

IMPACT OF BIOINFORMATICS TOOLS IN GENOMIC BIOMEDICINE

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, software in Bioinformatics etc.

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

1.0 INTRODUCTION

Bioinformatics is an interdisciplinary field of science consisting of computer science, statistics, engineering and mathematics, which are tools it uses to address the issues of biological data Research. As a biological study, computer programming constitutes a veritable tool to its understanding. However, genetics and genomics disciplines are parts of its tools also, as the later, (genomics) provides new insights into the genome biology of all organisms, one of which is genomic medicine. Thus, the paper focuses on the field of Bioinformatics as an essential aid for Genomic data analysis. Nevertheless, some Bioinformatics tools could be sited in the internet and the World Wide Web [1]

Bioinformatics tools in the opinion of Vondran Florence [2], consist of products that store, organize, evaluate, integrate, analyze and/or distribute biological data. Persidis [3], posits that Bioinformatics developed after automated protein and DNA sequencing technologies in the mid 70's. Nevertheless, mid-to late 80's saw the usage of computers as a control sequence repositories.

Furthermore, it was Zagursky and Russel [4], who analyzed the field of Bioinformatics to comprise the following: Computational genomics; Proteomics; transcriptional profiling, Functional genomics and structure function determinations.

The vital objectives of Bioinformatics, according to Himanshu [5], include the following:

- i. Organization of data in a user – friendly way;
- ii. Development of tools and resources that aid data analysis
- iii. Using data to interpret analyzed data

Relevant studies on modeling have been done using the concept of databases and some Bioinformatics tools, according to Himashu, like BLAST, FASTA, CLUSTALW, etc and the modeled structures are visualized using visualization tools such as Weblab, MOLMOL, Rasmol, etc. It is worthy of note here that modern biology and molecular medicine had taken their foothold in genomic information analysis study before the complete introduction of assembled human genome.

However, new approaches to high thorough put genomic analysis is obtainable now due to exponential rise of sequence entries in the major genomic databases, thus widening the gap between deposited sequence data and molecular biology analysis. Presently, the application of bio-computational approaches is the only affective way to keep abreast of the dramatic increase in sequencing and related information [6]

There is no gain-saying the fact that Bioinformatics has today become an important part of various fields of biology because genetics and genomics [6], which are biological entities helps in sequencing and annotating of genomes and their observed mutations. The tools associated with Bioinformatics had generally assisted in the understanding of the evolutionary aspects of molecular biology and in the comparison of genetic and genomic data.

Again, Bioinformatics techniques of image and signal processing had found application in experimental molecular biology, where extraction of useful results from large amount of raw data was possible. Besides, it had played a role in text mining of biological literatures and ontologies of biological genes' data query.

Furthermore, in the analysis of gene and protein expression cum regulation, Bioinformatics has found relevance. In the analysis and cataloging of biological pathways and networks, which are important units of system biology, Bioinformatics is also relevant. In structural biology, it aids the simulation and modeling of RNA, DNA, protein structures and interaction of molecules.

Bioinformatics has been found useful also in Genomic research with respect to information extraction and usage in molecular biology databases.

Nevertheless, in the design of bioinformatics tools' software, some of these factors may be considered; (a) End-User; and (b) Software tools [6]. The designed tools should be available and user-friendly.

In the case of data-mining and visualization tools, for example, customization of some of the Bioinformatics tools may be necessary. Although, in recent years, the field of bioinformatics has rapidly developed into an essential aid for genomic data analysis and many of these tools are readily available [7] through the internet. Therefore genome browsers, genomic databases, sequence alignment and prediction tools and algorithms can be sited.

2.0 Bioinformatics and Technological Development

New technological developments and methodologies prior to pilot surveys and large-scale approaches are necessary for the realization of some critical areas of genomics-based vision for biomedical researches. Recognizing the immensity in achieving development in this field, an organization called; National Human Genomic Research Institute (NHGRI) [8], put up a technology-based support aimed at realizing the goals of genome-based biomedical researches. Its areas of intervention includes;

- DNA sequencing
- Genetic variation
- Proteomics
- Genome "Parts Lists"
- Networks and Pathways and
- Genetic researches in the areas of health

diseases and drug responses [8]. The strength or achievements of these interventions are well-known. The organization, however, seeks for partnership and collaborative efforts in achieving these goals aimed at stimulating research approaches in their areas of interventions on genomic biomedical researches.

