ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

Gensolution: Website through Bioinformatics Approach

Dr. Mohammad Alamgeer

Assistant Professor, Department of Information Systems, King Khalid University, Kingdom of Saudi Arabia (KSA)

ABSTRACT: It is difficult to recall that only in 1953 was the famous double helical structure of DNA determined. Since then a stupendous series of discoveries has been made. The unraveling of the genetic code was only the beginning. Learning the details of genes and their discontinuous nature in eukaryotic genomes like ours has led to the ability to study and manipulate the material of that abstract concept of Mendel's, the gene itself. Learning to read the genetic material more and more rapidly has enabled us to attempt to decode entire genomes [1, 2].

The rate of innovation in molecular biology is breathing. The accumulation of data has necessitated international databases for nucleic acids, for proteins, and for individual organisms and even chromosomes. The crudest measure of progress, the size of nucleic acid databases, has an exponential growth rate. Consequently, a new subject or, if that is too grand, a new area of expertise is being created, combining the biological and information sciences called Bioinformatics - Marriage of Biology and Computer Science [3, 4, 5].

Bioinformatics represents a new, growing area of science that was computational approaches to answer biological questions. Answering these questions requires that investigators take advantage of large, complex data sets in a rigorous fashion to reach valid, biological conclusions. The potential of such an approach is beginning to change the fundamental way in which basic science is done. The term Bioinformatics is relatively new, and as defined here, it encroaches on such term as "Computational Biology" and others [6, 7].

Increasingly Bioinformaticians are interested in developing analytical tools and database that help scientists interpret experimental data especially in the content of biological systems. Such analytical tools have broad application throughout R & D, from validating targets by uncovering disease related genes and pathways to predicting pathways perturbed by therapeutic compounds.

The GenSolution, supports query search on Inflammatory Genes and PolyHydroxyAlkanoates biosynthetic genes by maintaining a comprehensive, non-redundant, well organized and freely available database. Presently the entries in this database are clustered in proper hierarchy. GenSolution also supports algorithms and tools for Motif Search, Sequence Comparison, Longest common sub-sequence finder, Promoter mapping of AlgT (ECF subfamily), Restriction Mapping, Open Reading Frame finder, GenCalculator, and Parser.

KEYWORDS: Bioinformatics, Computational Biology, Genetics, Molecular Biology, GenSolution, Inflammation, PolyHydroxyAlkanoates, Motif Search, Sequence Comparison, Longest common subsequence, Promoter mapping, Restriction Mapping, Open Reading Frame, Parser

I. OVERVIEW

Computational biology is a highly interdisciplinary field of biology, relying on basic principles from computer science, biology, physics, chemistry, mathematics, and statistics. Bioinformatics, a popular term in this era of large-scale DNA sequencing, is only a subset of computational biology - the part concerned with the storage, organization, curation and annotation of biological data.

Computational biology extends beyond bioinformatics into the realm of sequence analysis: finding genes and ascertaining their function; predicting the structure of proteins and RNA sequences; and determining the evolutionary

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

relationship of proteins and DNA sequences. Computational method of molecular biology deals development of database for extracted genomic information and implementation of tools & algorithms to solve the biological problems. By keeping all such type of scientific requirements and computational biology research work, the website ***GenSolution*** is designed which is freely accessible at "http://www.gensolution.org" from any internet portal worldwide. At present, tools and databases which are designed and developed are: Motif Search (Motif finding, Splice junction, Promoter mapping), Sequence Comparison (Score matrix, Sequence alignment, Longest common sub-sequence finder), Restriction Mapping (Single digest, Double digest) & Open Reading Frame finder, GenCalculator (Base pair determination, G+C percentage calculator, Hydrophobicity 2D-HP), Database (Inflammatory genes database, Polyhydroxyalkanoates), Parser (GenBank parser), and some more such as Nucleotide to Amino acid translator). This also provided link of other web portal related to Bioinformatics databases, Biological Databank & Bioinformatics tools and some more which is beneficial to the users. Few algorithms which developed for these tools are of NP complete and NP hard type.

The databases of inflammatory genes & pathways involve narrowing down all the genes responsible for inflammation on the basis of microarray experiment analysis and pathway mapping of the short listed genes in existing metabolic / regulatory pathway.

