International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 5 Issue VII, July 2017- Available at www.ijraset.com

Study on Forensic Analysis using Bioinformatics

Devi Selvam[1], S. Gayathri[2], P. Divya[3]
[1,2,3]*Assistant Professor, Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology*
Member of CSI

Abstract: Bioinformatics is an interdisciplinary field mainly involving molecular biology and genetics, computer science and statistics. Large-scale biological problems are addressed from a computational point of view. The common problems are modelling biological processes and making inferences from collected data. It involves the following steps to provide bioinformatics solution: First step is that Collection statistics through biological information and data. Second step is Building a computational model and to Solve a computational modelling problem. Test and evaluate a computational algorithm. In this paper it provides introduction to bioinformatics, biological terminology and then discussing some classical bioinformatics problems .Sequence analysis is the analysis of DNA and protein sequences for clues includes identification of homologues , multiple sequence alignment, searching sequence patterns, and evolutionary analyses. Protein structures are three-dimensional data and analysis of protein structures for clues regarding function, and structural alignment. Gene expression data is represented as matrices and analysis of microarray data involves statistics analysis, classification and clustering approaches. Biological networks such as gene regulatory networks, graph theoretic approaches are used to solve problems such as construction and analysis of large-scale networks.

Keywords: DNA APP, Entrez ,DNA, Forensic, Analysis

I. INTRODUCTION

To analyse and study how normal cellular activities are altered in different disease states, the biological data must be combined to form a comprehensive picture of these activities. The bio informatics field has evolved and now involves the analysis and interpretation of various types of data. This includes nucleotide and amino acid, protein structure protein domains. The process of interpreting and analysing data is referred to as computational biology. Bioinformatics and computational biology has important sub-discipline it includes: Implementation and Development of computer programs that enable efficient use , access to, and management of, various types of information. Statistical measures and Development of new algorithms that assess relationships among members of large data sets. For the instance, there are methods to locate a gene within a sequence, to cluster protein sequences into families of related sequences and to predict protein structure and/or function. Increase the understanding of biological processes is the primary goal of bioinformatics. It differs from other approaches, however, is its focus on developing and applying computationally intensive techniques to achieve this goal. Examples include: visualization, machine learning algorithms, data mining, pattern recognition .Some of the Major research efforts in the field include modelling of evolution and cell division , prediction of genome-wide and protein–protein interactions**,** association sequence alignment, gene finding, genome assembly, drug design, drug discovery, protein structure alignment, protein structure prediction. Bioinformatics has insists the theory to solve formal and practical problems arising from the management and analysis of biological data and creation and advancement of databases, algorithms, computational and statistical techniques. For the last few decades, developments in information technologies and other molecular research technologies have combined to produce a tremendous amount of information. It is related to rapid developments in genomic and molecular biology. Bioinformatics has some mathematical and computing approaches used to glean understanding of biological processes. Bioinformatics include Common activities such as creating and viewing 3-D models of protein structures, aligning DNA and protein sequences to compare them and mapping and analysing DNA and protein sequences.

II. RELATION TO OTHER FIELDS

The normal cellular activities are changed in different disease to study, analyze, and form complete picture of these activities possible only when the biological data is combined. The bioinformatics field has part in various domains. It involves analysis and interpretation of various forms of data. This contains nucleotide and amino acid, protein structure protein domain. Computational biology is the process of interpreting and analyzing data in the bioinformatics. Bioinformatics and computational biology has sub-category They are: Implementation and Development of computer programs that provides efficient use ,access to, and management of, various forms of information. Developing new algorithms and Statistical measures that makes relationships

©IJRASET (UGC Approved Journal): All Rights are Reserved

between the members of large data sets. For instance, there are methods to locate agene within the sequence, to predict protein structure or function and to combine protein sequences into families of same sequences. Maximizing the understanding of biological processes is the main goal of bioinformatics. It varies from others, but it focus on developing and applying computationally intensive techniques to achieve. For Example: visualization, machine learning algorithms, data mining, pattern recognition .Some of the main research efforts include modeling of evolution and cell division , prediction of genome-wide and protein-protein interaction, association sequence alignment, gene finding, genome assembly, drug design, drug discovery, protein structure alignment, protein structure prediction. Bioinformatics has made the theory to solve formal and practical problems coming from the management and analysis of biological data and creation and advancement of databases, algorithms, computational and statistical techniques. For the last few series, developments in information technologies and other molecular research technologies have combined to produce a large amount of information. It is related to very high developments in genomic and molecular biology. Bioinformatics field has some mathematical and computing approaches used to thorough understanding of biological processes. Bioinformatics include Common activities such as creating and viewing 3-D models of protein structures, aligning DNA and protein sequences to compare them and mapping and analyzing DNA and protein sequences.

