DNA Sequence Analysis and Matching for Forensic Science Used in Crime Department

Dr. A.V. Senthil Kumar
Head and Associate Professor, Department of Computer Applications, Hindusthan College of arts and Science, Coimbatore, Tamil Nadu, India.

E. Siva Prakash
PG Student, Department of Computer Applications, Hindusthan College of arts and Science, Coimbatore, Tamil Nadu, India.

Abstract – Crime investigation and forensic science activities are having tremendous growth now a days, every day crime rate increases and thus creates a complication for the investigation department in criminal detection process. The proposed system integrates the crime department and bio-informatics to handle the above problem. The application performs DNA (Deoxyribonucleic Acid) sequence analysis and matching process to find the DNA matching profiles in the crime department. To perform criminal shortlist process and biological relationship analysis for forensic science, a new application is developed. In this paper, we proposed a new sequence matching algorithm for effective DNA profile matching.

Index Terms – Bio-Informatics, DNA, Forensic Science, Criminology, Sequence matching, Pattern analysis.

1. INTRODUCTION

The major scientific development in the area of DNA technology and its facts revelation has solved many interacting crime related mysteries because of natural or human agencies or in solving civil dispute specially related with the paternity of a child and in finding the identity of an individual .It has also been used in solving the cases of exchange of babies in hospital wards and also protection of farmers rights and biodiversity. The modern day biology is seeking new and better ways to enhance our quality of life through the application of technology, which is called as Biotechnology and rapid progress in research on human genome [1]. The currently developing technique of DNA promises a degree of accuracy greater even than current method of finger printing suspect. DNA profiling has been used extensively for paternity testing as well as for the criminal investigation. DNA profiling has particular application to the criminal law because of the possibility that it offers of determining whether blood or semen deposits located at the scene of a crime come from a person suspected of having committed the crime [2]. The focus of most criminal investigations is on linking the evidence from the crime scene to suspects, and for more than a century, science has played an increasingly important role in this process. Fingerprinting [3], Genetic engineering [4], sequence of whole genome or exploitation of the differences between the DNA of the male and the female have thus all been historical imperative.

1.1. DNA Structure

The structure of DNA varies from individual to individual. DNA structure determines human character [5], behaviour and body characterization. The DNA sequences are generated and are used to match with other DNA profiles. The followings are the sample structure of DNA.

```
Sample DNA results
>sampled sequence 1 consisting of 50 bases.
aatatacggcgactaactgtgtcaaattgacgtcaggcatgtagcggtag

>sampled sequence 2 consisting of 50 bases.
gactgaattagatccgtttgggctcggatcgcaaccccgcaacgctgcga

>sampled sequence 3 consisting of 50 bases.
cgatcgtactcagctgagtgatacattatagatttcagttagtactgtaa

>sampled sequence 4 consisting of 50 bases.
cttcgggtttgggcggcctaacaaaattgaggataattctccagcctact

>sampled sequence 5 consisting of 50 bases.
caactcccctcccgattaagccttgcacacgacgtttcggagagcccatg
```

Fig 1.0 Sample DNA Sequence

The above fig 1.0 shows the sample DNA results with 5 sequence samples. The above data includes 50 bases in the sample DNA sequences. In the proposed system, the criminal DNA profiles are gathered and maintained in the DNA warehouse. This will compared with the requestor sample for suspected list generation.

1.2. DNA for Crime Investigation

The history of crime has started with the history of civilization. Every society has been enacting laws to combat crimes and

criminals since long time. From times Immemorial, the single quest of any legal system has been the Quest for truth. Author in [6] has assumed sacred proportions as regards to the Indian psyche. Crime is a social and economic phenomenon distressing the whole human community. With sweeping changes in society and in economic activities, the face of crime has also changed. Today the sophisticated technology in the hands of criminals is creating mayhem. The law enforcement agencies need to be a step ahead of the lawbreakers.