The database of Polyhydroxyalkanoates is designed to hold genes and relevant keys value which representing genomic characterization. These genes are responsible for synthesize biodegradable plastics.

Algorithm of longest common subsequence in multiple sequences and promoter mapping for ALgT (ECF sub-family of sigma factors) are NP complete type.

All applications of GenSolution are easily accessible at http://www.gensolution.org (right now merged in http://www.az-group.org). All comments, queries and suggestion should be sent by email to gensolution@az-group.org. For this, the website provides feedback and contacts link on home page.

II. OBJECTIVE

The development of GenSolution (website through bioinformatics approach) involves designing and creation of tools for solving the biological problems related to motif finding, sequence comparison and statistical solution of nucleotide and amino acid sequences. Database development is for storing extracted information of genes of Polyhydroxyalkanoates and Inflammatory genes & pathways by using relational data model. In addition, designing of user-friendly web pages would help in data manipulation information extraction through web technology. The main emphasis would be to develop algorithm of Longest Common Subsequence Problem and Promotor mapping for ALgT (ECF sub-family of sigma factors) which was NP complete problem. The website provides easy access of database and generates desire output in specific format.

Inflammatory Genes & Pathway database of GenSolution is a database of all inflammatory genes and pathways responsible for inflammatory action in Human, Mouse and Rat. Another database of website is of Polyhydroxyalkanoates which holds all the information including DNA & Amino acid sequences of those microbial genes, which are responsible for biodegradable plastic synthesis.

The GenSolution also providing links of other bioinformatics resources which are helpful for Bioinformatician to link with various commonly usable tools and databases. The GenSolution home page has drop down GUI based menu which makes the website easy accessible. Few features of GenSolution are protected from unauthorized people. To access such types of restricted features, author provides rights to access particular application with providing user id and password.

III. GENSOLUTION - WEBSITE THROUGH BIOINFORMATICS APPROACH

The entitled website – **GenSolution** is a website which developed through bioinformatics approach. Basically it provides Bioinformatics tools to solve basic biological problems. Out of these tools, two programs are new and was NP

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

– Complete problem, developed by the author and are publically accessible through internet. GenSolution website provides two large databases which are designed and developed after normalization of data through Bioinformatics technique. The entire Client/Server model of database provides two end programming, one for user access and another for database manipulation.

Tools and Databases:

3.1. Motif Search:

One of the most common things we do in bioinformatics is to look for motifs, short segments of DNA or protein that are of particular interest. They may be regulatory elements of DNA or short stretches of protein that is conserved across many species. The motifs we look for in biological sequences are usually not one specific sequence.

3.1.1. Motif Finding:

It is a tool which accepts string and motif which may be regulatory elements of entered DNA or short stretches of entered protein string. The options - case sensitive and remove space are to make case sensitive search and to remove the blank space present in entered string.

If searched motif is present in entered sequence, then it displays – how many times and at which positions (starting position of pattern) motif is present, else display message – motif not found.

3.1.2. Splice Junction:

This program automates splice junction identification. All locations where "GT" occurs in the string are noted and the probability of a splice junction occurring there is computed from the preceding nucleotides. The potential splice junctions are printed in order from most to least likely.

It accepts nucleotide sequence. After submitting a sequence, it displays sequence with indicating all positions where splice junction(s) is(are) present. It also displays associated probabilities of all potential splice junctions.

3.1.3. Promoter Mapping - AlgT (ECF sub family of sigma factors):

Pseudomonas aeruginosa is an opportunistic pathogen that causes chronic infections in Cystic Fibrosis patients. Frequent nosocomial infections are caused by this pathogen. Many clinical isolates in particular from Cystic fibrosis patients exhibit a mucoid phenotype. This is due to copious production of the polysaccharide alginate. Alginate is an important virulence determinant for Pseudomonas aeruginosa. It is believed that it inhibits phagocytosis and potentially limits antibiotic efficacy due to limited antibiotic penetration. We have been interested in understanding the regulation and production of alginate in P. aeruginosa. AlgT is a member of the ECF subfamily of sigma factors. It has been shown to control expression from the 18 kb biosynthetic operon for alginate. Members of the ECF family of sigma factors exist in a diverse group of organisms where they respond to various forms of extra cytoplasmic stimuli. Here I present a proteomic analysis to examine the regulon for AlgT. We have identified and demonstrated that dsbA is under transcriptional control of AlgT. I present here characterization of this gene and additional potential components of the AlgT regulon.

