Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 5, 2025

APPLICATIONS OF BIOINFORMATIC TOOLS IN THE ANALYSES OF BIOLOGICAL MOLECULES; A REVIEW

Sadaf Ilyas[1], Fiza Laraib[2], Aqsa Mumtaz[2], Muhammad Iltaf[3], Jamal Shah[2], Shahzada Khurram Syed[4], Naveed Munir[*5], Imtiaz Mahmood Tahir[1], Muhammad Riaz[6], Sheikh Ahmed Gull[2]

[1]*College of Allied Health Professionals, Directorate of Medical Sciences, Government College University, Faisalabad, Pakistan*
[2]*Department of Zoology, Faculty of Biological sciences, Quaid-i-Azam University, Islamabad, Pakistan*
[3]*Department of Zoology, Islamia College, Peshawar, Pakistan*
[4]*Department of Clinical Services, School of Health Sciences, University of Management and Technology, Lahore, Pakistan*
[*5]*Department of Biomedical Lab Sciences, School of Health Sciences, University of Management and Technology, Lahore, Pakistan*
[6]*Department of Allied Health Sciences, Sargodha Medical College, University of Sargodha, Sargodha, Pakistan*

DOI: https://doi.org/10.5281/zenodo.15550043

Keywords
Biotechnology, Bioinformatics, Giagnic structure, tools, sequence analysis

Article History
Received on 21 April 2025
Accepted on 21 May 2025
Published on 30 May 2025

Corresponding Author: *
Naveed Munir
naveedmunir215@gmail.com
naveed.munir@umt.edu.pk

Abstract
Biotechnology and Bioinformatics tools are the dire need of managing the huge magnitude of biological data present in databases. To cope with the hour of the need, scientists had to make some advancement and they ended up developing quite a few web-based online and offline tools to analyze the sequence. The pace, by which logical learning is being delivered and shared today, was never been so quick before. Diverse regions of science are getting nearer to each other to give rise new techniques. Bioinformatics is one of such recently developing fields, which makes utilization of computer softwares, science and insights in molecular biology to archive, retrieve, and analyse biological data. Although yet at earliest stages, it has turned out to be one of the quickest developing fields, and immediately settled itself as a necessary segment of any natural research action. It is getting famous because of its capacity to break down gigantic measure of organic information rapidly and cost-adequately. Bioinformatics can help a scientist to extricate significant data from natural information giving different web-or potentially PC based apparatuses, the larger part of which are unreservedly accessible. The present audit gives a thorough synopsis of some of these devices accessible to an existence researcher to investigate organic information. Only this audit will concentrate on those zones of organic research, which can be significantly helped by such devices like investigating a DNA and protein building block to recognize different highlights, expectation of 3D structure of protein atoms, to think about sub-atomic cooperations, and to perform reenactments to impersonate a natural marvel to separate valuable data from the natural information

INTRODUCTION

Bioinformatics is a emerging and developing field in the modern age. It has given many benefits to scientist in their research projects. It provides this ease by using different methods to store, retrieve and

analyses the data either it is of biology or mathematics. Biological data may be of DNA, RNA and protein and all these biomolecules contain unique sequences which are related to their functions. Bioinformatics approaches are emerging day by day which allows scientist to measure the changes and regulation in genome wide genes simultaneously. Scientist has devised many bioinformatics tools to analyses sequence of different nature and all these tools serve different purposes.

Sequence analyses usually refer to collect that information of the biomolecule (nucleic acid and protein) which gives it to its unique function (Hoersch et al. 2000; Hogeweg 2011; Mehmood et al., 2014). Different tools are used for different purposes. Nature of analyses determines which type of tool should be used. Sequences are subjected to these tools which predict its function according to their nature and features which are very much related to biomolecules function. The objective of this review is to encompass most the tools being used for biological sequence analysis which have been described in text and well as summarized in Figure 1 nad Table 1.