III. FORENSIC ANALYSIS

The field of forensic science is increasingly based on bio molecular data and many European countries are establishing forensic databases to store DNA profiles of crime scenes of known offenders and apply DNA testing. The field is boosted by statistical and technological advances such as DNA microarray sequencing, TFT biosensors, machine learning algorithms, in particular Bayesian networks, which provide an effective way of evidence organization and inference. The aim of this article is to discuss the state of art potentialities of bioinformatics in forensic DNA science. We also discuss how bioinformatics will address issues related to privacy rights such as those raised from large scale integration of crime, public health and population genetic susceptibility-to-diseases databases. Bioinformatics and forensic DNA are inherently interdisciplinary and draw their techniques from statistics and computer science bringing them to bear on problems in biology and law. Personal identification and relatedness to other individuals are the two major subjects of forensic DNA analysis. Typical contexts for forensic analysis are disputes on kinship; for example paternity disputes, suspected incest case, corpse identification, alimentary frauds (e.g. OGM, poisonous food, etc), semen detection on underwear for suspected infidelity, insurance company fraud investigations when the actual driver in a vehicle accident is in question, criminal matters, autopsies for human identification following accident investigations. Genetic tests have been widely used for forensic evidences and mass-fatality identification (terrorist attacks, airplane crash, tsunami disaster). Genetic testing results are integrated with information collected by multidisciplinary teams composed of medical examiners, forensic pathologists, anthropologists, forensic dentists, fingerprint specialists, radiologists and experts in search and recovery of physical evidence. Large scale tissue sampling and long-term DNA preservation under desiccation conditions with potential applications in mass fatalities has been recently described.

IV. BASIC FORENSIC DNA PROCEDURE

Some of the characteristics used for personal identification such as biological, physiological, behavioral. The biological contains DNA, blood, saliva. The physiological contains fingerprints, eye irises and retinas, hand palms and geometry and facial geometry. The behavioral contains dynamic signature, gait, keystroke dynamics, and lip motion. It is the combination of physiological and dynamical characteristics such as the voice. The most important personal identification is DNA. The expressed regions of DNA (genes) and some segments of DNA with no known coding function which are being all genetic differences are the pattern of inheritance which can be monitored. Because it can be used as markers. Consider any two humans who are greater than 99% identical in their DNA sequences but they are still having millions of genetic differences. It making them different in their risk of getting certain diseases and response to environmental factors. The copy number variation (CNV) , large genomic regions are the most important sources of genetic variations. SNP(single nucleotide polymorphisms) which is generally defined as single base difference among two different individuals of the same species. An average of every 1/2000 bases will be occurring in humans SNPs. The genome of human is one of the highly repetitious thing. In most of sequence length scales, number and dispersion repetitions will occur. Homo- and di-nucleotide repeats (microsatellites), and families of interspersed are such examples of repetitions in which elements hundreds of base pairs long such as the ALU sequences. There are more than one million ALU sequences in the human genome. In each 300 bases long, in other parts of the genome and generate mutations which are able to copy themselves. It allows detection of genetic variations among humans, usually short, repetitive loci variable number of tandem repeats

©IJRASET (UGC Approved Journal): All Rights are Reserved

(VNTRs) polymorphism in forensic DNA typing often requires the use of above techniques that allow the were used till few years ago. The core units three, four or five nucleotides long are composed such loci and the number of repeated segments at a locus varies between individuals. Therefore 17 bp sequence of DNA repeated between 70 and 450 times in the genome is considered as one VNTR in humans. The total number of base pairs at this locus could vary from 1190 to 7650. The replacement of VNTR is by STR (short tandem repeats). CNV are become useful in forensic science and supposed to be major determinants of human traits.