Forensic science provides investigators or crime fighters that weapon, the most significant being DNA profiling or DNA test. DNA technology is one such tool, which upholds truth. The focus of most criminal investigations is on linking the evidence from the crime scene to suspects, and for more than a century, science has played an increasingly important role in this process. Fingerprinting, Genetic engineering, sequence of whole genome (be they of men, animal, plant or microorganism) or exploitation of the differences between the DNA of the male and the female (for example of the X and Y sperm) have thus all been historical imperative.

Fingerprinting was applied to criminal investigation beginning in the 1880's. Shortly after the principle of ABO blood group typing was reported in 1900. Its relevance to forensic investigation became apparent. In the 1960's human leukocyte antigen (HLA) typing became the premier serologic tool for personal identification although in practice, it was useful for only a small percentage of samples.

Finally, the 1980's ushered in the age of DNA testing, which permits investigators to perform almost unbelievable feats of identification with current techniques[7]. It is possible for a single person to be differentiated from all the people that have ever lived using DNA from a single hair root .The principles and techniques used for forensic DNA typing are also quit useful for other purpose.

DNA profiles are widely used in resolving issues of parentage in man and animal, and are rapidly replacing serologic analysis (i.e. blood typing) for that purpose[8][9][10]. Additionally, DNA testing is an indispensible tool for positional cloning, a technique by which a previously unknown gene is identified by finding association or links between DNA markers and the inheritance of a disease.

2. PROBLEM DEFINITION

There are only few application performing DNA sequence matching analysis, however, those tools are not completely integrated with the forensic science and investigation. The existing system follows a DNA test and fingerprint tests for criminal detection. But due to the lack of tools, this can't be retrieved quickly. The existing system may have n number of criminal records, every record contain basic information's rather than the biological sequence. The existing system suffers due to the manual and inconsistent works on the biological tests.

3. PROPOSED SYSTEM

The paper is developed a new DNA matching tool, which is developed for a DNA lab. This includes the number of test requests from the criminology department. The authority can match the DNA report with the suspected criminal list. If the DNA matches, then the reports will be informed to the criminology department. This paper is the integration of bio lab and criminology department. All the activities of the criminology and bio-lab will be maintained in the website.

This paper presents a laboratory information management system for DNA sequences created and based on the needs of a criminology department. DNA bar-coding is a global initiative for species identification through simple DNA sequence markers. This aim sat generating sequence code and data for all test samples. The website has been developed to better manage large amounts of sequence data and to keep track of the whole experimental procedure. The system has allowed us to classify strains more efficiently as the quality of sequence data has improved, and as a result, up-to-date taxonomic names have been given to strains and more accurate correlation analyses have been carried out.

The proposed system created with the intension of developing a new and secure online application, which combines the bio-informatics and criminal department together for effective crime detection. Forensic Bioinformatics promotes establishing of forensic databases to store DNA sketches of crime scenes of known offenders and apply DNA testing. Strength in terms of accurate results is boosted by statistical and technological advances such as DNA microarray sequencing,

Genetic tests have been widely used for forensic evidences and mass-fatality identification. Genetic testing results are integrated with information collected by multidisciplinary teams composed of fingerprint specialists, medical examiners and forensic pathologists. In several countries, new rules could allow fingerprints and DNA samples to be taken from anyone they arrest, whether they are charged or not. The DNA sequence will be matched with the suspected criminals DNA online and the matching percentage will be displayed to the crime department.

Phase1: Crime and bio-service authority endorsement: This phase collects the authorized user details before providing the webpage. This page verifies and controls unwanted logins from their credentials. The phase has a registration and login process, where the administrator will create login authentication to the crime and bio-service authorities. The randomly created id and password will send to the authorized person via email. The authority can active their login credential after that.

Phase 2: Forensic DNA profiling : The crime department collects and sends the bio-information's to the lab for DNA test. The process begins with a sample of an individual's DNA, typically called a reference sample. The most desirable method of collecting a reference sample is the use of a buccal swab as this reduces the possibility of contamination. When this is not available (e.g. because a court order may be needed and not obtainable) other methods may be used to collect a sample of blood, saliva, semen, or other appropriate fluid or tissue from personal items such as toothbrush, razor or from stored samples like banked sperm or biopsy tissue. Samples obtained from blood relatives (biological relative) can provide an indication of an individual's profile, as could human remains which had been previously profiled.