It accepts nucleotide sequence in text area. If promoter for AlgT present in entered sequence then it displays desire result.

3.2. Sequence Comparison:

Mutation in DNA is a natural evolutionary process: DNA replication errors cause substitutions, insertions, and deletions of nucleotides, leading to "editing" of DNA texts. Similarity between DNA sequences can be a clue to common evolutionary origin (as with the similarity between globin genes in humans and chimpanzees) or a clue to common function (as with the v-sys oncogene and the growth-stimulating hormones).

3.2.1. Scoring Matrices:

The scoring matrix is used to score each nongap position in the alighnment. For nucleotide sequence alignments, generally Identity, BLAST or Transition Transversion scoring matrices are used. BLAS - a commonly used tool for aligning and searching nucleotide sequences is very simple matrix that assigns a score of +5 if the two aligned

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

nucleotides are identical, and -4 otherwise. Similarly, the Identity matrix assigns a score of 1 and 0 for identical and non-identical aligned nucleotides respectively. The Transition Transversion matrix provides a mild reward for matching nucleotides, a mild penalty for transions-substitutions in which a purine (A or G) is replaced with another purine or a pyrimidine (C or T) replaced another pyrimidine and a more severe penalty for transversions, in which a purine is replaced with a pyrimidine (C or T) or vice versa.

GenSolution provides Scoring Matrices program which accepts two sequences of nucleotides (for Identity, BLAST, or Transition Transversion matix) or amino acids (for PAM or BLOSUM matrix) and generates one matrix of type Identity / BLAST / Transition Transversion / PAM / BLOSUM in 2-D form with best score value.

3.2.2. Two Sequence Alignment:
GenSolution provides tools for Local and Global alignment of two sequences. The local alignment tool also provide BLAST and FASTA based algorithm. It is a program which accepts two sequences and users need to select types of alignment (Local alignment, BLAST - Ungapped local alignment, FASTA-Gapped local alignment, Global alignment). After submitting these parameters (values) it generates desired result.

3.2.3. Longest Common Subsequence:
A simple dynamic programming algorithm to compute common subsequence between two strings has been discovered independently by many authors. But there was no solution to compute LCS in multiple sequences.

Website author aggregate GenSolution – LCS recursive program using a specific technique. In this technique, once pattern of a subsequence not matches with sub-parts of any sequence, escape this subsequence and select next possible subsequence.

It accepts list of sequences (atleast two) in specific format in text area. Each input sequence starts with '>Sequence #' as first line and terminated with '//' symbol as last line. After submitting list of sequences, it generates longest common subsequence present in all entered sequences with indicating position in each sequence.
The result page of longest common subsequence program also provides link to generate ORF and protein sequence for all entered sequences.

3.3. Restriction Map & ORF:
3.3.1. Restriction Maps:
Computing restriction maps is a common and practical bioinformatics calculation in the laboratory. Restriction maps are computed to plan experiments, to find the best way to cut DNA to insert a gene, to makes a site-specific mutation, or for several other applications of recombinant DNA techniques.

3.3.1.1. Single Digest:
In single digest, entered sequence can be cut by one restriction enzyme. If the restriction site(s) of selected enzyme is(are) present in entered sequence, then it cuts the sequence from all positions of restriction site and displays all fragments after cutting. The result page also shows the exact position from where each fragment obtains after action of restriction site.

3.3.1.2. Double Digest:
Just like single digest, it accepts nucleotide sequence but it provides option to select two different restriction enzymes from drop down list. After submitting, program searches restriction sites for both enzymes. If sites present for either one or both enzymes, it displays appropriate result otherwise it displays message.

