# Bioinformatics Applications in life Sciences: Concepts and Stance

**Bhagavati Parekh**

Asst. Professor,Junagadh Kelavani Mandal Managed BCA College Meghraj House, Gandhi Chawk,Junagadh.

**ABSTRACT**

After the introduction and success of the World Genome Project, there were drastic changes in the field of Biology and Life Sciences with respect to the research activities and organization of the data thereof. Today, tremendous data flow produced with new innovations has increased the need of computational support to collect, store, retrieve, analyze, and correlate huge data sets of complex information. At the same time, the growth of the computational power for processing and storage also has increased the necessity for deeper knowledge in the field. New interdisciplinary approaches have opened the arena of naval innovations which otherwise could not have been done, had each of the discipline been working on its individual account. Researchers are addressing this challenge by adopting mathematical and statistical software, computer modeling, and other computational methods, as a result, bioinformatics has become the latest engineering discipline and an inevitable discipline to solve and sort the life science concerns.

**KEYWORDS** | Bioinformatics, Computational Biology, Bioinformatics software, Bioinformatics packages, Bioinformatics work-flow management systems.

## 1 Introduction:

Modern scientific research depends on computer technology to organize and analyze large data sets. In fact, nowadays it is difficult to imagine an area of knowledge that can continue developing without the use of computers and informatics [1]. Tremendous data flow produced with new innovations has increased the need of computational support to collect, store, retrieve, analyze, and correlate huge data sets of complex information. At the same time, the growth of the computational power for processing and storage has also increased the necessity for deeper knowledge in the field. New interdisciplinary approaches have opened the arena of entirely naval innovations which otherwise could not have been done, had each of the discipline been working on its individual account. Nevertheless, the crux has opened up new doors of development; it has come with the inherent challenges of new methods of statistical analysis and modeling though. Indeed new technologies are producing data at a rate that outpaces our ability to analyze its biological meanings. Researchers are addressing this challenge by adopting mathematical and statistical software, computer modeling, and other computational and engineering methods. As a result, bioinformatics has become the latest engineering discipline.

## 2 The Basics of Bioinformatics:

### 2.1. The origin:

Infact, the idea behind keeping the records of information dates about a century back. An Australian monk named Gregor Mendel cross fertilized different colours of same species of flowers and kept records of all the flowers which produced other colors to explain and prove that the inheritance of traits could be more easily explained if it is controlled by the factors passed down from generation to generation. Since then, the genetic record keeping has come a long way.

### 2.1.1. The History:

In the year 1990, the Human Genome Project was started by Human Genome Organization after completing and publishing the first complete genome map of bacteria Haemophilus Influenza. By 1991, a total of 1879 human genes had been mapped. In France, in 1993, Genethon, a human genome research center produced a physical map of the human genome. Three years later , Genethon published the final version of the human genetic map. This concluded the end of the first phase of the Human Genome Project.[2]. This required the creation of huge datasets, which ultimately resulted in the repository of big databases. Such information repository and processing

was later on referred to as "Bioinformatics" by Paulin Hogeweg around 1970 [3]. Using the computer technologies huge datasets were managed in the tabular formats and regarded as the pioneer databases to provide platform for the further research. These databases were Genbank, EMBL and DNA Database of Japan to store and compare the DNA sequence data erupting from the human genome and other genome sequencing projects.

### 2.1.2. Computing in biology:

Due to the massive volumes of biological data and the required preciseness in analysis, it became mandatory to use computers to achieve the dual goal of storage and retrieval upon requirement. Here, the power of computer took up the prominent stance in the process due to its virtue of introducing ease in handling the large quantities of data and probe the complex dynamics observed in the nature.

Today, the biological data are being produced on a phenomenal rate [4]. Anthony Kerlavage of Celera has cited that an experimental laboratory can produce over 100 gigabytes of data a day with ease [5]. It is even impossible to envisage the rough estimation of total data flow occurring at a particular time period. This enormous data flow is matched with the incredible processing power of computer technology; thus making it possible to record and process the gigantic biological data repository.

Thus bioinformatics deals with designing and deploying efficient software tools for accomplishing the above quoted tasks in a fast and precise manner. So, bridging the gap between the real world of biology and precise logical nature of computers requires an interdisciplinary perspective.

### 2.2. The Definition and the concept:

The classical definition of Bioinformatics is "the mathematical, statistical and computing methods that aim to solve the biological problems using DNA and amino acid sequences and related information".

In a broader sense, it can be defined as the study of information content and information flow in biological systems and processes.

Hence, Bioinformatics is concerned with creation and maintenance of databases of biological information whereby researchers can access existing information and make additions

of new information.

### 2.2.1 Aims and Objectives:
The aims of bioinformatics are basically three-fold. They are

1. Organization of data in such a way that it allows researchers to access existing information & to enter new information as they are produced.
2. To develop tools and resources that help in the analysis of data.
3. Use of these tools to make analysis of the individual process in detail, check and produce the resemblances, if any.

### 2.2.2. Scope of Bioinformatics:
Bioinformatics implements its applications in the areas of computer science, information science, computer and information technology, communication technology to solve complex problems in life sciences and particularly in biotechnology. Bioinformatics makes it possible to create multimedia databases, provides tools to carry out data analysis and facilitates modeling of molecules and biological systems on computer workstations as well as in a network environment.