2. Bioinformatics Tools For Biological Seq Analysis

2.1. Basic Local Alignment Search Tool (BLAST)

It is sequence similarity search tool which can be use through web to compare users query to databases of sequence. Different types of blast compare combination of nucleotide or protein queries to databases of nucleotide and protein. It usually finds similarity between two sequences and then align these sequence it tells about the alignment score expect value or E value of alignment. BLAST usually calculates statical significance by comparing nucleotide and protein sequences to database sequence. It can be implemented in a number of ways to find out similarity between different sequences (Johnson et al. 2008).

2.2. Clustal Omega

Multiple sequence alignment is very important in bioinformatics because it compares homologous sequences. we have an accurate tool for this purpose known as Clustal Omega. It can align sequence of any size. It has generated alignments of over 190,000 sequences in very few hours. So basically it can align very large number of sequences accurately. It has a number of features of adding new sequences to the existing alignments or adding existing alignments into the new sequences which help to align new sequences of DNA/RNA or protein. It allows user to specify a profile "HMM" which is basically obtained from an alignment of the sequences that are similar to the input set. Then sequences are aligned to these 'external profiles' which help them align to the rest of the input set. This tool has a full confidence on the accuracy of the alignments which are made by using it. That's why it is most widely used tool for multiple seq alignment of nucleotide or protein (Sievers et al. 2011).

2.3. GENSCAN

It is a tool which is used to identify the gene structures in genomic DNA. It predicts the location of exons and introns in the genomic DNA sequences from a variety of organisms. It resolved the fundamental biochemical issues of specifying the accurate sequence determinants of transcription translation and RNA splicing. It also identifies the whole structure of exons and introns in genomic DNA.This tool can also estimate the multiple genes in a sequence to deal with partial plus complete genes and to forecast the consistent set of genes occurring on either or both strands of DNA. It has more accuracy than existing methods, with 70-80% exons identified accurately. It is expected that the statistical analysis of genes may give some clues to the biochemical processes such as transcription, translation and RNA splicing which define genes biologically (Burge et al. 1997).

2.4. "CARNA"—alignment of RNA structure ensembles

Several approaches are use now a day for RNA analysis. These approaches compare the RNA sequences with already predicted or single RNA structure simultaneously. But we need another tool for multiple sequence alignment of RNAs available, where these approaches are not good to use and have limitations. We introduced another tool for the multiple sequence alignment of RNA with riboswitch or pseudoknot structures and this tool is named as "CARNA"this tool supports multiple RNAs with conserved structure and aligns these

pseudoknots originally. So CARNA is basically useful for aligning RNA riboswitch, which have more than one stable structure. We have to put RNA sequence as input and analyze base pair probability matrices and align the sequences on the basis of full ensembles of structures. So this tool is specialized to align RNA with conserved structures (Sorescu et al. 2012).

2.5. "IPknot": IP-based prediction of RNA pseudoKNOTs

It is used for the prediction of RNA secondary structures with pseudokots based on the accuracy of predicted structures. It breaks down the RNA structure with pseudoknot into a set of pseudoknot free structures and predict base pairing probability distribution that considers pseudoknot.It is an IP(integer programming) based prediction tool for RNA having pseudoknot .it is more accurate than previously used IP based methods. It computes the base pair possibilities used in the IP objective function and solve the IP problem to estimate the optimal pseudoknoted RNA secondary structure. It can take a single sequence or aligned sequence as input. Prediction accuracy of IPknot depends on its scoring functions as this method uses approximate possibilities distribution of pseudoknoted structures. Another fact is that IPknot can run quite fast long sequences less than one thousands bases. So it is a fast and accurate computational prediction tool for both single sequence analysis and comparative sequence analysis (Sato et al. 2011).

2.6. "Splign" alignment tool

Local alignment tool like BLAST quickly identify but these tools do not consider the nonaligning intronic segments in the alignment. So these tools are lee precise at splice junctions. To overcome this problem this problem we need a tool that combines the local and global alignment algorithm to produce accurate eukaryotic gene models from transcript. A new NCBI spliced alignment tool have all these features, named as Splign.It calculates cDNA to genome alignments. It gives information about exon-intron boundaries, spliced junctions and possible frame shift. It is power full tool for eukaryotic genome annotation processes and alternative splicing studies. It is gradually developing over the past five years. It can also tolerate sequencing errors and polymorphic sites because it use true optimal alignment algorithm (Kapustin et al 2008).

