

An Integrative Study on Bioinformatics Computing Concepts, Issues and Problems

^{1*}Muhammad Zakarya, ²Izaz Ur Rahman, ¹Nadia Dilawar & ¹Reshma Sadaf

ABSTRACT: *Bioinformatics is the permutation and mishmash of biological science and ⁴IT. The discipline covers every computational tools and techniques used to administer, examine and manipulate huge sets of biological statistics. The discipline also helps in creation of databases to store up and supervise biological statistics, improvement of computer algorithms to find out relations in these databases and use of computer tools for the study and understanding of biological information, including ¹DNA, ²RNA, protein sequences, gene expression profiles, protein structures, and biochemical pathways [1]. The study of this paper implements an integrative solution. As we know that solution to a problem in a specific discipline may be a solution to another problem in a different discipline. For example entropy that has been rented from physical sciences is solution to most of the problems and issues in computer science. Another example is bioinformatics, where computing method and applications are implemented over biological information. This paper shows an initiative step towards that and will discuss upon the needs for integration of multiple discipline and sciences. Similarly green chemistry gives birth to a new kind of computing i.e. green computing. In next versions of this paper we will study biological fuel cell and will discuss to develop a mobile battery that will be life time charged using the concepts of biological fuel cell. Another issue that we are going to discuss in our series is brain tumor detection. This paper is a review on ³BI i.e. bioinformatics to start with.*

Keywords

¹DNA Deoxyribose Nucleic Acid
²RNA Rybo Nucleic Acid
³BI Bio Informatics
⁴IT Information Technology

-
- 1 Department of Computer Science, Abdul Wali Khan University (AWKU), Mardan, Khyber Pakhtun Khwa (KPK), Pakistan
2 School of Information Systems, Computing and Mathematics, Brunel International University, London, UK

* Corresponding Author

I. INTRODUCTION & CONCEPTS:

Bioinformatics is new research vicinity in Pakistan and Abdul Wali Khan University, Mardan (AWKUM) is one of those institutes who are interested and initiating this activity. SAIMS (Society for Advancement & Integration of Multiple Sciences), a very babyish research squad i.e. a sub domain of iFuture (a leading researcher's forum) in the Department of Computer Science (AWKUM), aims to set up Bioinformatics activity, in partnership with Department of Botany, Biology, Physics, Chemistry, Mathematics, Statistics, and even social sciences.

Bioinformatics is the arithmetical, statistical, algebraic and computing routines that endeavor to explain biological issues and problems using DNA, amino acid sequences and interrelated information. Usually speaking bioinformatics is the establishment and development of sophisticated information and computational technologies for troubles in biological science most frequently molecular biology and gradually more in other areas of biology as well. As such it deals with schemes for storing, accessing and analyzing biological data and information such as DNA, RNA, protein sequences, protein structures & functions and genetic interactions. Computer databases play a chief role in bioinformatics. Similarly the information retrieval mechanisms, techniques and algorithms that are applied in mathematical, statistical and computer sciences are also applicable to biological data.

Molecular biology is the understanding of biological processes at the molecular level i.e. it deals with the biological activity at the molecular level. Molecular biology is an amalgamation of genetics and biochemistry. Without Molecular biology, there would not have been bioinformatics. DNA holds the genetic information of all plants, animals, bacteria and some viruses. A property of living organism is that DNA is reproduced and migrated on to the subsequent generations. It contains instruction for making proteins. DNA contains a base plus a deoxyribose sugar and phosphate. The four bases of DNA are cytosine (C), adenine (A), thymine (T) and guanine (G). RNA contains ribose instead of deoxyribose and uracil (U) instead of thymine (T). The three major classes of RNA are transfer RNA (tRNA), messenger RNA (mRNA) and ribosomal RNA (rRNA). Gene is a segment of DNA

representing nucleotides that requisite for the fabrication of a functional protein or a functional DNS, RNA molecule.

The rest of the paper is structured as follows. In next section i.e. Section II we are going to define the terms like DNA, RNA, and Gene etc. In section III we have a brief discussion on bioinformatics and bioinformatics computing. In section IV we give an overview of the use of computing methods and application to biological data like databases. In section V a sketch is given on issues, hurdles and problems that computer scientist or biologists are or will face in bioinformatics. In section VI, the role of computer networks and distributed computing is discussed in term of biological data. In subsequent section VII an overview of bioinformatics literacy in Pakistan is sketched. We conclude in section VIII, with some future directions and work in subsequent section IX.