Phase 3: Service requisition: The service requisition phase allows the crime department authorities to request the bio-service, which includes the DNA test and DNA sequence generation process. The crime department can send numerous users DNA for test. Before sending it, the authority should send a request regarding the test to the bio-service authorities.

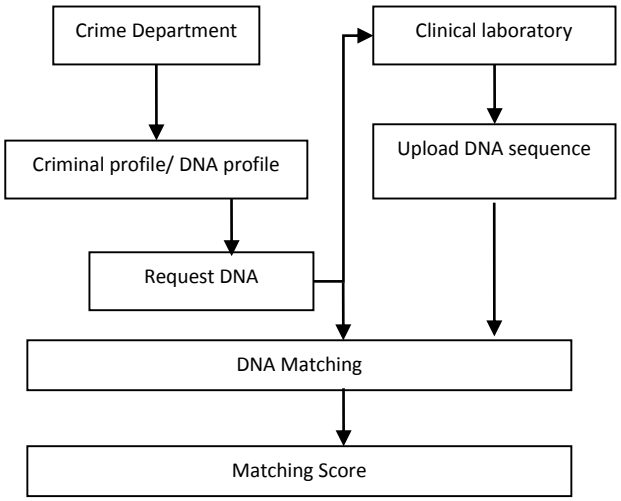

Fig 2.0 Proposed Architecture

Phase 4: DNA profiling and profile statistics: This phase will be performed by the bio-lab authorities. DNA profiling also called DNA testing, DNA typing, or genetic fingerprinting is a technique employed by forensic scientists to assist in the identification of individuals by their respective DNA profiles. DNA profiles are encrypted sets of numbers that reflect a person's DNA makeup, which can also be used as the person's identifier. DNA profiling should not be confused with full genome sequencing.

Phase 5: DNA sequence matching and reporting: The final phase reports the top most matched profiles according to the DNA sequence. The received DNA sequences will be matched and compared with others. The crime authority can start the investigation with the top matched individuals.

The DNA casework unit analyzes biological samples from crime scene evidence and reference samples from known individuals. The forensic and known sample DNA profiles are compared to determine possible sources of the DNA. The analysts in the casework unit issue scientific reports on their findings and are often called upon to testify as to the findings in criminal proceedings.

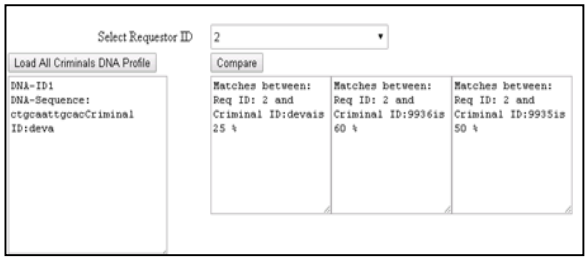

Fig 3.0 results of matching

Sequence matching techniques:

Sequence alignment is a process of comparing two or more DNA sequences. High sequence similarity in DNA samples usually implies significant structural and functional similarity between two profiles. In biological process a sequence determines structure and structure determines function of a bio-molecule. Therefore, by using sequence alignment we try to discover or validate similarity in structure and function between two human DNA profiles. Similarity measure represents both quantitative and qualitative resemblance in sequences. The percentage of matches among the sequences denotes quantitative measure. The degree of similarity and differences among the sequences denotes qualitative measure in a sequence alignment. In the proposed system, the sequence matching techniques are used to find the suspected list of criminals which matches in the given DNA profile.

Algorithm: DNA sequence matching

Input: DNA profiles of requestor and criminals

Output: matching score and suspected list

Steps:

1. Read the DNA profiles from criminal data D.
2. Read DNA sequences from the requestor.
3. Split the whole data into n sequence set n.