3.3.2. Open Reading Frame (ORF):
The biologist knows that, given a sequence of DNA, it is necessary to examine all six reading frames of the DNA to find the coding regions the cell uses to make proteins. If you don't know where the translation stats, you have to consider the six possible reading frames. Since the codons are three bases long, the translation happens in three "frames", for instance starting at the first base, or the second, or perhaps the third. The fourth would be the same as starting from the first.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

It is therefore quite common to examine all six reading frames of a DNA sequence and to look at the resulting protein translations for long stretches of amino acids that lack stop codons. The stop codons are definite breaks in the DNA to Protein translation process. During translation (actually of RNA to protein), if a stop codon is reached, the translation stops, and the growing peptide chain grows no more. Long stretches of DNA that don't contain any stop codons are called open reading frames (ORFs).
GenSolution provides tool to make ORF of entered sequence. After submitting a nucleotide sequence, it displays all possible ORFs present in 1 to 6 (first 3 from original sequence and last three from its complimentary sequence). If more than one start to stop regions are found in any reading frame, it shows all fragments is result page.

3.4. Gen Calculator:

3.4.1. Base Pair Determination:

It is a program which used to count total number of individual characters either nucleotide or amino acids present in entered respective sequence. From user it accepts sequence of nucleotide or protein and asks to select sequence type from drop down option. After submitting these parametric values, user gets desire result.

3.4.2. G + C percentage calculator:

It is a program which is use to count G+C contents of entered nucleotide sequences. It accepts one or more than one nucleotide sequence. If user enter more than one sequence, then each sequence must be starts with symbol '>' followed by Sequence name as first line and end with a line of symbol '//'. After submitting sequence(s) it will displays GC1 (G+C content of first codon position), GC2 (G+C content of second codon position), GC3 (G+C content of third codon position), and GCT (total G+C contents) counts for each sequence.

3.4.3. Hydrophobicity (2D – HP):

One major simplification that can be made to the process of protein folding is to consider only hydrophobic interactions. Taking thing a step further, suppose that all amino acids must fall onto the intersections of equally spaced lines in a grid. For a further simplification, consider only a two-dimensional grid, because it is easy to draw examples for such a grid.

The tool of GenSolution reads an entered sequence consisting of classifications of amino acids and directions to get t the location of the next amino acid.

3.5. Databases:

3.5.1. Inflammatory Genes & Pathways Database:

The databases of inflammatory genes & pathways involve narrowing down all the genes responsible for inflammation on the basis of microarray experiment analysis and pathway mapping of the short listed genes in existing metabolic / regulatory pathway. This is a superset database for inflammatory pathways and genes. The entire list of genes has been narrowed down from whole genome list of Human, Mouse and Rat on the basis of microarray experiment result and gene's description. Two global databases (KEGG & Biocarta) have a collection of pathways. The NCBI database resource helps to collect all information of genes. To fetch the query result from database, three major independent modules (Single Gene Query Search, Multi Gene Query Search, and Pathway Model Search) are designed which are useful to analyze the microarray data and genes involved in inflammatory pathways.

All pathways have collection of genes, clustered on the basis of similarity of its biological and chemical properties (downloaded from global databases). And each gene on pathway is with hyperlink. On clicking this gene, user can know details of same gene. Each result page of query search contains hyperlink of global database (NCBI, Biocarta and KEGG), by clicking on which the user can jump on this database for resultant gene and finally will get present time updated result. All modules are structurally tested on design time and the final system is functionally tested.

3.5.2. Polyhydroxyalkanoates Database:

Purpose of developing such a database is to establish the database of list of organisms and list of genes and its related physical and chemical properties of genes which are able to synthesize biodegradable plastic. This database becomes helpful for scientists who are working on such biodegradable plastic synthetic project.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

Polyhydroxyalkanoates Database is a single repository of genes and its genomic information responsible for Polyhydroxyalkanoates to synthesize biodegradable plastics. It is based on Genomic characterization of intermediates of Polyhydroxyalkanoates (CAB genes, responsible for biodegradable plastic synthesis) metabolic pathway. To fetch the query result from database, two major independent modules are designed which are useful to analyze the CAB genes of organism.

To access database, the web page is designed which is good to look at, user friendly and completely validated which provide easy to query search on database.
This home page of Polyhydroxyalkanoates database provides facilities to query search by entering search key of organism's name or by clicking on alphabetic list of organisms / taxons.

3.6. Parser:
Parser is a program or technique through which we can extract or parse different annotation or specific information from lengthy flat file sequence. Different bioinformatics existing databases generate the result in specific format and user needs to extract required information from huge collection of lines. Common file formats of bioinformatics database are GenBank, PDB and BLAST from which we need to parse annotations and sequence.