II. RELATED WORK & GENERAL TERMS

i. Biology

Biology is the study of different living organisms like plants, humans etc. Biology is the learning of life. In conjunction with chemistry & physics, biology is one of the biggest and most imperative branches of science. At the uppermost level, biology is divided into different sections based on the kind of mortal being considered for study. For example zoology is the study of animals, microbiology is the study of microorganisms and botany is the study of plants. More short and snappy classifications based on the grouping of organism being studied, biology restrains numerous other expert sub-disciplines, which may spotlight on just one kind of organism or address organisms from diverse categories [5]. This take account of biochemistry that is the crossing point and edge between biology and chemistry; ecology which studies the interactions between organisms themselves, molecular biology which looks at life on the molecular level, cellular biology which studies diverse types of cells and how they work, physiology which looks at organisms at the level of tissue and organs, ethology which studies the deeds of animals particularly complex animals, and genetics overlapping with molecular biology, which studies the rules & regulations of life i.e. DNA, RNA etc [7].

ii. Molecular Biology

Molecular biology is the understanding of biological processes at the molecular level i.e. it deals with the biological activity at the molecular level. Molecular biology is an amalgamation of genetics and biochemistry. Without Molecular

biology, there would not have been bioinformatics. It deals with the study of molecular building blocks such as nucleic acid and proteins [5].

iii. Cell

Cell is the structural or functional unit of the body or living organism. There are two components of cell i.e. nucleus and cytoplasm. Nucleus has nuclear materials which have chromosomes. There are 46 chromosomes which has 23 pairs in a human body. Chromosomes are the combination of DNA. Some organisms are unicellular like bacteria and protozoa, which contain only a single cell. Others are multi cellular like plants, animals and fungi, which carry out thousands of biochemical reactions in each minute and produce a new cell each time.

iv. Enzymes

Special organic substance composed of polymers and amino acid which acts as a catalyst to police and regulates the speed of chemical reactions that are implicated in the metabolism process of living organisms.

v. DNA

DNA holds the genetic information of all plants, animals, bacteria and some viruses. Deoxyribo means sugar, nuclei means nitrogenous base and acid stands for phosphoric acid. DNA is the structural or functional unit of the cell. A property of living organism is that DNA is reproduced and migrated to the next generation. It contains instruction for making proteins. DNA contains a base plus a deoxyribose sugar and phosphate. The four bases of DNA are cytosine (C), adenine (A), thymine (T) and guanine (G). The chromosome composed of single molecule of DNA shaped like a twisted ladder that is composed of linked chemical compounds known as nucleotide, which consists of sugar, phosphates and bases. Bases consist on purine and pyrimidine. Pyrimidine consists of cytosine (C), thymine (T) while purine is a group of adenine (A) and guanine (G), where adenine (A) usually pairs with thymine (T) and guanine (G) pair with cytosine (C). Purine and pyrimidine are called nitrogenous bases. A nucleotide may also consist of pyrimidine, sugar and phosphate. The way in which all these bases or two groups i.e. purine and pyrimidine are combined and integrated to make a DNA are called pairing rules. It states that always purine make pair with pyrimidine i.e. adenine (A) with

thymine (T) and guanine (G) with cytosine(C). Always two nitrogenous bases make double or triple bond i.e. hydrogen bonds. In summary DNA is double standard and ladder type and it is composed of large number of nucleotides or nitrogen bases. The structure of the integration of DNA is called DNA sequence and it is unique in nature. DNA is also called chromosomal DNA.

DNA is the crucial foundation of genetic information in cells. Humans, plants, animals, and bacteria all living things contain DNA. DNA is physically passed from generation to generation, bestowing definite persona and character of parents to their kids. The cause why kids have bodily characteristics from each of their parents e.g. an infant or teen may have their mother's eye color and father's hair color is for the reason that they received half their DNA from each parent.

vi. RNA

RNA contains ribose instead of deoxyribose and uracil (U) instead of thymine (T). The three major classes of RNA are messenger RNA (mRNA), ribosomal RNA (rRNA) and transfer RNA (tRNA). Ribo means sugar, again nucleic is nitrogenous base and acid stands for phosphoric acid. RNA is single standard and it is synthesized from the DNA structure. RNA is important for transmitting of message or information within the cell and outside the cell. Bases consist on purine and pyronidine. Pyronidine consists of cytosine (C), uracil (U) while purine is a group of adenine (A) and guanine (G), where adenine (A) usually pairs with uracil (U) and guanine (G) pair with cytosine(C). Purine and pyronidine are called nitrogenous bases. The three different types of RNA have different functionality. For example mRNA transmit message to ribosome and rRNA reads the mRNA. tRNA makes proteins and transfers or sends it to ribosome. RNA is used in protein synthesis or production.

vii. Proteins

Proteins are fundamental parts of organism like nucleic acids that participate in nearly every process within the cell. Proteins define the structure and function of living materials. It contains four elements, carbon, hydrogen, oxygen and nitrogen. Most proteins restrain some sulphur. These elements are bonded together as a compound called amino acid. So proteins are made up of amino acids. So far 20 amino acids are commonly founded in proteins. Each amino acid has two groups, carboxyl group i.e. COOH and

amino group i.e. NH₂. These two groups are bonded jointly by peptide bond which produces polypeptide chain where the number of amino acid is from 40 to 500. Proteins are biochemical compounds consist of one or supplementary polypeptides usually folded into fibrous form or globular form facilitating a biological function.

a) Proteins Structure

The majority of proteins fold into distinctive 3 dimensional structures. The form into which a protein logically folds is know as its natives conformation. The structure of the protein depends on the number of polypeptides, amino acid type and sequences. Proteins molecules may be fibrous as keratins that are founded in hair, fur, nails, claws, outer skin, cornea of eye, bones and ligaments. Others are globular e.g. hemoglobin globular proteins often consist of two or more polypeptides.

