

Original Research Article

BIOINFORMATICS TOOLS IN GENOMIC BIOMEDICAL RESEARCH

Abstract

Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data

This paper was borne from the fact that some medical laboratories and related fields have failed to implement Bioinformatics tools as they relate to Genomic Biomedical experiments. The paper therefore seeks to study the extent of this involvement and in addition provide methods and resource reports that can provide novel information on technological tools that will be of interest to a broad readership in the area of Bioinformatics and Genomic research. The paper discusses the Computer as an Instrument in Bioinformatics research, basic Bioinformatics tools for Genomic research such as Genomic Databases, Genome browsers, tools for sequence alignment, softwares in Bioinformatics etc.

Keywords: *Bioinformatics Tools, Software, Genomic Browsers, Biomedical Database, Computer System*

1.0 Introduction

As an interdisciplinary field of science, bioinformatics combines computer science, statistics, mathematics, and engineering to study and process biological data. It is both an umbrella term for the body of biological studies that use computer programming as part of their methodology, as well as a reference to specific analysis "pipelines" that are repeatedly used, particularly in the fields of genetics and genomics. Genomic per se provides novel insights into the genome biology of all organisms, including advances in genomic medicine. This paper explores how the field of Bioinformatics has rapidly developed into an essential aid for Genomic data analysis and powerful Bioinformatics tools has been developed, many of them publicly available through the World Wide Web [1] .

On the advent of a completely assembled human genome, modern biology and molecular medicine stepped into an era of increasingly rich sequence database information and high-throughput genomic analysis. However, as sequence entries in the major genomic databases currently rise exponentially, the gap between available, deposited sequence data and analysis by means of conventional molecular biology is rapidly widening, making new approaches of high-throughput genomic analysis necessary. At present, the only effective way to keep abreast of the dramatic increase in sequence and related information is to apply biocomputational approaches [2].

Bioinformatics has become an important part of many areas of biology. For example, in the field of genetics and genomics [2], it aids in sequencing and annotating genomes and their observed mutations. Bioinformatics tools also aid in the comparison of genetic and genomic data and more generally in the understanding of evolutionary aspects of molecular biology. Other applications of Bioinformatics tools are in experimental molecular biology, where bioinformatics techniques such as image and signal processing allow extraction of useful results from large amounts of raw data. It plays a role in the text mining of biological literature and the development of biological and gene ontologies to organize and query biological data. It also plays a role in the analysis of gene and protein expression and regulation. At a more integrative level, it helps analyze and catalogue the biological pathways and networks that are an important part of systems biology. In structural biology, it aids in the simulation and modeling of DNA, RNA, and protein structures as well as molecular interactions.

As software designed for extracting the meaningful information from the mass of molecular biology / biological databases and to carry out sequence or structural analysis, Bioinformatics tools have been found useful in Genomic research and usage. According to Bioinformaticweb.tk [2], a Bioinformatics resource portal, factors that must be taken into consideration when designing bioinformatics tools, software and programs should include: the end user (the biologist), who may not be a frequent user of computer technology and the software tools, which must be made available over the internet given the global distribution of the scientific research community, among others.

Bioinformaticweb.tk further remarked that there were major categories of Bioinformatics Tools that included both standard and customized products to meet the requirements of particular projects such as data-mining software that retrieve data from genomic sequence databases and also visualization tools to analyze and retrieve information from proteomic databases. These can be classified as homology and similarity tools, protein functional analysis tools, sequence analysis tools and miscellaneous tools. Thus, over recent years, the field of bioinformatics has rapidly developed into an essential aid for genomic data analysis and powerful bioinformatics tools have been developed, many of them which are publicly available [3] through the World Wide Web. In this study, we summarize and describe the basic bioinformatics tools for genomic research such as: genomic databases, genome browsers, tools for sequence alignment, prediction, expression databases, and algorithms for promoter prediction.

2.0 Technology development

Many areas of critical importance to the realization of the genomics-based vision for biomedical research require new technological and methodological developments before pilots and then large-scale approaches can be attempted. Recognizing that technology development is an expensive and high-risk

undertaking, the National Human Genome Research Institute (NHGRI)[4] is nevertheless committed to supporting and fostering technology development in many of these crucial areas, including the following:

- *DNA sequencing.* There is still great opportunity to reduce the cost and increase the throughput of DNA sequencing, and to make rapid, cheap sequencing available more broadly. Radical reduction of sequencing costs would lead to very different approaches to biomedical research.
- *Genetic variation.* Improved genotyping methods and better mathematical methods are necessary to make effective use of information about the structure of variation in the human genome for identifying the genetic contributions to human diseases and other complex traits.
- *The genome 'parts list'.* Beyond coding sequences and transcriptional units, new computational and experimental approaches are needed to allow the comprehensive determination of all sequence-encoded functional elements in genomes.
- *Proteomics.* In the short term, the NHGRI expects to focus on the development of appropriate, scalable technologies for the comprehensive analysis of proteins and protein machines in human health and in both rare and complex diseases.
- *Pathways and networks.* As a complement to the development of the genome 'parts list' and increasingly effective approaches to proteome analysis, the NHGRI will encourage the development of new technologies that generate a synthetic view of genetic regulatory networks and interacting protein pathways.
- *Genetic contributions to health, disease and drug response.* The NHGRI according to [4], will place a high priority on creating and applying new crosscutting genomics tools, technologies and strategies needed to identify the genetic bases of medically relevant phenotypes. Research on the genetic contributions to rare and common diseases, and to drug response, will typically involve biological systems and diseases of primary interest to other NIH institutes and other funding organizations. Accordingly, the NHGRI expects that its involvement in this area of research will often be implemented through partnerships and collaborations. The NHGRI is particularly interested in stimulating research approaches to the identification of gene variants that confer disease resistance and other manifestations of 'good health'.