3.6.1. GenBank Parser:
GenBank flatfile (GBFF) is the elementary unit of information in the GenBank database. It is one of the most commonly used formats in the representation of biological sequences. GBFF is separated into three parts: the header, which contains the information that apply to the whole record; the features, which are the annotations on the record; and the nucleotide sequence itself.

GenSolution provides a tool - GenBank Parser, to prompt GenBank flatfile of GenBank database which after submission displays the all annotations and parametric value checked by the user. Here we can filter Locus, Definition, Accession, Version, Source, Organism, and references from header part; entire features or Source detail, Gene, cDNA, A.Acid sequence form features table; and Origin (Nucleotide sequence).

Parsing of annotations and sequence from GenBank flatfile concept of Regular expression is used. The pattern modifiers present on flatfile is helpful to solve such types of parser.

IV. VALIDATION

Each program on GenSolution is validated which accepts input for query in specified format. Most of the programs provide links which is helpful for end users to access the tools and database properly. It provides the confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled.

Before the calling of any event's action, the code of programs (modules) first checks the conditions, if satisfied then perform the command action otherwise give message to the user. In entire project development, most of the validations have been used in Client side programs (by using JavaScript) and some are in server side programs to validate the data in server database (by using ASP). Except some, most of the event's actions are with full validation, which provides quality of tool and better architecture model for the organization.

V. SECURITY

In this GenSolution, the Database is developed by using MS SQL. To protect such types of Database, more protection afford is required. Some features of GenSolution are restricted from unauthorized access. To access such types of features, author provides User ID and Password.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

VI. ADVANTAGES

- The GenSolution providing multiple solutions at single platform linked with home page of single web address. Entire development work is through Bioinformatics approach (solutions of biological problems through IT).
- LCS in multiple sequences and Promoter mapping for AlgT are golden solution done by the author which was NP complete problems.
- Motif search, Sequence alignment, GenBank parser, Sequence Translator, ORF finder, Restriction map, etc are tools to solve many problems rises during data analysis study.
- Documentation and abbreviations make easier to access. Link for many Bioinformatics websites make it popular by researcher and students.
- Splice junction, Hydrophobicity, G+C content and Base pairs counter are tools of GenSolution which helps in statistical calculation of biological nucleotide sequences.
- The GenSolution has two databases which are useful for R&D research in Pharmaceutical and Environmental Genomics organizations.
- Inspite of the complexity of algorithms, execution is fast.
- To established interface, the entire ASP programs and modules are based on DSN-less mode. Such model provides robustness to the entire website model.
- Author accepts feedback from users and welcomes suggestions for any error, incorrectness and for further improvement.

VII.FUTURE OF GENSOLUTION

- In coming days author has plan do make this website research oriented in which will help Biotechnologist, Pharmaceuticians, Chemists and Bioinformatician through Bioinformatics approach of Information technology.
- GenSolution's author has plan to develop few more biostatistical tools such as for calculation of atomic density, codon adaptation index (CAI) value, Chi squire test, etc.
- Author also developing a large database of genes of medicinal plants.
- Author also trying to develop tools for Multiple Sequence Alignment.
- In coming time, author will try to develop algorithm for Traveling Salesman Problem (NP Complete) based on Artificial neural network.
- Users will get more useful link on GenSolution home page time to time which will help for Bioinformatics work.

VIII. AVAILABILITY

All the tools and databases available at GenSolution are freely accessible at http://www.gensolution.org or http://www.az-group.org/GenSolution/home.asp or GenSolution link available on http://az-group.org home page. All comments, queries and correction should be sent by email to dralamgeer@az-group.org. For this, the website provides feedback and contacts link on home page.

IX. CONCLUSION

The website "GenSolution …through bioinformatics approach" is developed to outline author's research, and provide links to resources and other information related to genomics, biotechnology, and bioinformatics. The entire research work is encapsulated under the name of GenSolution which are - Sequence analysis and Motif finding; Sequence translator and ORF finder; Alignment; GenBank Parser; Base Pair, G+C content and Hydrophobocity Calculator; Algorithm of LCS and Promotor mapping of ALgT (ECF sub family of sigma factor); Database of inflammatory genes & pathways and Polyhydroxyalkanoates, etc.