b) Protein Sequences

Protein sequencing is a procedure to conclude the amino acid sequence of a protein in addition to which conformation the protein agreed to and the scale or degree to which it complexes with any non peptide molecule recovering the structure and function of proteins in living being. It is a significant tool & gizmo for understanding cellular process and allows drugs that target definite metabolic to be invented more easily. The two major direct methods of proteins sequencing are edman degradation reaction and the mass spectrometry. It is also achievable to produce an amino acid sequence from DNA & mRNA sequence encoding the proteins.

c) Functions of proteins

Proteins are extremely essential, significant and vital molecules in our cells. Every protein within the body has a definite unique function. Some proteins are involved in bodily movement while others are involved in structural support or in defense against germs and diseases. Antibodies proteins makes the defense system, contractile proteins are responsible for movement, hormonal proteins are messenger proteins help in coordination of certain bodily activities, enzymes proteins facilitate biochemical reactions, structural proteins provides support, storage proteins stores amino acids and transport proteins migrate molecules from one position to another around in the body.

- Proteins make the major component of cell.
- Proteins make muscles that help in movement.
- Proteins perform the function to carrying oxygen from lungs to different parts of body.
- Proteins help in food digestion
- Proteins are important source of energy
- It is the main component of some seeds e.g. pulses, peas etc
- Proteins help in blood clotting
- Many proteins act like enzymes that catalyze biochemical reactions and are fundamental to the process of metabolism.

viii. Gene

Gene is a segment of DNA representing nucleotides required for the fabrication and invention of a functional protein or a functional RNA molecule. Gene is the basic unit of heredity found in the cell of living organisms e.g. from bacteria to human. It determines the physical characteristics that an organism inherits such as color of hairs or shape of face. Genes composed of small segments of DNA (a molecule that form thread like structures called chromosome). The study of functions and behavior of genes is called genetics. A sequence of DNA is called gene and different cell DNA; they are different on the basis of arrangement not on the basis of their structure. Any change in of disarrangement of gene is called mutation. Mutations are of two types i.e. useful and abnormality.

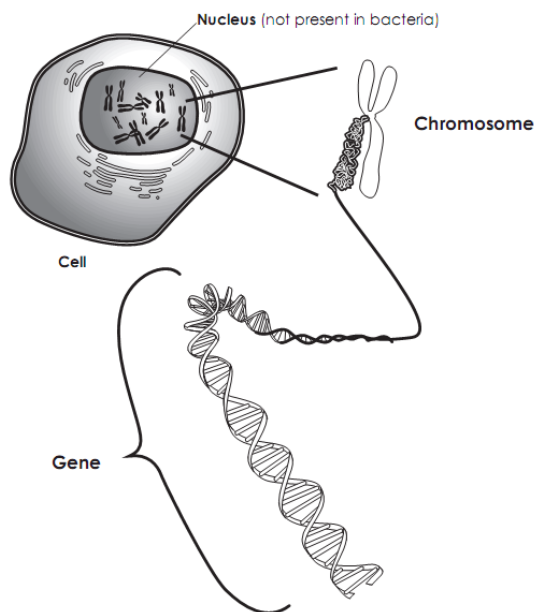


Fig 1: DNA, Cell, Chromosome, Gene [6]

ix. Chemoinformatics

Chemoinformatics is the application of computational, statistical and mathematical methods to chemical problems. Computer technology, particularly computational power, networking & storage space and capacity, has sophisticated to an arena that it is proficient of managing some of the recent disputes and hurdles posed by chemistry and biology [2]. This makes it achievable to grip the enormous quantity of data or information that is being generated as a result of the international genome project [3, 17]. JMOL [18] is a simulation environment that researchers from biochemistry, chemistry and bioinformatics departments use for 3D representation of molecular structure of different compounds. The tool provides an easy use for the beginners. Although all functionalities are not GUI based, but command prompt can be used to execute all commands, which are available. JMOL is a java based application that is portable and can be used on any platform or architecture.

III. BIOINFORMATICS:

Up-to-the-minute molecular biology and medical research encompasses an increasing quantity of information, as well as an increasing diversity of information. The usage of informatics to organize, manage, and scrutinize these information has subsequently become a significant element of biology and biomedical research. Bioinformatics is the mixture of computing, mathematics, and biology to address this requisite [4]. Data management and knowledge discovery are the two most important issues that are addressed in [4]. The integration of biological and computational sciences have resulted in this new era and story of bioinformatics, which is an interested area for students, researchers and scientist at most teaching departments and knowledge places and institutors of repute. Following are some common applications of bioinformatics.