2.7. PromoSer

It is a web service which provides an ease to retrieve a large number of proximal promoters. It functions to extract promoters of Human, mouse and rat and maps promoter sequence and transcription start sites. In this, you just have to provide list of accession numbers from which promoters are required. You can also provide sequence in FASTA format rather than accession numbers.PromoSer finds clusters which match to the provided sequences. At last PromoSer provides a list of promoters and information about them. It can also identify alternative promoters for each cluster of transcript. It mainly predicts or extracts possible promoters in the sequences which are provided as input (Halees et al 2008).

2.8. Primer 3

It is software which is used to design primers. Designing of primers through primer 3 is an easy and time saving process. And it designs primer (forward and reverse) for PCR technique which is a famous laboratory technique to amplify a fragment of DNA.This program to design primers is becoming extremely with the passage of time. Thermodynamic models and whole genome sequence of many species are improving the functionality of primer 3 to predict primer pair. Discovery of many molecular genetics techniques is bringing scientist to this program as it offers an easy way to design primers artificially. To design primers user have a upload a file of sequences from which you need forward and reverse primers. It gives output within few seconds. The result tells about the primer length, start sequence, GC content, Tm and product size. It also tells about the stability of ANY and END bimolecular interactions and hairpin structures as melting temperature (Untergasser et al 2012).

2.9. Prot Param

Protein ID and examination programming plays out a focal part in the examination of proteins from two-dimensional (2-D) gels and mass spectrometry. For protein recognizable proof, the client coordinates certain exactly obtained data against a protein

database to characterize a protein as definitely referred to or as novel. For protein examination, data in protein databases can be utilized to anticipate certain properties about a protein, which can be helpful for its observational examination. The two procedures are in this manner integral. Despite the fact that there are various projects accessible for those applications, we have built up an arrangement of unique apparatuses on account of a couple of fundamental objectives (Stephan et al 2005). Further, Stephan et al., (2005) described it in particular, these are:

1. To use the broad explanation accessible in the Swiss-Prot database (1) wherever conceivable, specifically the position-particular comment in the Swiss-Prot include tables to consider posttranslational alterations and protein handling.
2. To create apparatuses particularly, however not solely, appropriate to proteins arranged by twodimensional gel electrophoresis and peptide mass fingerprinting tests.
3. To make all apparatuses accessible on the World-Wide Web (WWW), and unreservedly usable by mainstream researchers.

2.10. novoSNP

Confinement site related DNA sequencing (RAD-seq), a cutting edge sequencing innovation, has incredibly encouraged hereditary linkage mapping ponders in outbred species. RAD-seq is fit for finding a huge number of hereditary markers for linkage mapping crosswise over numerous people, and can be connected in species with or without a reference genome. Albeit a few logical apparatuses are accessible for RAD-seq information, elective procedures are vital for enhancing the marker quality and henceforth the hereditary mapping exactness (Robert et al 2005).

2.11. HMMER web server

Well is a product suite for protein arrangement comparability seeks utilizing probabilistic strategies. Already, HMMER has chiefly been accessible just as a computationally concentrated UNIX charge line instrument, confining its utilization. Late advances in the product, HMMER3, have brought about a 100-overlap speed increase with respect to past forms. It is currently doable to make productive profile shrouded Markov show (profile HMM) looks by means of the web. A HMMER web server (http://well) has been outlined and executed to such an extent that most protein database looks return inside a couple of moments. Strategies are accessible for seeking either a solitary protein arrangement, various protein succession king through a protein grouping against Pfam. The web server is intended to take into account a scope of various client skill and acknowledges bunch transferring of numerous questions without a moment's delay. All inquiry strategies are additionally accessible as RESTful web administrations, in this manner enabling them to be promptly coordinated as remotely executed undertakings in privately scripted work processes (Robert et al 2005).