3.0 The Computer System – An Instrument Of Bioinformatics Tools

The Computer System, according to Ekwonwune [5], is an electronic device which takes in data, processes it and gives information as a result of that processing. The Computer System provides the platform through which the programs are designed for extracting the meaningful information from the mass of molecular Biology or Biological Databases and carry out sequence or structural analysis. This system makes it possible for the design of some useful Bioinformatics tools like BLAST, FASTA,

FMBoss, PROSPECT, COPIA. These are significant softwares applied in the area of Genomic Biomedical Research. The management of Bioinformatics tools and resources is performed largely by the System software. Ekwonwune added that, the system software is a significant determinant of whether a particular computer's resources are used optimally.

3.1 Databases

A database is an organized collection of data. Ekwonwune [6] posits that computers can store one record in a database that contain all the information it needs to do different task and can search its memory very quickly to find specific information. The NHGRI has been a major source of support for several major genetics/genomics-oriented databases, including the Mouse Genome Database *ibid*¹, the *Saccharomyces* Genome Database *ibid*, FlyBase *ibid*², WormBase *ibid*³, and Online Mendelian Inheritance in Man *ibid*⁴. The NHGRI will continue to be a leader in exploring effective solutions to the issues of integrating, displaying and providing access to genomic information.

4.0 Bioinformatics Tools

Bioinformatics tools are softwares that are designed for extracting the meaningful information from the mass of molecular biology / biological databases & to carry out sequence or structural analysis. According to Garry G. [7], a software is a set of rules/instructions given to computers to perform their tasks. These Bioinformatics tools are classified as application softwares which are designed to perform specific functions which according to Bioinformaticweb.tk are as shown below:

➤ **BLAST** (**B**asic **L**ocal **A**lignment **S**earch **T**ool) comes under the category of homology and similarity tools. It is a set of search programs designed for the Windows platform and is used to perform fast similarity searches regardless of whether the query is for protein or DNA. Comparison of nucleotide sequences in a database can be performed. Also a protein database can be searched to find a match against the queried protein sequence.

Depending on the type of sequences to compare, according to Bioinformatics, there are different programs:

- blastp compares an amino acid query sequence against a protein sequence database
- blastn compares a nucleotide query sequence against a nucleotide sequence database
- blastx compares a nucleotide query sequence translated in all reading frames against a protein sequence database
- tblastn compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames
- tblastx compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

- **FAST** homology searches All sequences. An alignment program for protein sequences created by Pearsin and Lipman in 1988. The program is one of the many heuristic algorithms proposed to speed up sequence comparison. The basic idea is to add a fast prescreen step to locate the highly matching segments between two sequences, and then extend these matching segments to local alignments using more rigorous algorithms such as Smith-Waterman.
- **EMBOSS** (European Molecular Biology Open Software Suite) is a software-analysis package. It can work with data in a range of formats and also retrieve sequence data transparently from the Web. Extensive libraries are also provided with this package, allowing other scientists to release their software as open source. It provides a set of sequence-analysis programs, and also supports all UNIX platforms.
- **Clustalw**: It is a fully automated sequence alignment tool for DNA and protein sequences. It returns the best match over a total length of input sequences, be it a protein or a nucleic acid.
- **RasMol**: It is a powerful research tool to display the structure of DNA, proteins, and smaller molecules. Protein Explorer, a derivative of RasMol, is an easier to use program.
- **PROSPECT** (Protein Structure Prediction and Evaluation Computer ToolKit) is a protein-structure prediction system that employs a computational technique called protein threading to construct a protein's 3-D model.
- **PatternHunter**, based on Java, can identify all approximate repeats in a complete genome in a short time using little memory on a desktop computer. Its features are its advanced patented algorithm and data structures, and the java language used to create it. The Java language version of PatternHunter is just 40 KB, only 1% the size of Blast, while offering a large portion of its functionality.

5.0 Application of Softwares in Bioinformatics

Softwares are applied in key areas of Bioinformatics to bring solutions to observed and anticipated problems in Biomedical technology. It has been emphasized by Ekwonwune E, that softwares are designed to perform specific functions. However, some of these softwares according to Wikipedia [8], the free encyclopedia, are as discussed below:

5.1 Java in Bioinformatics

Since research centers are scattered all around the globe ranging from private to academic settings, and a range of hardware and OS's are being used, Java is emerging as a key player in bioinformatics. Physiome Sciences' computer-based biological simulation technologies and Bioinformatics Solutions' PatternHunter are two examples of the growing adoption of Java in bioinformatics.