- Drug design process
- Heart modeling
- DNA sequencing
- Biological fuel cells
- Plants pathology
- Criminal investigations
- Health care
- Agriculture
- Brain tumor detection
- Environmental protection
- Crops / development of crops varieties
- Industrial applications etc

Thus bioinformatics is an interdisciplinary vicinity at the meeting point of biological, mathematical, statistical, computer, and information sciences essential and crucial to handle, route, and recognize bulky amounts of information, for illustration from the sequencing of the human genome, or from bulky databases holding information about animals and plants for utilize in discovering and developing new drugs [13].

IV. BIOINFORMATICS COMPUTING:

Bioinformatics Computing (BC) means to apply informatics and computational methodologies and techniques on biological information and data to process, manage and understand large amount of data i.e. biological information and to store these information in data stores i.e. databases in computational terms. Machines have played a major role in human life al the time. When biologist felt that there research data in enormous and hence needs to be properly managed and stored, they find computer and information technology that can easily handle their issues. Hence a new field was introduced i.e. bioinformatics and bioinformatics computing.

V. ISSUES & PROBLEMS IN BIOINFORMATICS COMPUTING:

Bioinformatics is an integrated science and hence face a lot of problems and challenges. Computer scientists often have no knowledge of biology and biologist's lacks computer programming and information theory as well. Therefore there is a need to minimize the gap between these scientists that cover a lot of other disciplines as well. So we have to develop some special teaching mechanism and course contents in academia. In [8] the author has sketched a blueprint of teaching mechanism and course contents. To congregate the requirements of computer scientists, critical facets of the chemistry of existence & life have to be presented, together with an introduction to procedures, issues, hurdles & challenges and scientific idea behind genomics, molecular biology, and biotechnology. To assemble the desires of biologists, essential skills and ethics in information technology and computational science must to be offered, together with the development and application of computer programming proficiency. We need critical bioinformatics experts in our academia but it's too difficult as most experts are hired by private organization where there is a restriction on every information to be disclosed to audience.

Another possible solution is to initiate some integrated programs that will address multiple subjects and hence providing a broad background in all facets of bioinformatics. In such a program students have more choices and options to pursue in the field. Our university has also started a BS (Integrated) 4 years program that we think, will results in scientists that will have knowledge of

every subjects and discipline, but still they need some better curriculum to address our future needs and requirements. Following are some major issues currently researchers and academia need to focus.

- a) **Gap between biological and computer sciences**
Because biology and computer sciences are two different fields, biologists don't know the complex computer programming languages and computer experts needs biological data and knowledge before writing programs [8].
- b) **explosion of data**
The steep amount of information in Genbank is now rising on exponential velocity and as data type beyond RNA, DNA and protein sequences being to undergo the same kind of sudden increase, simple managing, monitoring, accessing and presenting this data to the user in an intellectual form is a decisive and critical task [15].
- c) **large computation**
It requires a large scale computations (for scheduling, exact monitoring and fault tolerance purpose), as several fields and technologies are merged in bioinformatics.
- d) **quality of data**
The quality of large scale data is also notoriously variable, with relative high error rates and low productivity.
- e) **algorithms issues**
Bioinformatics algorithms need to encompass complex scientific assumptions that can do complicated programming and data modeling.
- f) **complex biological data (an opportunity for computer scientists)**
Biological data is extremely difficult and interlinked. A speck on DNA array for example, is linked not only to instant information about its intensity but to layers of information about DNA sequence, genomic location, structure, function and many more. Creating information systems that permits biological scientists to effortlessly track these links exclusive of getting in a sea of information, is also a gigantic opportunity for computer scientist [13, 15].
- g) **overlap graphs**
A memory proficient data structure representing accurate match overlap graphs with the application for subsequent generation DNA assembly.
- h) **availability of large tools**
The wide variety of the analysis tools available today means that it is often difficult for the non specialist to choose an appropriate tool for his

specific problem, in order to select the most suitable method for the particular problem.

i) education

Development of suitable bioinformatics course curriculum for secondary, under graduate and graduate education level is a major problem in different academia and laboratories.

j) database issues

The number of publicly available data repositories has grown in an exponential rate in 2002. There are thousands of publicly bimolecular data sources. These data sources are developed independently, the data they contains are represented in a wide variety of formats, annotated using a variety of methods and may or may not be supported by a database management system.

k) data integration issues

Many of the problems facing genomic data integration are related to data semantics. For example the difference between the meaning of the data representing in a data source and the difference between the semantics within a set of sources is a major issue.