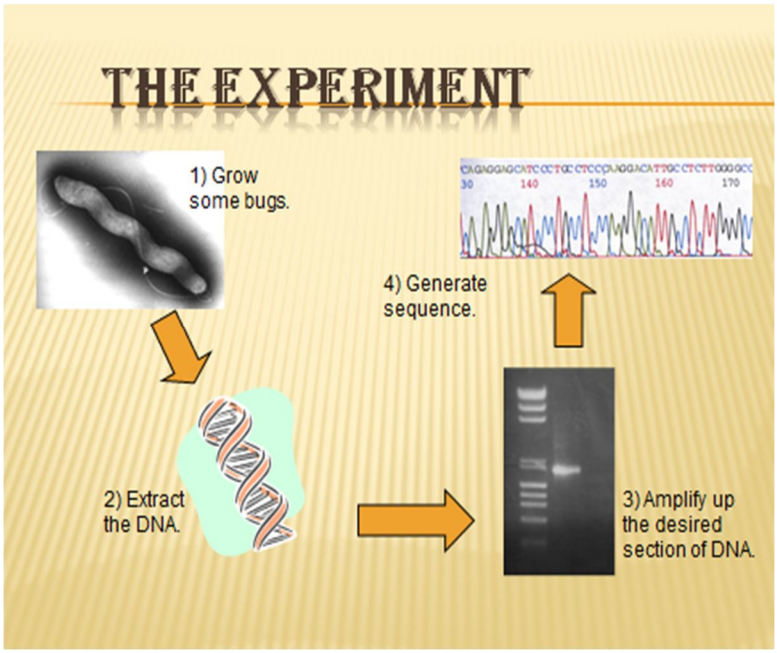

Fig.1 A example experiment for DNA Analysis

V. DATABASE

Biological database are big group of life science information. The data is gathered from scientific experiments, published literature, high-throughput experiment technology, and computational analysis. The databases consists of information gathered from research areas whichincludes genomics, proteomics, metabolomics, microarray gene expression, and genetics. The Information contained in biological databases may include gene function, structure, localization (both cellular and chromosomal), clinical effects of mutations as well as similarities of biological sequences and structures. Biological databases are mainly classified into categories, structure and functional databases. Nucleic acid and protein sequences comes under sequence databases and solved structures of RNA and proteins comes under structured databases. Functional databases issue done only for providing information regarding physiological role of gene products. For an instance, enzyme activities, mutant phenotypes, or biological pathways. Model Organism Databases are functional databases and these databases provides species-specific data. Databases are one of the most important tool . It is used for helping scientists in order to analyze and explain a host of biological phenomena from the bio-molecular structure to the metabolism of organisms and for understanding the species evolution. This knowledge is used to fight against diseases and to help in the development of medications and in predicting similar genetic diseases and in discovering basic species relationships in the life history. Biological knowledge is distributed among various general and specialized databases. At times this might lead to some difficult in ensuring the information's consistency. Integrative bioinformatics is the field to tackle this ultimate problem by providing unified access. The only remedy is that how biological databases are cross-referenced with other databases with accession numbers to link their related knowledge .Relational database concepts of computer science and informatics of digital libraries are very important for understanding biological databases. Relational database concepts of computer science and informatics reveal concepts of digital libraries are the important concepts for understanding biological databases. Biological database design, development, and long-term management is one of the concentrated area of the discipline in bioinformatics. Data contents contains gene sequences, textual descriptions, attributes and ontology classifications, citations, and tabular data. These are known as semi-structured data. These information can be represented in the form of tables, key delimited records, and XML structures.

©IJRASET (UGC Approved Journal): All Rights are Reserved

VI. ENTREZ TOOL

The Entrez Global Query Cross-Database Search System is one of the search engine and it is a ultimate web portal for users to search as many as discrete health sciences databases at the National Center for Biotechnology Information (NCBI) website. The NCBI which is one of the part of the National Library of Medicine (NLM) is itself a department of the National Institutes of Health (NIH).It is also one of the part of the United States Department of Health and Human Services. The name "Entrez" derives from a French word and its meaning is "Come in!". It was chosen to personify the spirit of welcoming the people for to search the content available from the NLM. Entrez Global Query is an integrated search and retrieval system which is used to access databases s with both a single query string and user interface. Entrez is used to retrieve related sequences, structures, and references in an useful manner. The Entrez system is used to provide results of gene and protein sequences and chromosome maps. The main pages of Entrez allows access to the global query. All databases that are mapped by Entrez can be searched through single query string supporting boolean operators and search term tags to limit parts of the search to certain fields. This returns with results on the next page, which shows the number of hits for thag search in all of the databases. Also this will link to actual search results for that exact database. Entrez tool provides an interface for searching every particular database and also for refining our search results. The Limits feature is a good feature which allows the user to shorten the sight of search through web forms interface. The History feature of Entrez makes a list of queries that are actioned recently by the user. The previous queries results can be referred by number and combined by boolean operators. Search results can be saved for short span of time in a Clipboard. The users with a My NCBI account is allowed to save queries as long as possible an for long span of time and it is also used to choose and have updates with new search results which will be mailed for saved queries of many of the available databases. It is broadly used in the biotechnology field as a reference tool for students as well as professionals.

