Journal of Computational Intelligence in Bioinformatics
ISSN: 0973-385X Volume 3 Number 1 (2010) pp. 41–48
© Research India Publications
http://www.ripublication.com/jcib.htm

Efficacy of Information Technology in Biotechnology and Bioinformatics

[1]Suman Deep Singh, [2]Kulpreet Kaur and [3]Kamal Kishore

[1]Jyoti Institute of Information Technology, Bareilly-243006 (Uttar Pradesh), India
[2]S R International Public School, Bareilly-243006 (Uttar Pradesh), India
[3]Associate Professor, Department of Pharmacy, M.J.P. Rohilkhand University, Bareilly-243006 (Uttar Pradesh), India
[3]Corresponding Author E-mail: kamalbareilly@yahoo.co.in

Abstract

In this era, biotechnology is a novel prospect for entrepreneur. Now scientists have make-believe where software companies sketch to revive their businesses writing algorithms for the human genome. Biotechnology companies plan to expand bioinformatics as an addition of their businesses. Bioinformatics is the association among computational and biological sciences. It is a new forte in biotechnology. However, bioinformatics is the liaison connecting computation and genetic sciences. The competence to sort and extract genetic codes from human genome database of 3 billion found pairs of DNA in a momentous technique is conceivably an uncomplicated form of bioinformatics. It is useful in mapping diverse human beings genomes, deriving difference in their inherent composition through prototype detection software. The feature what becomes more complex is to decode the genetic code to distinguish the differences in genetic framework between different people to translate it in terms of physiological traits. Also so far there is one more stage, which is even more elaborate that is the set of codes itself. The genetic code, essentially codes for amino acids and proteins and the explicit role played by every protein controls the state of our health. The function of each gene in coding for a definite protein, which regularizes a particular metabolic path, is explained as “functional genomes”. Computational stimulation of experimental biology is another important application of informatics. It is recognized as “in silico” testing. The development of this technique will take place in a productive way, both in animals and humans, to verify the certainty of medical behavior. It provides a break towards increasing insensitivity towards animal experimentation. On the other hand, research strategies will certainty come across computational representation to be a fundamental tool in ample

momentum to researcher with enormous cost benefits. Thus, bioinformatics offers to the researcher vast and exhilarating opportunities to learn how living organism metabolize, develop, fight disease, their reproduction and regeneration. All existing facts we know are like tips of various icebergs seen submerged in deep oceans. Gradually an encyclopedia is building which will help to unravel the mysteries of life. The significance of computational science in accumulating the information and its immediate analysis by biologists is the major philosophy of bioinformatics. A range of scholars and investigator professionally explain the numerous elements of bioinformatics to facilitate aspirant to figure out the ins and outs of this contemporary sphere of influence in their detection of strong probability.

Introduction

The expression bioinformatics has been confiscated by several unusual disciplines to mean rather diverse things. In its extensive intellect, the expression can be considered to mean information technology functional to the management and analysis of biological data. This has implications in diverse areas, ranging from Artificial Intelligence and robotics to Genome analysis. In the context of Genome initiatives, the expression was originally functional to the computational operation and analysis of biological sequence data (DNA or protein). At the beginning of 1998 available databases more than 300000 protein sequences have been deposited [1]. Information technology plays a vigorous role in the field of Bioinformatics, which is up-and-coming as one of the most extraordinary areas in advent years. This trend is transmittable, because bioinformatics is not directly linked to biology as compared to statistical computing. Bioinformatics has expanded its significance and recognition after the formulation of Human Genome Project in public. Bioinformatics was found in 1970. Due to informatics techniques, it works out complex problems that are increasingly faced in molecular biology and biotechnology. Bioinformatics has progressed as regulation due to the power of contemporary computational apparatus subsequent from arithmetic, statistical information, and computer discipline. Bioinformatics alias ‘in-silico’ biology, is biology in expressions of molecules and be relevant informatics techniques to fabricate up software mechanism, databases and management information systems to obtain and accumulate biological data molecules on a hefty range. Actually, the first signal of informatics has been focusing on data storage, data access and revealing the primary DNA sequence. Bioinformatics is meticulously computer oriented. Progress in DNA sequencing technology has created the demand for significant quantity of sequence data.