2.11. SEAVIEW AND PHYLO-WIN

It is a tool for comparison that uses CLUSTAL W interface and DOT PLOT approach.

It performs manual editing, manual alignment adjustments and external multiple sequence alignment.

It provides a unique function of creating subsets such as “SETS” and “SPECIES” while examining DNA sequence.

It also has the ability to change the alignment patterns of sequence by adding gaps, deletion, and insertion parallel just by dragging a specific sequence from complete set (Galtier et al., 1996).

2.12. DATAMONKEY2010

It is a phylogenetic analysis tool. It performs functions like

1: Natural selection: The sites in multiple sequence alignment may get deleted so approaches like SLAC, FEL, and REL are used to find them.

Selection at population level is done by FEL.

Selection of specific lineage is done by GA branch.

2: Recombinant detection: appropriate GAs are used to find recombinant DNA. Test like HASEGWA-KISHIN points out the topological differences. SQUEAL module helps finding pol gene of HIV.

3: Co-evolution between sites:(SPIDERMONKEY). It finds out the set of matching areas in alignment using mutations that occur in same sets while evolving (Delport et, al 2010).

2.13. TIGR (TGICL)

It is a tool which analyses the EST and mRNA based on clustering systems of pairwise and individual alignment to produce elongated sequence.

1: A single multi-FASTA file containing cap3 sequence and singletons is taken and run in TGICL, but before this it is cleaned from containments like vectors, adapters ETC using SEQCLEAN. Seed clustering is then done to produce small, purified clusters.

After that it runs MGBLAST for pairwise similarity search. Large databases are also partitioned in a separate file.

Clustering utilities like TCLUST, SCLUST, and NRCL are used for expression of sequences.it is done in graphs where nodes represent sequence and edges represent alignment (PERTEA et al., 2003).

2.14. RSAT

It is a software that is used for detection of CIS-REGULATORY regions in a gene sequence and helps in scanning sequences with (PSSM) and conserved regions are also detected in orthologous gene.

1: Sequence retrieval: non coding sequences present on upstream of gene are retrieved

2: Pattern discovery: OLIGO-analysis tool detects the oligomers on protein sequences

3: Footprint discovery: in this multiple query genes are run and ORTHOLOGS with overrepresented elements on Promoters are discovered (Thomas chollier et, al 2008).

2.15. VECTOR NIT

It is a Well Balanced all in one sequence analysis suite.

It have 5 modules where main modules perform multiple sequence alignment through ALIGN X, or sequence creation and analysis through VECTOR NIT. BIOANOTATORS are used for protein and nucleotide sequence analysis.

FUNCTIONS:

1: MOLECULAR DISPLAY: its primary function is to display and manipulate DNA/RNA and PROTEIN using a three pane system. Text pane contains molecular information, graphic pane displays graphics, and Sequence pane contains the sequence file. These three combine to interpret the results.

2: Molecule construction: plasmid and new design are constructed.

3: Bio annotators: In this graphical representation of DNA/RNA and PROTEIN is done (Stewart et al., 2005).

2.16. SOAP2: short read alignment tool

It is a short oligonucleotide alignment program that reduces memory and increases alignment speed. This tool determines an exact match by constructing harsh table to accelerate searching for location of a read in the BWT index. BWT is a burrowed wheel transformation. It is an alternative of speed algorithm (Li et al., 2009). More bioinformatics tools have been summarized in figure 1 and table 1 below.