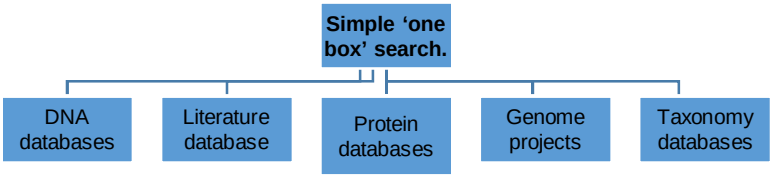

Fig.2 Entrez search toolbox

A. Dna and Agricultural Fields

Every one might be wondering how DNA could be possibly relate to agricultural yields. But the thing is actually quite interesting one which means it is a curious mix of nature and technology. Agricultural yields can be increased through the use of genetic modification (GM), which relies on taking coveted properties from one organism and inserting those genes into another organism in order to create those same desired effects. The idea behind increasing agricultural yields is to provide more crops with the reduced costs. Generally, While genetically modified crops have been created to feed people in developed nations, research has been suggested that their use in developing countries has been far more beneficial.

1) *Benefits of Improving Agricultural Yields:* One of the key benefits of improved agricultural yields has been that they are thought to help battle poverty and hunger in developing. This is mainly used to support third world countries. In fact, in one trial of cotton yields, yields were eighty percentage higher in GM crops when compared to non-modified crops. Even every researchers were surprised at the enormous difference between the two types of crops. Therefore, increase in agricultural yields is considered a sustainable form of development and it also carries the benefit of reduced pesticide use because GM crops are sprayed less frequently than non-GM crops.
2) *Concerns Regarding Technology in order to Increase Crop Yields*: However, still there are many concerns regarding the practice of increasing agricultural yields by means of genetic modification. In many environmental crusaders cite ,this type of technology has many flaws and which in turn may lead to any increase in yield which is only temporary because insects and similar pests will be eventually becoming resistant to the GM crops. Still, others argues which is surrounding the concepts that even if the benefit is only a short-term, it is still a valuable one for farmers in developing nation as a developed nation in the

©IJRASET (UGC Approved Journal): All Rights are Reserved

world. Many farmers are not affording with the constant increasing cost of pesticides and their yields were jeopardised. By using GM crops, the pesticide resistance can be allowing them to provide greater yields at a reduced cost in terms of pesticides and enhancing the profit percentage for the farmers. It is also important to note that tropical climates could be bringing them even more difficult in terms of pest challenges. This means that pesticide resistant crops become more important in this climate.

3) *Future of GM Crops and Improving Agricultural Yields:* The controversy of GM crops will not likely end soon, at least not until long-term studies have been performed to assess their merits, efficacy and safety. The improvement of agricultural yields is a constant and important issue for farmers. Indeed, their profits and livelihood depend on the provision of substantial agricultural yields each year. In fact, the livelihood of the entire community will also depend, in part, on adequate crop yields - particularly in developing nations where staple crops may ultimately provide most of the population's food. The ability to improve these yields through GM crops is an important issue to consider but the long-term implications regarding pesticide resistance must also be acknowledged and addressed. Research will also hopefully include these concerns and perhaps better GM crops will be developed that can afford consistent and lengthy pesticide resistance. Until that time, alternatives to GM crops should still be considered, specifically ones that do not entail the numerous challenges, questions and controversies of GM crops.