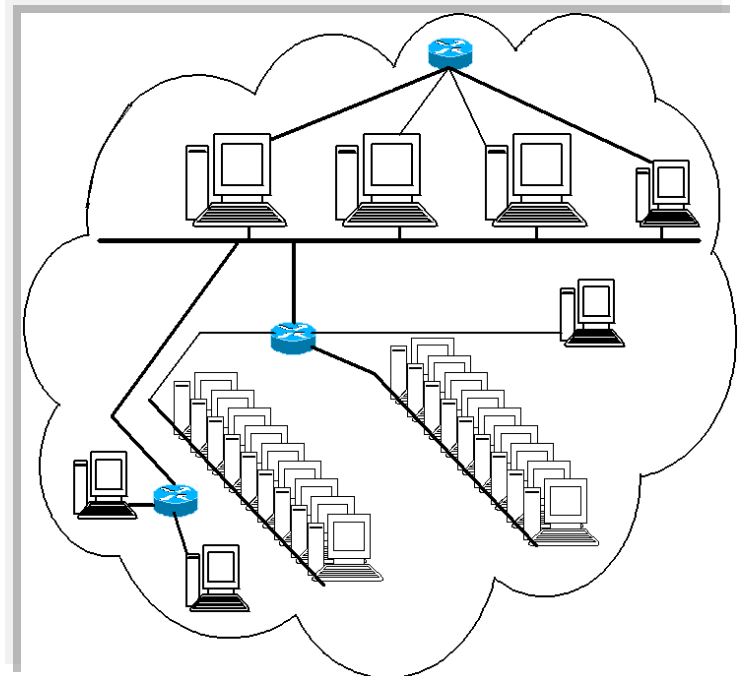


Fig 2. High Performance Computing

VI. ROLE OF COMPUTER NETWORKS IN BIOINFORMATICS:

New generations of networked computing systems, aided by the global existence of the Internet infrastructure, are playing an increasingly important role in the society and civilization. Distributed computing has altered the face of computing and presented rapid and clear-cut solutions for a diversity of complex issues & problems for diverse disciplines and fields. With the rise of new computing era distributed systems like Grid, Cloud systems, there is a need to implement cloud computing over the sheer amount of data in bioinformatics. Grid computing is term used for On Demand Computing, where services are provided to customers on demand base. The underlying communication media in Grid computing is local area networks, high speed Ethernet. It means Grid may span specific and limited geographic area. Cloud computing indicates the infrastructure as a “Cloud” from which businesses and customers are competent and capable to access & use applications from everywhere in the world using on demand techniques. Storage, processing, managing and monitoring of huge amount of data will be easy if Grid or Cloud i.e. distributed concepts are applied and implemented over bioinformatics [9, 10, 11, 12].

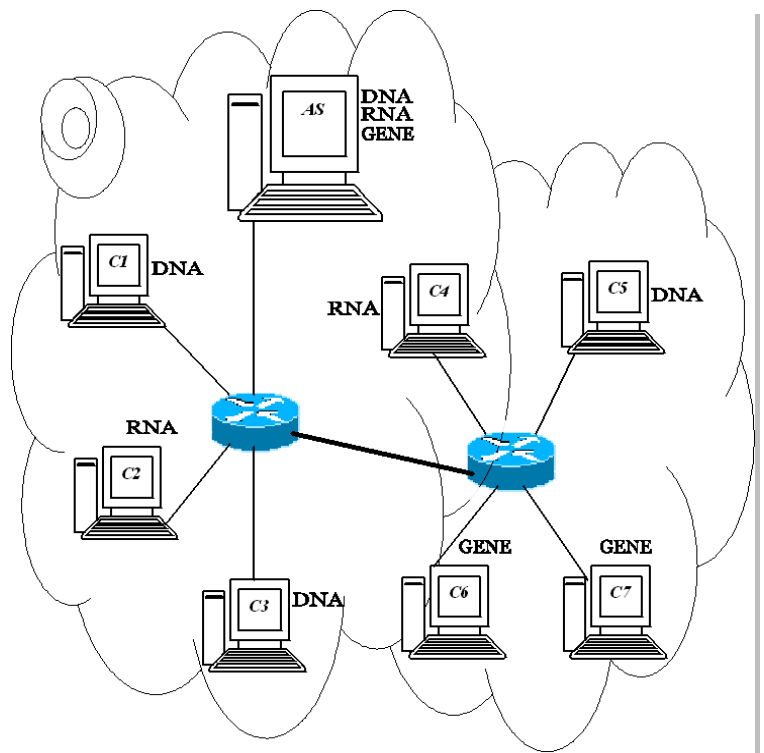


Fig 3. HPC in Bioinformatics

VII. HPC OVER BIOINFORMATICS:

High Performance Computing can be implemented over the huge amount of biological data, to achieve High availability and better performance. Fig 3 shows a HPC environment where DNA, RNA and GENE related database is replicated in the network. Fig 2 shows HPC environment where more than one systems are trying to find a solution for a specific problem. Fig 3 shows how biological databases are replicated over the network and how they interact each other to access and service user in a distributed fashion. The AS is responsible to authenticate and authorize users and to provide an easy way for accessing and storing biological data. AS maintains all the replication information and can disconnect users depending on server load to balance the workloads i.e. load balancing. The architecture provides the following benefits.

- *Limited & Local Security Policy*
- *Little computation as compared to Global security policy*
- *User accesses AS, strong authenticated & authorization check*
- *Performance Scalability*
- *Load balancing*
- *Quality of Service*
- *No need for resources to check the user identity*
- *Local & Quick allocation of resources by AS*
- *No Single point of failure, affect some part of the HPC*
- *Fault Tolerance*

One of the main disadvantage of this architecture is that AS are required to inform all corresponding AS in case of new node to any geographical community means if some database is updated, all the replicated databases are also needed to be updated as soon as possible. If [20] the authors have proposed a distributed architecture CISM, which is still under development. This agent based systems showed in its initial study and results that it is adequate for building and designing of complex systems taking the advantage of composite services. CISM provide a robust, flexible, modular and adaptable solution that can cover most requirements of a wide diversity of distributed systems. [19] describe in full detail the opportunities for the Pakistani IT industry within bioinformatics, and have sketched a very knowledgeable summary of current bioinformatics activities in Pakistani universities, academia and laboratories. Different institutes of repute are considered for their bioinformatics literacy and goals.