5.2 Perl in Bioinformatics

String manipulation, regular expression matching, file parsing, data format inter conversion etc are the common text-processing tasks performed in bioinformatics. Perl excels in such tasks and is being used by many developers. Yet, there are no standard modules designed in Perl specifically for the field of bioinformatics. However, developers have designed several of their own individual modules for the purpose, which have become quite popular and are coordinated by the BioPerl project.

Table 1: Bioinformatics softwares

Software	Description	Operating System
.NET Bio	A language-neutral bioinformatics toolkit built using the Microsoft 4.0 .NET Framework to help developers, researchers, and scientists.	.NET Framework
Biochemical Algorithms Library (BALL)	C++ library and framework for molecular modeling and visualization designed for rapid prototyping	Linux, Mac OS X, Windows
Bioclipse	Visual platform for chemo- and bioinformatics based on the Eclipse Rich Client Platform (RCP).	Platform
EMBOSS	Suite of packages for sequencing, searching, etc.	Platform
FACS	FACS machine output analyzer. Percentage to absolute cell numbers.	Linux, Mac OS X, Windows
Galaxy	Scientific workflow and data integration system	Unix-like
GenePattern	Scientific workflow system that provides access to more than 150 genomic analysis tools	Unix-like (public server); Linux, Mac OS X, Windows
GMOD	Toolkit for addressing many common challenges at biological databases.	Unix-like (server), Web browser (client)
GenGIS	Application that allows users to combine digital map data with information about biological sequences collected from the environment.	Windows, Mac OS X
GENTle	An equivalent to the proprietary Vector NTI, a tool to analyze and edit DNA sequence files	Platform
Integrated Genome Browser	Java-based desktop genome browser	Linux, Mac OS X, Windows

InterMine	Extensive data warehouse system for the analysis and integration of biological datasets	Cross-platform
LabKey Server	platform for integrating, analyzing and sharing data	Linux, Mac OS X, Windows
LeoTask	A lightweight & reliable framework for running computational tasks in parallel and aggregating the results (MapReduce) on a multicore computer	Linux, Mac OS X, Windows
STAMP	Software package for analyzing metagenomic profiles that promotes 'best practices' in choosing appropriate statistical techniques and reporting results.	Windows, Mac OS X

Source: Bioinformaticsweb.tk

6.0 Conclusion

bioinformatics in our interdisciplinary field of science which combines computer science, statistics, mathematics and engineering to study and process data. It embraces body of biological studies that use computer programming as part of their methodology and the fields of genetics and genomics. The study stresses the computer as a veritable instrument (Sine qua non) of Bioinformatics tools because of its major contribution to the design and development of technological tools like BLAST, FAST, EMBOSS, CLUSTLAW, RASMOI, etc the study also emphasized the computer as instrumental to the design and development of softwares in Bioinformatics like .NETBIO, Biochemical Algorithms, Bioclipse, FACS, Galaxy, etc. The Computer System provides the platform through which these programs are designed for extracting information from the mass of Molecular Biology or Biological Databases and carry out Sequence or Structural Analysis.

Analyzing biological data to produce meaningful information involves writing and running softwares – program that uses algorithms from Graph Theory, Artificial Intelligence, Soft Computing, Data Mining, Image Processing and Computer Simulation.

Preparing a vision for the future of Bioinformatics Tools in genomics research can be both daunting and exhilarating. This article captures the sense of excitement of the new discipline of Bioinformatics in genomics research. In this unique presentation of a revolutionary field of science, we introduced the domains of application for genomics and bioinformatics, such as sequencing and the comparative sequence analysis of model organisms, investigating genetically influenced diseases, the development of new active substances, agri-food and environmental genomics [12].

References

- [1] Teufel .A et al (2006), Current Bioinformatics tools in genomic biomedical research (Review), Department of Medicine, Johannes Gutenberg University, Mainz, Germany.
- [2] Bioinformaticsweb.tk. *Bioinformatics resource portal*
- [3] Hogeweg .P (2011), The Root of Bioinformatics in Theoretical Biology. *PLoS Computational Biology*. Inside Wiktionary, the free dictionary.
- [4] Francis S. et al () The National Human Genome Research Institute (NHGRI), National Institute of Health, Bethesda, Maryland, USA
- [5] Ekwonwune E. (2010), Basic Information System: *An Introduction*. Celbez Publishers, Owerri, Nigeria
- [6] Ekwonwune E. (2007), Information Technology: *A practical Approach*. Celbez Publishers, Owerri, Nigeria
- [7] Garry .G (1986), Computer Literacy: *Applications and Programming*. Design office, San Francisco.
- [8] Wikipedia the free encyclopedia
- Ibid¹ <http://www.informatics.jax.org/mgihome/MGD/aboutMGD.shtml>)
- Ibid² <http://genome-www.stanford.edu/Saccharomyces>
- Ibid³ <http://flybase.bio.indiana.edu>
- Ibid⁴ <http://www.wormbase.org>