Fig.3 Agricultural field

B. Dna Emergent Technologies

In 1984 Sir Alec Jeffreys first reported his DNA profiling technique given for crime solving. This issue mainly focuses on the latest DNA technologies, designed specifically to address these issues and increase the efficiency of DNA evidence. “DNA Fingerprinting Comes Of Age” discusses how the recent years automation suppliers have started to provide smaller, more affordable instruments that enable these labs to automate many rate-limiting steps: extracting the DNA sample from group of cards, taking the cell to extract DNA, transferring liquid reagents, and preparing for the STR- and PCR-based workflow. ”This technology also allows DNA evidence to be applied in a big range of cases (property crime cases, for instance) that would have previously been thought too minor to warrant processing DNA evidence.“ Development of an Innovative DNA Quantification and Assessment System: Streamlining Workflow Using “Intelligent Tools” gives the way to next generation STR systems drive the need for improved quantification systems. When a sample can be processed in just over an hour, collected DNA at the crime scene can be immediately compared to DNA in databases or that of suspects whose DNA was taken upon booking. As discussed in “DNA First”, when a myriad of samples are collected at the scene, Rapid DNA technology can allow the CSI to determine which samples have enough genetic material to warrant further analysis at the lab. These new technologies will allow more DNA evidence to be processed more efficiently, reduce backlogs, and especially in the case of Next Generation Sequencing technologies, help process more complex samples.

Fig.4 DNA profiling technique

©IJRASET (UGC Approved Journal): All Rights are Reserved

C. DNA Application

DNA app is an online store for information about your genes will make it cheap and easy to learn more about your health risks and predispositions. Other technologies DNA finger printing DNA bar coding Human genomes hold information about our health risks, our physical traits. Helix which calls the idea "sequence once, query often." (The company says customers would find these apps on websites and possibly in the Android and Apple app stores.) Kao says the tactic will make genetic information available to consumers "at an unprecedentedly low entry price. The engine to power the app store is being assembled a mile from Illumina's San Diego headquarters, in a building where workmen were still bending sheet metal and laying floor tiles in January. Customers will control their data by deciding who sees it. There's even a "nuclear button" to erase every A, G, C, and T. But key details are still being sorted out. As with browsing on Amazon, he thinks, people will discover things they "didn't know they needed but that [are] targeted to them, and that they want." A looming question mark is the U.S. The bottom line is going to be: What are the regulatory constraints on information that is truly useful?" says MirzaCifric, CEO of Veritas Genetics. Cifric hasn't decided whether to create an app with Helix, but he says he shares its core belief: "The genome is an asset that you have for life, and you'll keep going back to it."

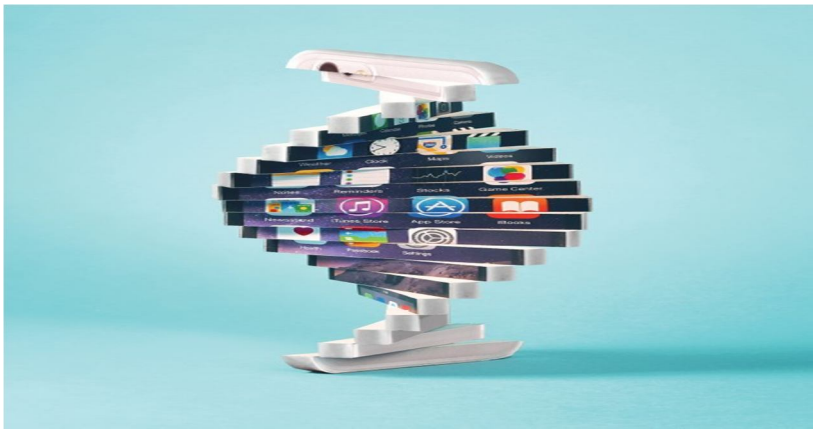

Fig.5 DNA app

VII. CONCLUSION

Analyses in bioinformatics is mainly focusing on three types of large data sets that are available in molecular biology: macro molecular structures, genome sequences, and the results of functional genomics experiments (e.g. expression data). Additionally it includes the text of scientific papers and "relationship data" from metabolic pathways, taxonomy trees, and protein-protein interaction networks. Bioinformatics is employed with wide a wide range of computational techniques including sequence and structural alignment, database design and data mining, macro molecular geometry, phylogenetic tree construction, prediction of protein structure and function, gene finding, and expression data clustering. Approach is mainly based on integrating a variety of computational methods and heterogeneous data sources. Finally, bioinformatics is a practical discipline and We are surveying some representative applications, such as finding homologues, designing drugs, and performing large-scale censuses.

REFERENCES

[1] www.kinexus.ca/kinetica/weblinks/bioinformatics
[2] www.elsevier.com/journals/genomics...and-bioinformatics
[3] https://en.wikipedia.org/wiki/Bioinformatics

©IJRASET (UGC Approved Journal): All Rights are Reserved