VIII. BIOINFORMATICS AND PAKISTAN:

Two fields biology and computer sciences merged with each other and introduced a new area called bioinformatics, which is mostly interested area for students,

researchers and teachers in academia, organizations and laboratories. In Pakistan bioinformatics was 1st introduced in 2003 with the help of two institutes i.e. COMSATS Institute of Information Technology and Muhammad Ali Jinnah University, but nowadays approximately 12 or more institutes are working in this field. At these institutes and universities researchers and bioinformatics expertise were focusing on guess & forecast of the functions of the newly recognized genes, finding the protein interaction, restriction analysis, identification of mutations, determination of the active sites of enzymes and etc. In Pakistan biotechnology is used from corner to corner to an extensive range of applications including agriculture, health care, crime investigation, environmental protection, industries, development of new crops varieties, drug designing, finding targets of drugs and many more. In Pakistan LIMS (laboratory Information Management Systems) and medical informatics and clinical trails management are used for laboratory tests to improve the results of databases.

As bioinformatics is mostly interested area for researchers, students and teachers in Pakistan. Students and researchers have made a group that has produced useful platform for young bioinformaticians. The group is called Regional Students Group (RSG-Pakistan), which came into being in January 2010. The aim of this platform is to develop the new generation, help students in their projects, encourages students and researchers to come & join, to train students and provides opportunities to commit with experts in computational biology.

A brief survey at different institutes, research forums and organization were taken into consideration. Fig 4 shows computer and internet availability and use to researchers, students and employers. In some cases we were unable to get the opinions. In Fig 5 we have sketched the bioinformatics tools that researchers, students and employers at different institutes, research forums and organizations uses for analysis of there problems and research issues. Fig 6 shows Jmol i.e. a java base simulator for representation of 3D molecular structure of different compounds. Mostly students related to biochemistry and chemical sciences were observed that were some what familiar with Jmol. Fig 7 is short survey of PROSPECT. PROSPECT stands for PROtein Structure Prediction and Evaluation Computer ToolKit that is a protein structure prediction system that make use of a computational procedure called protein threading to assemble a protein's 3-D model. We observed that students at most universities and institutes are not properly aware of bioinformatics literacy and its role in our society. Most of the students were in favor of complexity of bioinformatics. Students of computer literacy were founded that they have no interest in bioinformatics computing, although a few were well educated on bioinformatics.

iFuture a leading Research Group in Computer Science Department is also working on bioinformatics computing and arranging different seminars and trying to explore this new era of computational and biological

sciences. This work is a result of the efforts of this leading research group.

Government of Pakistan needs proper attention to the lose development in bioinformatics literacy. They needs to offer scholarship for higher study in bioinformatics which will encourage our students and will increase their number and interest in this new and immature field of science & technology in Pakistan. Educational institutes need to arrange workshops and seminars on this growing field. Similarly the experts are required to discuss some easy bioinformatics applications and their uses in rural areas, which will force the people on thinking and hence this field of biotechnology will make a place in our developing country.