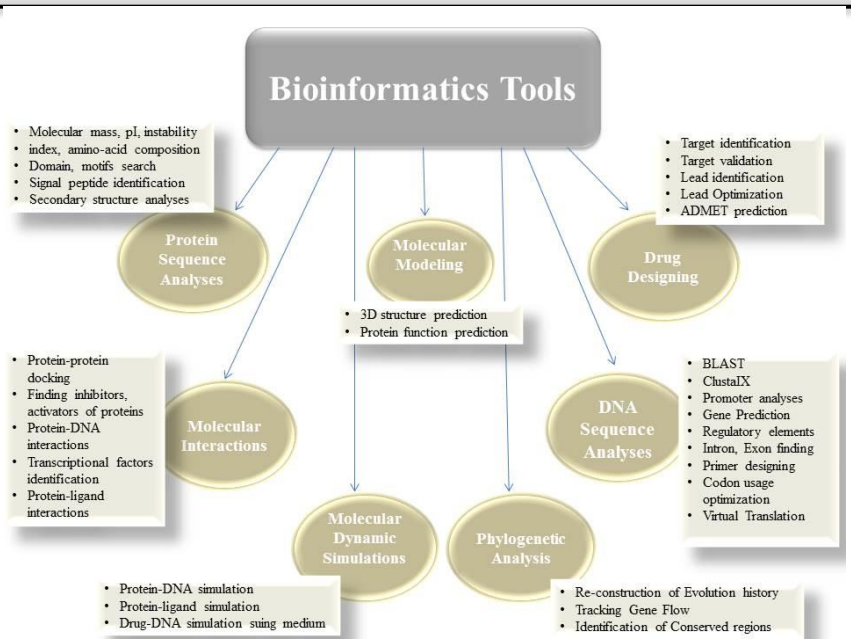

Figure 1: Different types of Bioinformatics Tools and their applications

Table 1: Different types of Bioinformatics Tools and their applications (Summary)

Sr#	Tool name	Function of tool	References
1	**BLAST**	It is use to find out the similarity between sequences and analyze the query sequence.	Johnson et al., 2008
2	**Clustal Omega**	It is used for the multiple sequence alignment and align the sequences based on HMM profile.	Sievers et al., 2011
3	**GENSCAN**	It is used to identify the gene structures in the genome and also predicts the location of exons and introns in genome.	Burge et al., 1997
4	**CARNA**	It is used for the multiple sequence analysis of RNA with riboswitch or pseudoknoted structures.	Sorescu et al., 2012
5	**IPknot**	It is used for the prediction of RNA secondary structures with pseudoknots based on the accuracy of predicted structures.	Sato et al., 2011
6	**Splign**	It calculates cDNA to genome alignments also gives information about exon-intron boundaries and spliced junctions.	Kapustin et al., 2008
7	**PromoSer**	It is a web service which provides an ease to retrieve a large number of proximal promoters.	Halees et al., 2008
8	**Primer 3**	It is software which is used to design primers artificially for different purposes.	Untergasser et al., 2012
9	**SEAVIEW**	Manual editing, manual adjustments, Creation of subsets, Addition of gaps and del+insertion	Galtier et al., 1996
10	**DATAMONKEY**	Natural selection, Recombinant detection Co-evolution between sites	Delport et al., 2010

			datamonkey2010
11	**TIGR(TGICL)**	Pairwise similarity searches Clustering Sequence cleaning Graphical representation	PERTEA et al., 2003
12	**RSAT(REGULATORY SEQUENCE ANALYSIS TOOL)**	Sequence retrieval Pattern discovery Footprinting	Thomas chollier et al., 2008
13	**Vector NTI**	Molecular display Molecular construction Bio-annotators Align-x and align x blocks	Stewart et al., 2005
14	**SOAP2**	Harsh table construction Determine exact match	Li et al., 2009 ,

Conclusion

Bioinformatics is a rapidly growing young discipline which from last few years is progressing surprinsingly. Bioinformatic tools are highly cost effective because these tools predict a lot of informations prior running huge and expensive experiements. However, now a days a vast number of tools to analyse genomes, predicting structures, molecular simulations, proteomes and rational drug designing are available with specific URL but none of them is 'perfect'. So it is the need of current era to develop and find more bioinformatics tools, for this purpose bioinformatics and field of science have to struggle parallel to flourish for the welfare of humanity.

Conflict of Interest

Not conflict of interest

REFERENCES

Hoersch S, Leroy C, Brown NP, Andrade MA, Sander C Trends Biochem Sci. 2000 Jan;25(1):33-5.