Information Networks

The internet is a global network of computer networks that links government, educational and business institutions. In order to work competently networks share a communication protocol, called Transmission Control Protocol/Internet Protocol, known as TCP/IP. In the internet, the mainly recognizable services comprise

electronic mail, news groups, file transfer, and distant computing. Software tools can assist the biologist to foresee the metabolic trail to comprehend the proportional, architectural, serviceable, and evolutionary role of genetic factors. As it necessitates making obtainable work for an array of techniques of sequence analysis, software packages have developed to bring numerous methods simultaneously. Gizmo's for searching all the key sequence databases, BLAST - Basic Local Alignment Search Tool, Software packages, SWORDS - Statistical Software for Analyzing Words in a DNA sequence. CINEMA - Color Interactive Editor for Multiple Alignment is a java configuration editor that integrates services for database searching, construction visualization etc. In fact, probability theory has contributed considerably to bioinformatics. Individual has to depend typically on widely accessible databases rather than any database. Oracle has also included particular database tools exclusively to beat the Bioinformatics market.

The European Molecular Biology network (EMBnet)-The internet is energy for worldwide communication and reserve centralization. This became mainly vital in the mid-1980s as biological databases were starting to flourish and users were challenging further competent way of contact to more modern data. In 1988, a network was recognized to association European laboratories that use bio computing and bioinformatics in molecular biology research. The network, identified as EMBnet, predicted as a way of providing information. The enterprise of such centralized countrywide services was a vital pace, as the ground of computational environmental science expanded; it is possible for entity institutions to remain modern copies of a variety of biological databases. EMBnet work 34 nodes, 20 are intended National Nodes. These are allotted through the governments of their particular nations, and contain a consent to make available databases, software and on-line services (including sequence analysis, protein modeling, genetic mapping etc.) to recommend user support and working out, and to take on Research and Development, such as the design of the Sequence Retrieval System (SRS). These are educational, trade research centers, which considered having scrupulous data of precise areas of bioinformatics. They are mainly accountable for the upholding of biological databases and software: e.g., the EBI maintains the EMBL nucleotide database, the ICGEB maintains the SBASE annotated domain database, and so on [2].

The UK MRC Human Genome Mapping Project-Resource Center-The UK Medical Research Council funds the HGMP-RC. It is lively in the Human and Mouse Genome Projects, playing a task as equally as resources and service giver. It moreover provides an online computing service.

The European Bioinformatics Institute (EBI)-Recognized in 1994, the European Bioinformatics Institute is an outstation of the European Molecular Biology Laboratory (EMBL), an intercontinental research institute with its headquarters in Heidelberg, Germany. The vital activities of the EBI are expansion and sharing of the EMBL Nucleotide Sequence database [3].

The Martinsried Institute for Protein Sequencing (MIPS)-IT is the European associate of the International Protein Sequence Database. The task of MIPS is to gather, allocate, and sustain modern protein sequence data within Europe [4].

University college of Londan (UCL)-The Biomolecular Structure and Modeling unit at University College London is a biocomputing center with proficiency in two central areas of bioinformatics, mainly in protein sequence and structure analysis. Programs are available for direct use via web, with a range of database search and sequence analysis tools, sequence alignment and structure visualization software. There are servers which provide these facilities, the DbBrowser Bioinformatics Web Server [5].

The National Center for Biotechnology Information (NCBI)-American information provider, the National Center for Biotechnology Information, was established in 1988 as a division of the National Library of Medicine (NLM), and is situated on campus of the National Institutes of Health (NIH) in Bethesda, Maryland. The NLM was selected to host the NCBI because of its experience in biomedical database maintenance, and because, as part of the NIH, it might set up a research program in computation biology. The function of the NCBI is to build up new information technologies to assist a kind of the molecular and genetic processes that lie beneath health and diseases. Its precise aim contains the formation of computerized systems for storing and analyzing biological information for the expansion of sophisticated methods of computer support information dealing out the facilitation of user access to databases and software and the management of hard work jointly to make biotechnology information universal.

Biological Database

The whole genomes of a range of organisms have known to a protein sequence/structure deficit; essentially, because it is an easier quickly to construct huge amounts of sequence information then it is to verify protein 3D structure in atomic particulars. We have obtained the utmost advantage from the overflow of sequence information, to pact with it in a concentrated technique. This means establishing, maintaining and disseminating databases given that easy to use software to access the information they have and designing state of the art analysis apparatus to dream of and understand the structural and functional clues underlying in the data. The first step, then, in analyzing sequence information is to accumulate it into central, sharable resources, i.e., databases. Databases are successfully electronic filling cabinets, a suitable and competent technique of storing huge amounts of information. There are lots of unusual database types, depending together on the environment of the information being stored (e.g. sequences or structures, 2D gel or 3D structure images and so on) and on the way of data storage (e.g. whether in flat-files, tables in a relational database, or objects in an object-oriented database). In the perspective of protein sequence analysis, we will come upon main, composite and secondary databases. Such resources accumulate diverse levels of information in extremely diverse formats. In the history, this has led to a diversity of communication troubles, apart from rising computer technologies are starting to make available solutions, permit flawless, clear access to dissimilar, dispersed data structures over the internet. Primary and Secondary databases used to deal with diverse aspects of sequence analysis, because they accumulate diverse levels of protein sequence information.