3.0 The Computer System – An Instrument Of Bioinformatics

The Computer System, according to Ekwonwune [9], 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, FMBOS, PROSPECT, COPIA as shown in www.bioinformaticsweb.net. 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 [10] 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*², FlyBase *ibid*³, WormBase *ibid*⁴, and Online Mendelian Inheritance in Man *ibid*⁵. The NHGRI will continue to be a leader in exploring effective solutions to the issues of integrating, displaying and providing access to genomic information.

Databases in the words of SRS [11], and Entrez [12], contain volumes of information such as nucleotide and protein sequences, metabolic pathways and single nucleotide polymorphism, inter – alia. Contributing, Hamosh et al [13], posits that most of these databases have direct relevance with online Meridelian Inheritance in man (OMIN). But Thorisson and Stan [14], said that databases were important resources for pharmacology nomics [15], and cardiovascular research.

Algorithm for Data Interpretation

Ekwonwune posits that algorithm involves an outline of a set of instructions for solving problems, which cover broad range of areas such as sequence analysis, visualization, Statistics, Modeling and Simulation of cellular analysis. However, some of them were developed to handle new challenges and issues arising from genome – wide data sets production. Some of these algorithms can also handle large amount of data. It is note – worthy to note that traditional algorithms may force problems of cluster algorithms used for the identification of related groups of genes in DNA microarray experiments [16].

4.0 Advanced Bioinformatics Tools

Advanced Bioinformatics deals with the use and development of bioinformatics and biostatistical tools in the analyses of clinical and biological data since large scale technology generates great amounts of data that must be processed for extracting relevant information. Therefore, reliable tools for dissection and handling of data are of great importance for the outcome of large scale research. In all, the following are some of the objectives of advanced bioinformatics, which according to [17] include the following:

- I. To develop statistical models, inference procedures and tools useful in a wider biomedical context;
- II. To address the specific and computational issues raised by the scientific activities;
- III. To develop image analysis tools for nucleomics and nucleotyping and
- IV. To use bioinformatics infiltration of genome information.

Some of the Advanced Bioinformatics Software in the opinion of [18] are as stated below:

- a) MPromDb – Mammalian Promoter Database
- b) IsoformEx – Isoform level gene expression estimation from RNA – Seq Data;
- c) TPD – Modeling Transcription Factor Binding Site Profiles from ChIP – Seq Data;
- d) NPEBseq – Differential Expression analysis based on RNA-seq data;
- e) PIGExClass – Platform – independent Isoform – level Expression based classification – system
- f) Lumi, a NU initiative that develops, maintains updates and distributes software libraries in the Bioconductor R package for Illumina expression and methylation microarray data analysis and genotyping estimation
- g) GeneAnswers, a NU initiative that facilitates gene – concept network analysis and generation of protein – protein interaction networks, which focuses on clinical and biological interpretation based on given genes.
- h) Metacore
- i) The Database for Annotation, Visualization and Integrated Discovery (DAVID)
- j) Weight gene Co – expression Network Analysis (WCGNA)
- k) Gene Set Enrichment Analysis (GSEA)

Advanced Bioinformatics Applications

Advanced Bioinformatics has so many applications. On Genetic research, it describes how information technology is used in biological research or how bioinformatics is used to perform genetic research in which case bioinformatics tools of BLAST, ORFind, ClustalW and Cn3D are used to analyze genetic bar-coding” gene that allow for identification of animal species, where one examined DNA sequences from different animal species, investigate the relationship between protein structure and function, and explore evolutionary relationships among eukaryotic organisms [19].

Application of Advanced Bioinformatics for Gene Expression

For example, Data Acquisition and Basic analysis, the use of advanced scientific algorithms, Affymetrix GeneChips Command Console (AGCC) provides several different analysis applications whose analysis

options include conversion of intensity data into expression results, allele detection, singular nucleotide polymorphism detection, and nucleotide analysis, among others [20].