### 3. Software and Tools
Bioinformatics basically deals with database creation, data analysis and modeling. Tools for bioinformatics range from simple command-line tools, to more complex graphical programs and standalone web-services available from various bioinformatics companies or public institutions.

### 3.1. Open-source bioinformatics software
Many free and open-source software tools have existed and continued to grow since the 1980s [6]. They can be enlisted as mentioned below.

### 1. Bioconductor:
This is the suit of applications that provides tools for the analysis and comprehension of high throughput data.

### 2. BioPerl:
It is thriving open source project that indulges in the development and support of various computational algorithms that create the biological applications.

### 3. FASTA AND SSEARCH:
FASTA provides a heuristic search with a protein query. FASTX and FASTY translate a DNA query. Optimal searches are available with SSEARCH (local), GGSEARCH (global) and GLSEARCH (global query, local database).

### 4. BLAST :
Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches.

### Matlab:
It is a high-level language for numerical computation, visualization, and programming. It allows users to analyze data, develop algorithms, and create models for the biologically generated data.

### 3.2. Various Web services in bioinformatics
A number of application are designed to configure themselves on SOAP(Simple Object Access Protocol) and REST (Representational State Transfer) based interfaces.

### 1. SSS (Sequence Search Services), MSA (Multiple Sequence Alignment)   and BSA (Biological Sequence Analysis)
These resources are service-oriented bioinformatics applications, which exhibit the applicability of web-based bioinformatics solutions. They cover the range of standalone tools with a common data format under a single, standalone or web-based interface, to integrative, distributed and extensible bioinformatics workflow management systems.

### 2. Cytospace:
It is an open source software platform for visualizing complex networks and integrating these with any type of attribute data.

### 3. Galaxy:
Galaxy is an open, web-based platform for data intensive biomedical research. Whether on the free public server or the own instance, reproduction and sharing of complete analysis is made possible by Galaxy.

### 3.3. Bioinformatics workflow management systems
Any workflow management system should be designed specifically to make up and perform a set of computational or data manipulation steps or a workflow. A Bioinformatics application is the specialized work flow management system that allows integrative data capture, efficient and pertinent work flow management processes and composite reports in the prescribed format.

### Various platforms for running the Work flow Management System:
The platforms on which the Work Flow Management Systems can run are:

### 1. Galaxy:
In this context, Galaxy provides a scientific workflow, data integration and data and analysis persistence and publishing platform to make the computational biology accessible to those which don't have computer related experience.

### 2. Kepler:
Kepler helps in process and data monitoring, provenance information, and high-speed data movement solutions. The Workflows are represented as directed graphs, in which the nodes represent discrete computational components, and the edges represent paths, along which data and results can flow between components.

### 3. Anduril:
Anduril is an open source component-based workflow framework for scientific data analysis. The data analysis can be done in systematic, flexible and efficient way, particularly in the field of high-throughput experiments in biomedical research.

### 4. Various Bioinformatics projects worldwide:
The management and, more importantly, the dissemination of information to the world are implemented today with various Bioinformatics Projects. Some of them are:

### 1. BioJava
BioJava Project is built up with the java unique feature of platform independent application deployment, which uses various Java tools for processing biological data that includes objects for manipulating sequences, dynamic programming, file parsers, simple statistical routines, etc.

### 2. BioPerl
The BioPerl project is an international association of developers of Perl tools for bioinformatics and provides an online resource for modules, scripts and web links for developers of Perl-based software.

### 3. BioXML
A part of the BioPerl project, this is a resource to gather XML documentation, DTDs and XML aware tools for biology in one location.

### 4. bioperl-db
Bioperl-db is a relatively new project intended   to transfer some of Ensembl's capability of integrating bioperl syntax with a standalone Mysql database to the bioperl code-base.

### 5. Biopython and biojava
Biopython and biojava are open source projects with very similar goals to bioperl. However their code is implemented in

python and java, respectively. With the development of inter-face objects and biocorba, it is possible to write java or python objects which can be accessed by a bioperl script, or to call bioperl objects from java or python code.

**Conclusion:**

With its ability to organize, sort, reproduce and analyze the bi-ological data with high throughput simulations, Bioinformatics has become an inevitable discipline to solve and sort the life science concerns. Various challenges emanating from the cross currents of various life science research activities can be well managed by this discipline with a broad application spectrum and continuously developing and growing mechanism. With the bright future prospects, the bioinformatics has reserved its inevitable place in the research areas of various life sciences.

**REFERENCES**

1. Heitor silverio Lopes, Computational Biology and Applied Bioinformatics, InTech, p.no 2  | 2. Bioinformatics web-Comprehensive educational resource on Bioinformatics www.geocities.com/bioinformaticsweb | 3. Hogeweg, P. (2011). "The Roots of Bioinformatics in Theoretical Biology". In Searls, David B. PLoS Computational Biology 7 (3): e1002021. | 4. Reichhardt T. It's sink or swim as a tidal wave of data approaches. Nature 1999;399(6736):517-20. | 5. Drowning in data, The Economist 26 June 1999. Extracted on 25/03/20 | 6. "Open Bioinformatics Foundation: About us". Official website. Open Bioinformatics Foundation. |