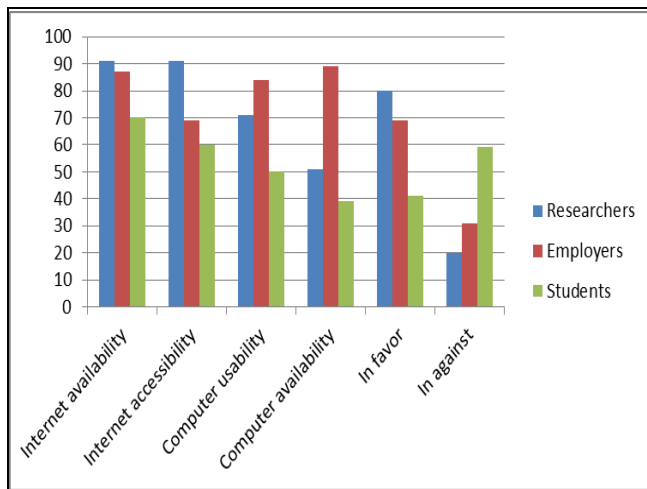


Fig 4. Bioinformatics Literacy ratio in Pakistan

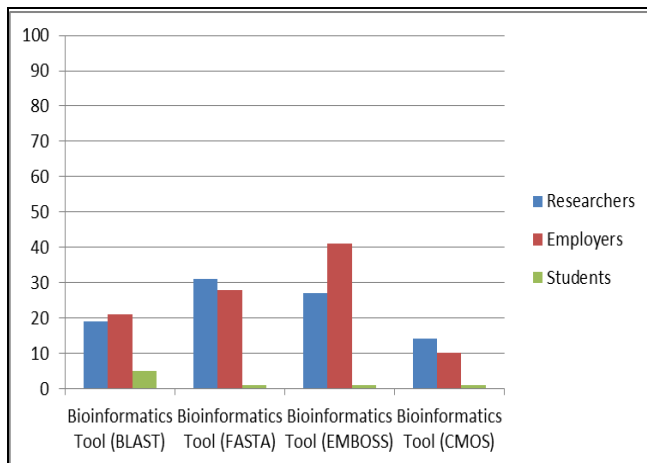


Fig 5. Bioinformatics Literacy ratio in Pakistan

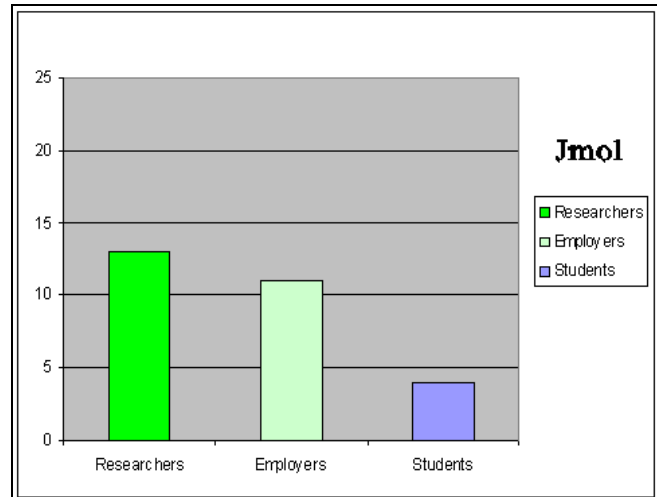


Fig 6. Jmol usage

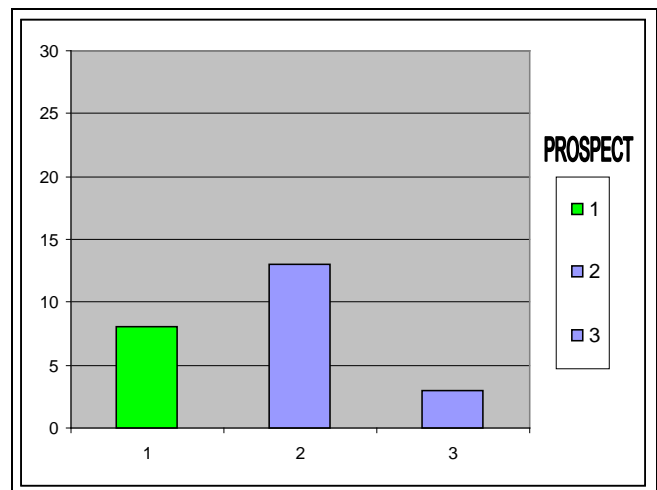


Fig 7. PROSPECT usage in Pakistan

IX. CONCLUSION:

Biology is problematical and complex field of science. In other fields such as physics, chemistry or engineering sciences and Information Technology, applications are developed from well-characterized ethics and ideology. With biotechnology on the leading edge of molecular biology research, it can be intricate or even impossible to predict the conclusions of manipulations and they can have surprising consequences and cost. For the reason that it is impossible to wholly guess the result of these events, scientists must carry out experiments, take observations, purify theories, and to finish develop functional applications. This is why biotechnology research is so difficult, time uncontrollable, and loaded with astonishing setbacks and frustrations.

In this article we have discussed some bioinformatics issues and proposed the need for HPC to efficiently maintain, access and store all biological data and information. Our next step is to implement HPC over bioinformatics and will study the proposed architecture statically and will have some simulations results.

X. FUTURE WORK:

This article discusses bioinformatics on a very basic level that is easily understandable by computer scientist. The purpose is to draw a sketch and some overview of the challenges that need to be addressed. This work is basically an initiative step to implement the cloud computing environment over bioinformatics. Biological data is so huge that a single processor cannot be addressed for its execution neither can be stored on a single system. Therefore High Performance Computing like Clusters, Grids and Clouds are needed to implement for easy data storage, access and management.

ACKNOWLEDGMENT

This work is fully supported by Abdul Wali Khan University, Mardan, Khyber Pakhtun Khwa (KPK), Pakistan. The author(s) of this article are greatly thankful to SAIMS i.e. Society for Advancement & Integration of Multiple Sciences for full guidance and major support. iFuture is also given a number of credits for arranging seminars on the subject matter. iFuture is a Research Group at the Department of Computer Science. The authors are very thankful to ¹Dr. Zidong Wang, who helps us and guided us for the entire period of this survey.