There is also advanced data analysis for Genotyping. For example, in the ArrayAssist Expression Software powered by avadis, there is some ease of usage, while still offering a full spectrum of advanced analytical features such as new statistical tests (multi – way ANOVA and non – parametric statistical test), data mining tools (classifiers for machine learning – based prediction) and sophisticated, dynamic, interactive visualization options. For advanced users, a new scripting engine is featured in all products of the ArrayAssist line. The engine allows scripting of complex functions and user - defined workflows and is fully integrated with external programs such as the R statistical software.

There is also this Exemplar Genotype Solutions from Sapio Sciences, which is an innovative, advanced genetic analysis tool. The exemplar Genotyping Analysis Suite was devised to model the genetic roots of disease using advanced machine learning technology combined with traditional statistics, where analysis solution takes raw genotype data all the way through the analysis pipeline to biological interpretation. Exemplar provides a consistent and easy – to – use approach of creating projects and grouping experiments within those projects. New experiments are created with the help of a wizard to walk through the options for various types of experiments, which can then be run at the click of a button through [20].

Classification Of Bioinformatics Tools Software

As application software meant to perform specific tasks, Bioinformatics' tools software may be classified as follows:

(i) Basic Local Alignment Search Tool [BLAST]

Under Homology and similarity tools, this software is a set of search programs designed for windows platform used to perform fast similarity searches irrespective of whether the query is of protein or DNA origin. It can also be used to compare nucleotide sequence in databases; search protein database for a match against queried protein sequence and so on.

Kinds Of Blast

According to Bioinformaticsweb.net, different software or programs on BLAST exist based on the type of sequence comparison analysis.

- (a) BLASTp which compares Amino Acid Query Sequence Against Protein Sequence Database
- (b) BLASTn which compares Nucleotide Query Sequence Against Nucleotide Sequence Database
- (c) BLASTx which compares translated nucleotide query against protein sequence database.
- (d) EBLASTx that compares the six-frame translations of a nucleotide query sequence against the six-frame translation of a nucleotide sequence database
- (e) EBLASTn – compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames.

- (ii) **Fast Alignment Search TOOL (FAST):** This is a homology search tool that searches ALL sequences. It was an alignment for protein program created in 1988 by Pearsin and Lipman as

an heuristic algorithm meant to speed up comparison of sequence - in Bioinformatics. Here, basically, is an idea to add a fast pre-screen step to locate the highly matching segments between two sequences, which could be extended to alignments using more rigorous algorithms in the likes of that of Smith – Waterman.

(iii) **European Molecular Biology Open Software Source (EMBOSS)**

This software works with a range of formats to transparently retrieve sequence data from the web. Its extensive libraries' characteristics enable other scientists release their software as open source language providing a set of sequence analysis programs. It supports all platforms of UNIX operating system.

(iv) **Protein Structure Prediction and Evaluation Toolkit (PROSPECT)**

This is a protein-structure prediction system software that employs a computational technique called protein threading to construct a 3-D model.

(v) **RasMol:** This powerful research tool displays DNA structure, proteins and smaller molecules. It is a user-friendly software.

(vi) **Clustlaw:** A fully-automated sequence alignment tool for protein and DNA sequences that returns the best match over a total length of input sequences, be it a protein or nucleic acid.

(vii) **PatternHunter Software Toolkit:** It is a built on Java-based platform with about 1% BLAST size that offers a large portion of its functionality using little computer Ram memory. It can identify all approximate repeats in a complete genome.

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.

5.0 Application of Softwares in Bioinformatics

Software designed for extracting meaningful information from masses of molecular biology cum biological databases for the sake of carrying out sequence or structural analysis is called a Biometrics tool. Garry G. [21] defined such software as a set of rules or instructions given to the computer to perform given tasks. 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, that software are designed to perform specific functions. However, some of these software according to www.bioinformaticsweb.net, are as discussed in table 1 below:

5.1 Java Language Software in Bioinformatics

Java language software is a user-friendly, open source and object-oriented program used in research centres all over the globe at private, academics, hardware and operating systems' establishments, including Bioinformatics players. Two examples of Bioinformatics companion that uses Java Language are physiome computer-based biological simulation technologies and Bioinformatics Solution PatternHunter. (http://biotech.fyicenter.com/resource/bioinformatic_tools_and_software_2.html)