The tools - motif finder, splice junction finder, and promoter mapper for AlgT are based on pattern matching and are helpful in respective aspects to the genomics analysts. Related to sequence compression - score matrix, sequence alignment, and longest common subsequence finder are helpful for compression of nucleotide sequences. Restriction mapping program (promoter mapping of single and double digest) helps to cut nucleotide sequences with different restriction enzymes to get fragments of sequence in dry lab. Through open reading frame user can get all possible

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

frames of entered gene. gencalculator of GenSolution provides tools for base pair count, g+c contents counter, hydrophobicity of nucleotide sequence, etc to perform different statistical operations on nucleotides and amino acids. GenBank parser helps to parse different annotations of gene bank flatfile. There is a tool to translate nucleotide sequence into amino acid sequence.

The databases of inflammatory genes & pathways involve narrowing down all the genes responsible for inflammation on the basis of Microarray experiment analysis and pathway mapping of the short listed genes in existing metabolic / regulatory pathway. This is a superset database for Inflammatory pathways and genes. The entire list of genes has been narrowed down from whole genome list of Human, Mouse and Rat on the basis of microarray experiment result and gene's description. Two global databases (KEGG & Biocarta) have a collection of pathways. The NCBI database resource helps to collect all information of genes. To fetch the query result from database, three major independent modules are designed which are useful to analyze the microarray data and genes involved in inflammatory pathways.
The database of Polyhydroxyalkanoates is designed to hold genes responsible for Polyhydroxyalkanoates to synthesize Biodegradable plastics. It is based on Genomic characterization of intermediates of polyhydroxyalkanoates (CAB genes, responsible for Biodegradable plastic synthesis) metabolic pathway. To fetch the query result from database, two major independent modules are designed which are useful to analyze the CAB genes of organism.

At GenSolution collection of various web resources are also linked on a single website. Website of GenSolution provides collection of links for Biological databank and Bioinformatics tools which is helpful to the Bioinformatician and Bioinformatics students. All the programs of search tool and database manipulation are with full validation. Inspite of complexity of algorithm, the execution is fast.

X. ACKNOWLEDGEMENTS

The author thanks Dr. Moinuddin Khan, Dr. Abdul Ilah, Dr. D.C.Upadhaya, Professor S. I. Ahson, Mr. Sayed Zeeshan Hussain, and Dr. Kamal Raval for discussions and encouragements. The author also thanks Dr. Kulvinder Singh Saini and Dr. V.C. Kalia for giving me opportunity to realize such types of biological problem during the research training at Ranbaxy and IGIB research center respectively. The author deeply indebted to Honorable Shri D.C. Singhania (Chancellor, Singhania University) and Shri Rajkumar Yadav (Chairman of Singhania University), for giving me opportunity to work and avail the facilities of the Bioinformatics center of University Campus. Author also thanks to parents, family members, relatives, villagers, and friends for their support, freedom and motivation.

REFERENCES

1. Achuthsankar S Nair Computational Biology & Bioinformatics - A gentle Overview, Communications of Computer Society of India, January 2007
2. Aluru, Srinivas, ed. *Handbook of Computational Molecular Biology*. Chapman & Hall/Crc, 2006. ISBN 1584884061 (Chapman & Hall/Crc Computer and Information Science Series)
3. Baldi, P and Brunak, S, *Bioinformatics: The Machine Learning Approach*, 2nd edition. MIT Press, 2001. ISBN 0-262-02506-X
4. Barnes, M.R. and Gray, I.C., eds., *Bioinformatics for Geneticists*, first edition. Wiley, 2003. ISBN 0-470-84394-2
5. Baxevanis, A.D. and Ouellette, B.F.F., eds., *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, third edition. Wiley, 2005. ISBN 0-471-47878-4
6. Baxevanis, A.D., Petsko, G.A., Stein, L.D., and Stormo, G.D., eds., *Current Protocols in Bioinformatics*. Wiley, 2007. ISBN 0-471-25093-7
7. Claverie, J.M. and C. Notredame, *Bioinformatics for Dummies*. Wiley, 2003. ISBN 0-7645-1696-5