
- Difficulty in interpreting data
- Need for new paradigm for comparing with data & extracting new knowledge from it.
- Greater integration
- More freeware and greater web-accessibility
- Greater use of text mining and machine learning methods
- Focus on predictions

Challenges in Bioinformatics for statistical Analysis

Statistics is the science of learning from data includes data collection storage, management and drawing inferences from numerical facts which is directly related to Bioinformatics concepts. Statistics is needed for data mining i.e. process of exploration and analysis by automatic or semiautomatic means of large quantities of date in order to detect meaningful pattern & rules or to predict novel structure patterns, motif or relationships) from databases.

In simple terms data mining signifies knowledge discovery from the data an essential component of Bioinformatics which means to extract the hidden information from the biological databases.

Foremost challenge in front of statistician is to develop methodologies for structural and expression analysis of sequences/data visualization etc. As such statistical data mining approaches are ideally suited for Bioinformatics task as it is data rich but lacks a comprehensive theory of organization of the molecular model.

Data mining in Bioinformatics is lamp red by many facts of biological databases.

- Size
- Number
- Diversity
- Lack of standard ontology to aid the querying of them. Due to above mentioned reasons the proper integration of biological database thus creating a problems in property handling of queries. Both statistician and Bioinformatician have then own perspectives as mentioned below:

From Statistical Data Mines

Most Bioinformatician tend to be ignorant of statistical data mining, are too impatient for solutions expect statistical data mines to know the solutions before they have any data.

Statisticians and bioinformatician both should understand that the present time need is to understand the professionals have the some admit leaning from data" or turning "Data into information and information to knowledge". Therefore both have to work in synergy with the objectives as mentioned below:

- To examine, important research issues in Bioinformatics and develop new data meaning methods for sealable and defective analysis.
- A maturity challenge for statistical date mines and Bioinformaticians is to under their focus until time collaboration and unlocking of the smarts of the cell become reality.

References

1. Tarczy-Hornoch P, Minie M. Bioinformatics Challenges and Opportunities. Medical Informatics Integrated Series in Information Systems. Springer, Boston, MA. 2005, 8.
2. Jonathan CF, Pierre K, Holger D, Kristoffer F,

Alexandros S, Joseph B, *et al*. Biggest challenges in bioinformatics. EMBO Rep. 2013; 14(4):302-304. doi: 10.1038/embor.2013.34.

3. Al Kawam, Sen A, Datta A, Dickey N. Understanding the Bioinformatics Challenges of Integrating Genomics Into Healthcare. IEEE Journal of Biomedical and Health Informatics. 2018; 22(5):1672-1683. doi: 10.1109/JBHI.2017.2778263.
4. Guy Haskin Fernald, Emidio Capriotti, Roxana Daneshjou, Konrad J Karczewski, Russ B Altman. Bioinformatics challenges for personalized medicine, *Bioinformatics*. 2011; 27(13):1741-1748. https://doi.org/10.1093/bioinformatics/btr295
5. Khalid Raza. Application of data mining in bioinformatics; Indian Journal of Computer Science and Engineering. 1(2):114-118
6. Mohammed J Zaki, George Karypis, Jiong Yang. Data Mining in Bioinformatics (BIOKDD) Algorithms Mol Biol. 2007; 2:4. Published online 2007 Apr 11. doi: 10.1186/1748-7188-2-4 PMCID: PMC1852315