5.2 Perl Language Software in Bioinformatics

This language is associated with string manipulation, regular matching expressions; file processing and data format inter-conversions, which are some of the text processing tasks done in Bioinformatics. Although Perl – specific standard design module is yet to be developed in Bioinformatics but individual

220 modules in the platform of Perl language have been developed by players in this language to deal with
 221 BioPerl projects. (http://biotech.fyicenter.com/resource/bioinformatic_tools_and_software_2.html)

222 **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]

6.0 Conclusion

Bioinformatics is an interdisciplinary field of science which combines computer science, statistics, mathematics and engineering to study and process data which 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 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. This piece of work discussed 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.

References

- [1] Teufel .A et al (2006), Current Bioinformatics tools in genomic biomedical research (Review), Department of Medicine, Johannes Gutenberg University, Mainz, Germany.
- [2] Vorndran .C and Florence .R (2002), Bioinformatics: Patenting the Bridge Between Information Technology and the Life Sciences *IDEA - The Journal of Law and Technology*, vol. 4
- [3] Persidis .A, (1999) Data mining in biotechnology," *Nature Biotechnology*, vol. 17
- [4] Zagursky .R and Russel .D (2001), Bioinformatics: Use in Bacterial Vaccine Discovery, *Biotechniques*, vol. 31, pp. 636-659,
- [5] Himanshu S. (2016), International Journal of PharmTech Research, CODEN, USA
- [6] Bioinformaticsweb.tk. *Bioinformatics resource portal*

- [7] Hogeweg .P (2011), The Root of Bioinformatics in Theoretical Biology. *PLoS Computational Biology*. Inside Wiktionary, the free dictionary.
- [8] Francis S. et al, The National Human Genome Research Institute (NHGRI), National Institute of Health, Bethesda, Maryland, USA
- [9] Ekwonwune E. (2010), Basic Information System: *An Introduction*. Celbez Publishers, Owerri, Nigeria
- [10] Ekwonwune E. (2012), Information Technology: *A practical Approach*. Celbez Publishers, Owerri, Nigeria
- [11] Zdobnov, E. M., Lopez, R., Apweiler, R., and Etzold, T., (2002) The Ebi Srs server—new features, *Bioinformatics*
- [12] Wheeler, D. L., Church, D. M., Federhen, S., Lash, A. E., Madden, T. L., Pontius, J. U., Schuler, G.D., Schriml, L. M., Sequeira, E., Tatusova, T. A., and Wagner, L. (2003), Database resources of the national center for biotechnology, *Nucleic Acids Res.* **31**.
- [13] Hamosh, A., Scott, A. F., Amberger, J., Bocchini, C., Valle, D., and McKusick, V. A. (2002), Online Mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders, *Nucleic Acids Res.* **30**:52–55.
- [14] Thorisson, G. A., and Stein, L. D. (2003), The SNP Consortium Website: past, present and future, *Nucleic Acids Res.* **31**:124–127.
- [15] Altman, R. B. and Klein, T. E. (2002), Challenges for biomedical informatics and pharmacogenomics, *Annu. Rev. Pharmacol. Toxicol.* **42**:113–133.
- [16] Van Kampen A.H and Homevoets A.J. (2005), The Role of Bioinformatics in Genomic Medicine, New York
- [17] Advanced Bioinformatics and development of Imaging Tools for Cancer Research” 20Biomedicine.htm; Published Jan 2012.
- [18] Ramana Davuluri, PhD: “Advanced Bioinformatics and Bio – Computation (ABBC) Core” institute.htm Jan 2017
- [19] Advanced Bioinformatics: Genetic research: WABR.ORG.htm#overview; Jan 2017
- [20] “Advanced Bioinformatics - Solutions for Gene Expression” Jan 2017
- [21] Garry .G (1986), Computer Literacy: *Applications and Programming*. Design office, San Francisco.
- 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>
- Ibid⁵ <http://www.ncbi.nlm.nih.gov/omim>