Primary Sequence Databases

Early 1980s, sequence information started to turn into extra rich in the scientific literature. Realizing this some laboratories saw that there might be reward to harvesting and storing these sequences in central repositories. Therefore, some primary database projects began to change in diverse parts of the globe.

A few of the mainly vital nucleic acid and protein sequence databases:-

***DNA sequence databases*-**EMBL, the nucleotide sequence database from the European Bioinformatics Institute (EBI), includes sequences together from straight author submissions and genome sequencing groups, and from the scientific literature and copyright applications [6]. The database is shaped in alliance with DDBJ and GenBank; the participating groups bring together a section of the entire sequence data reported global. Information has retrieved from EMBL using the SRS Sequence Retrieval System [7]. DDBJ is the DNA Data bank of Japan, which started at 1986 with the collaboration of EMBL and Genbank [8].

***GenBank*-**GenBank, the DNA database from the National Center for Biotechnology Information (NCBI), includes sequences from publicly accessible sources [9], mainly from straight writer submissions and major sequencing projects [10]. To assist make certain complete coverage, the resource connections data with equally the EMBL Data Library and DDBJ. A GenBank release includes the sequence files; indices formed on a variety of databases fields (e.g. GenPept, a database of translated coding sequences in fastA format). At first, GenBank was accessible on CD-ROM, which proved to be a suitable mechanism for general and comparatively economical allocation. On the other hand, as the volume of the database grew, the number of CDs essential to hold becomes clumsy (12 for the last available CD release) both for the producers and for the users. Today, GenBank offered only through FTP [11].

***cDNA Libraries and ESTs*-**The method for detaining an expression outline is easy. A sample of cells has obtained, and then RNA extracts from the cells and is steady by using overturn transcriptase to run off cDNA from RNA pattern. The cDNA has renovated into a library appropriate for use in quick sequencing experiments. A sample of clones is chosen from a library at unsystematic-e.g. 10000 from a library with a complication of 2 million clones. A significant automated sequencing operation is necessary to make 10000 sequencing reactions and run these on automated sequencers. The resultant data has been downloaded to computers for additional analysis.

Pair wise Alignment Techniques

Database inquiry can receive the appearance of text queries or sequence similarity searches e.g. given the sequence of human adrenergic receptor displayed all related sequences in the databases. This is a positive use to attempt, if the reader has modest knowledge in database interrogation, as these two questions construct slightly diverse outcome.

Blast-The BLAST (Basic Local Alignment Search Tool) algorithm was described by Altschul et al. [12]. BLAST calculates all the segment pairs between the query and the database sequences.

Comparing two sequences-Shaping the resemblance between two sequences, every selected from an alphabet of complexity 20. The raw approach is to line up the sequences beside each other and insert extra characters to carry the two strings into vertical alignment

Unaligned

```
Sequence 1      (query)   AGGVLIIQVG
                          | | | | ||
Sequence 2      (subject) AGGVLIQVG
```

Aligned

```
Sequence 1      (query)   AGGVLIIQVG
                          | | | | |  || | |
Sequence 2      (subject) AGGVL-IQVG
```

In the above figure, the use of gap character '-' is to bring two sequences into alignmen t. Vertical bars indicate identical matches – six in the first alignment, nine in the second.

Commercial Databases

Majority of biological databases are available from publicly accessible servers on internet. The industrial approach to information technology is the desire to purchase databases to well-defined problems, rather than the more exploratory academic approach. In industry, if services exist to develop and maintain databases, and can be purchased finance and human resources can then be released for the more exacting scientific task of searching and analyzing the data. Major releases of DNA and protein sequence databases occur every three to four months. In the meantime, newly determined sequences add to daily-updated files.

Comprehensive packages

GCG-Commercially available sequence analysis software is the GCG suite. This was developed by the Genetics Computer Group at Wisconsin [13], primarily as a set of analysis tools for nucleic acid sequences, but which in time included additional facilities for protein sequence analysis.