REFERENCES:

- [1] Bourne, Philip E., and Helge Weissig, eds. Structural Bioinformatics. Methods of Biochemical Analysis 44. Hoboken, NJ: Wiley-Liss, 2003. BA Call Number: 572.8733
- [2] Hwa A. Lim, Bioinformatics and Chemoinformatics in the Drug Discovery Cycle, Santa Clara, California 95051- USA 1997
- [3] Mapping and Sequencing the Human Genome, (National Research Council, National Academy Press, Washington, D.C., 1988).
- [4] Vladimir B. Bajic, Vladimir Brusic, Jinyan Li, See-Kiong Ng, Limsoon Wong, From Informatics to Bioinformatics, Australian Computer Society 2003
- [5] Nikola Kasabov, Igor A. Sidorov, Dimiter S. Dimitrov, Computational Intelligence, Bioinformatics and Computational Biology: A Brief Overview of Methods, Problems and Perspectives, Computational and Theoretical Nanoscience Vol.2, 473–491, 2005
- [6] Yali Friedman, Ph.D., Building Biotechnology, Third Edition, ISBN-13, Hardcover: 978-0-9734676-5-9, Sofcover: 978-0-9734676-6-6
<http://www.wisegeek.com/>
- [7] Gustavo Caetano-Anollés, Teaching Bioinformatics: Needs and Challenges, gca@uiuc.edu
- [9] Yanpei, Vern Paxson, Randy H.Katz, What's New About Cloud Computing Security, University of California at Berkeley, January 20, 2010
- [10] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Above the Clouds: A Berkeley View of Cloud Computing, University of California at Berkeley, February 10, 2009
- [11] Muhammad Zakarya, Izaz Ur Rahman, Mukhtaj Khan, Cloud QoS, High Availability & Service Security Issues with Solutions, BUJICT July 2011
- [12] Muhammad Zakarya, Ayaz Ali Khan, Hameed Hussain, Grid High Availability & Service Security Issues with Solutions, ICIIT 2010, IEEE
- [13] Bryan Bergeron, Bioinformatics Computing, Prentice Hall of India, 2006
- [14] Stephen A.Krawetz, David D.Womble, Introduction to Bioinformatics – A theoretical and practical approach, Humana Press 2009
- [15] See-Kiong Ng and Limsoon Wong, Accomplishments and Challenges in Bioinformatics, 1520-9202/04/\$20.00 © 2004 IEEE
- [16] Lawrence Hunter, Molecular Biology for Computer Scientists - ARTIFICIAL INTELLIGENCE & MOLECULAR BIOLOGY, Chapter 1
- [17] C. Stan Tsai, An Introduction to Computational Biochemistry, ISBN 81-265-1105-2
- [18] JMOL online & offline documentation
- [19] Dr. Ahmad Ijaz Gilani, Sadia Zaidy, Opportunities for the Pakistani IT industry within bioinformatics, Ministry of IT, Government of Pakistan, Islamabad, December 2005
- [20] Javier Bajo, Carolina Zato, Fernando de la Prieta, Ana de Luis, Dante Tapia, Cloud Computing in Bioinformatics, Springer-Verlag Berlin Heidelberg 2010

Keywords

¹School of Information Systems, Computing and Mathematics, Brunel International University, UK

- [21] Anthony Finkelstein, James Hetherington, Linzhong Li, Ofer Margoninski, Peter Saffrey, Rob Seymour, Anne Warner, Computational Challenges of Systems Biology, 0018-9162/04/\$20.00 © 2004 IEEE
- [22] Atul J. Butte, Challenges in bioinformatics: infrastructure, models and analytics, TRENDS in Biotechnology Vol.19 No.5 May 2001
- [23] Peter Tarczy-Hornoch, Mark Minie, BIOINFORMATICS CHALLENGES AND OPPORTUNITIES, MEDICAL INFORMATICS
- [24] See-Kiong Ng and Limsoon Wong, Accomplishments and Challenges in Bioinformatics, 1520-9202/04/\$20.00 © 2004 IEEE
- [25] N.M. Luscombe, D. Greenbaum, M. Gerstein, What is bioinformatics? An introduction and overview, Yearbook of Medical Informatics 2001
- [26] Mario Cannataro, Carmela Comito, Filippo Lo Schiavo, and Pierangelo Veltri, Proteus, a Grid based Problem Solving Environment for Bioinformatics: Architecture and Experiments, IEEE Computational Intelligence Bulletin 2004

ABOUT THE AUTHOR(S):



Currently the author of this paper is working as lecturer in computer science department of Abdul Wali Khan University Mardan, Pakistan. The author of this paper is a new researcher to the field of new emerging computing technologies like Grid, Cloud and Green Computing. He has done MS in Computer Science and is interested for a doctorate degree in computer engineering. Currently the author of this paper is working on security issues in Grid and Cloud computing i.e. distributed systems. The author is interested in implementing Cloud computing over bioinformatics. He is also working and interested to Green the Cloud.



Izaz Ur Rahman is working as Lecturer in the same university and department. He is enrolled for a PhD study in Brunel International University, UK. His research area is distributed systems and bioinformatics.

Nadia Dilawar and Reshma Sadaf are students of master in Computer Science, at Computer Science Department of Abdul Wali Khan University. Both of them are interested in bioinformatics computing. This work is an initiative step toward the implementation of a new research area i.e. implementation of Grid, Cloud on demand computing environment over bioinformatics