N.Galtier1, M. Gouy and C. Gautier, CABIOS 1996 SEA VIEW and PHYLO_ WIN: two graphic tools for Sequence alignment and molecular phylogeny

Wayne Delport1, Art F. Y. Poon2,*BIOINFORMATICS APPLICATIONS NOTE 2010 Vol. 26 no. 19 2010, And doi:10.1093/bioinformatics/btq429Data monkey 2010: a suite of phylogenetic analysis tools for evolutionary biology*

Geo Pertea 1, Xiaoqiu Huang2, and BIOINFORMATICS APPLICATIONS NOTE Vol. 19 no. 5 2003, DOI: 10.1093/bioinformatics/btg034 TIGR Gene Indices clustering tools (TGICL)

Morgan Thomas-Chollier, Olivier Sand, Jean-Vale´ RY , Nucleic Acids Research, 2008, Vol. 36 doi:10.1093/Nar/gkn304
RSAT: regulatory sequence analysis tools

HENRY STEWART PUBLICATIONS 1467-5463. BRIEFINGS IN BIOINFORMATICS. VOL 5. NO 4. 378–388. DECEMBER 2004 Vector NTI, a balanced all-in-one sequence analysis suite

Ruiqiang Li1, Chang Yu1,†, Yingrui Li ,*BIOINFORMATICS APPLICATIONS NOTE , Vol. 25 no. 15 2009, doi:10.1093/bioinformatics/btp336*

Hogeweg P. PLoS Comput Biol. 2011 Mar;7(3):

Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics Methods and Protocols. Totowa, NJ: Humana Press; 2000. pp. 365–386

Mark Johnson Irena Zaretskaya Yan Raytselis Yuri Merezhuk Scott McGinnisThomas L. Madden *Nucleic Acids Research*, Volume 36, Issue suppl_2, 1 July 2008, Pages W5–W9,

Fabian Sievers, Andreas Wilm, David Dineen, Toby J Gibson, Kevin Karplus, Weizhong Li,[5] Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, Julie D Thompson, and Desmond G Higgins Mol Syst Biol. 2011; 7: 539.

Burge C[1], Karlin S. J Mol Biol. 1997 Apr 25;268(1):78-94.

Dragoş Alexandru Sorescu, Mathias Möhl,[1] Martin Mann, Rolf Backofen, and Sebastian Will Nucleic Acids Res. 2012 Jul; 40(Web Server issue): W49–W53.

Kengo Sato,[,] Yuki Kato,[,] Michiaki Hamada, Tatsuya Akutsu, and Kiyoshi Asai Bioinformatics. 2011 Jul 1; 27(13): i85–i93.

Yuri Kapustin, Alexander Souvorov, Tatiana Tatusova, and David Lipman Biol Direct. 2008; 3: 20.

Anason S. Halees and Zhiping Weng Nucleic Acids Res. 2004 Jul 1; 32(Web Server issue): W191–W194.

Yuki Kato, Kengo Sato, Kiyoshi Asai, and Tatsuya Akutsu Nucleic Acids Res. 2012 Jul; 40(Web Server issue): W29–W34

Sugimoto, Masahiro; Kawakami, Masato; Robert, Martin; Soga, Tomoyoshi; Tomita, Masaru Current Bioinformatics, Volume 7, Number 1, March 2012, pp. 96-108(13)

Andreas Untergasser, Ioana Cutcutache, Triinu Koressaar, Jian Ye, Brant C. Faircloth, Maido Remm, and Steven G. Rozen Nucleic Acids Res. 2012 Aug; 40(15): e115.

Stefan Weckx, Jurgen Del-Favero,[1] Rosa Rademakers, Lieve Claes, Marc Cruts, Peter De Jonghe, Christine Van Broeckhoven, and Peter De Rijk. Genome Res. 2005 Mar; 15(3): 436–442.

Robert D. Finn,[*] Jody Clements, and Sean R. EddyGenome Res. 2005 Mar; 15(3): 436–442..

Mehmood, M. A., Sehar, U., & Ahmad, N. (2014). Use of bioinformatics tools in different spheres of life sciences. Journal of Data Mining in Genomics & Proteomics, 5(2), 1.