EGCG-Extended GCG, or EGCG, began life at EMBL in Heidelberg as a collection of programs to support EMBL's research activities. Since 1988, through a collaboration of groups within EMBnet and elsewhere, further additions have been made to the suite in order to provide support both for core analysis activities at the Sangar Centre, and for the entire user base of EMBnet national nodes. There are more than 70 programs in EGCG, covering themes such as fragment assembly, mapping, database searching, multiple sequence analysis, pattern recognition, nucleotide and protein sequence analysis, evolutionary analysis, and so on.

Staden-The Staden Package is a set of tools for DNA and protein sequence analysis. It does not provide databases, but the software works with the EMBL database and other databases in a similar format. The package has a windowing interface for UNIX workstations. Amongst its range of options, the suite provides utilities to define and to search for patterns of motifs in proteins and nucleic acids.

Laser gene-Laser gene is a PC- based package that provides facilities for coding analysis, pattern and site matching, and RNA/DNA structure and composition analysis, restriction site analysis, PCR primer and probe design, sequence editing, sequence assembly management, multiple and pair wise sequence alignment, protein secondary structure prediction and hydropath analysis, helical wheel and net creation, database searching.

Internet packages

CINEMA-CINEMA is a Color Interactive Editor for Multiple Alignments, written in java [14], the programs allow creation of sequence alignments by hand, generation of alignments automatically, and visualization and manipulation of sequence alignments currently resident at different sites on the Internet. CINEMA provides links to the primary data sources, thereby giving ready access to up-to-date data, and a gateway to related information on Internet.

Conclusion

To bring to a close, a more numerical approach of thinking to standardize and appropriately guide our role in biology should be embraced else there may be non-identical complexity. For that we could do with to get back to study biology, not so in a great deal but to master all of it with the object of adequately enlightening ourselves for gaining assured level of essential familiarity and insight. In order, the part of bioinformatics has agreed a latest sanction of life to the discipline of statistics over and above computer science. The two filament of double helical arrangement of DNA correspond to information and computer data connecting which biology is surrounded as instigation of nucleic acids. Undeniably, the prospect for statistical process is bioinformatics and at this point it is within our approach and it is supposed to be now our initiative to throw our gates wide open and give it the welcome it duly deserves.

References

[1] Baguski, M.S., Tolstoshev, C.M. and Bassett, D.E. 1994, “Gene discovery in DBEST,” Science, 265, pp. 1993-1994.

[2] EMBL, http://www.ebi.ac.uk/ebi_docs/embl_db/ebi/topembl.html

[3] Emmert, D.B., Stoehr, P.J., Stoesser, G. and Cameron, G.N., 1994, “The Europeon Bioinformatics Institute (EBI),” Nucleic Acids Res., 22(17), pp. 3445-3449.

[4] MIPS, http://www.mips.biochem.mpg.de/

[5] Michie, A.D., Jones, M.L. and Attwood, T.K., 1996, "DbBrowser: integrated access to databases worldwide," TIBS, 21(5), pp. 191.

[6] Stoesser, G., Moseley, M.A., Sleep, J., McGowran, M.,Garcia-Pastor, M. and Sterk, P., 1998, "The EMBL Nucleotide Sequence Database," Nucleic Acids Res., 26(1), pp. 8-15.

[7] Etzold, T., Ulyanov, A. and Argos, P., 1996, "SRS-Information retrieval system for molecular-biology data-banks," Methods in Enzymol., 266, pp. 114-128.

[8] Tateno, Y., Fukami-Kobayashi, K., Miyazaki, S., Sugawara, H. and Gojobori, T., 1998, "DNA Data Bank of Japan at Workon genome sequence data," Nucleic Acids Res., 26(1), pp. 16-20.

[9] Wheeler, D.L. and Chappey, C., 2000, "Database resources of the national center for biotechnology information," Nucleic Acid Res., 28(1), pp. 10-14.

[10] Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J. and Oullette, B.F.F., 1998, "Genbank," Nucleic Acids Res., 26(1), pp. 1-7.

[11] GenBank, http://www.ncbi.nlm.nih.gov/web/Genbank/

[12] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990, "Basic local alignment search tool," J. Molecular Biol., 215, pp. 403-410.

[13] Devereux, J.,Haeberli, P. and Smithies, O., 1998, "A Comprehensiveset of sequence-analysis programs for the VAX," Nucleic Acids Res., 12(1), pp. 387-395.

[14] Parry-Smith, D.J., Payne, A.W.R., Michie, A.D. and Attwood, T.K., 1998, "CINEMA-A novel Color Interactive Editor for Multiple Alignments," Gene, 211(2), pp. 45-56.