   N=split(DNAseq)
4. Match the DNA with each sequence set.

   Ms=Match (Ri..Ci)
5. Return score Ms.

The above sequence matching steps allows the user to match the given DNA with the existing DNA profiles. The algorithm returns the matching score of two given sequences. Based on the score, the percentage will be calculated.

4. IMPLEMENTATION AND RESULTS

The Experimental analysis is indented to be of use to show the results and findings of the research work. It has two goals: first, to provide a useful guide to new experiment lists about how such work can best be performed and written up and the next one is to challenge current researchers to think about whether their own work might be improved from a scientific point of view. Efficient implementations allow one to perform experiments on more and/or larger instances or to finish the study more quickly. The proposed DNA sequence matching framework is implemented using .Net framework. The system has taken n number of DNA sequences from the ncbi website. The followings are the sample records used for experiment.

ACAAGACGGACAGTGAGAATGAATGCACGG ATTCGAGGGGCATAAG

Table 1.0 Sample DNA Sequence

The Table 1.0 shows the extracted DNA sequence. For the implementation, several DNA profiles are extracted and converted into the proper dataset format.

The first set of experiments is to compare the performance of different combinations of DNA test, similarity and relationship based strategies with and without using similarity scores. All apps are compared with one another. There are totally 10 criminal records are taken for the comparison. In more specific this paper particularly interested to find the time taken to find matching score for the total number of DNA profiles and attributes. And matching prediction from the set of profiles and its accuracy of the detection are summarized.

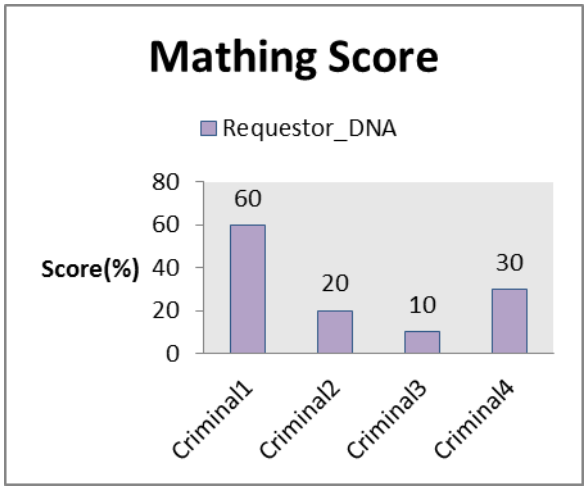

Fig 4.0 DNA Matching Score Analysis

After implemented the proposed system, the system matches the requestor DNA with the criminal list and the matching score is plotted in fig 4.0.

5. CONCLUSION

This paper proposes a new framework to calculate DNA similarity between two more DNA profiles. It combines DNA sequence relationship and feature similarity and sequence matching process and finds the criminals who involved in the activity. By this process, the relationship between two profiles can be identified. Using the proposed system, the department can find the person age from their DNA profile. This helps to the police/crime department to improve the investigation accuracy. This paper also makes two improvements on existing DNA analysis process. One is to make it high-quality even with weak initial parameters and minimum sequences. The other is to reduce DNA sequence generation time and increases the accuracy.

REFERENCES

[1]. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. and Gocayne, J.D., 2001. The sequence of the human genome. *science*, *291*(5507), pp.1304-1351.

[2]. Ribaux, Olivier, Simon J. Walsh, and Pierre Margot. "The contribution of forensic science to crime analysis and investigation: forensic intelligence." *Forensic Science International* 156.2 (2006): 171-181.

[3]. Dolf, Gaudenz. *DNA fingerprinting: approaches and applications*. Vol. 58. Birkhäuser, 2013.

[4]. Butler, John M. *Forensic DNA typing: biology, technology, and genetics of STR markers*. Academic Press, 2005.

[5]. Kieff, Elliott, et al. "Epstein-Barr virus: structure of the viral DNA and analysis of viral RNA in infected cells." *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* 560.3 (1979): 355-373.

[6]. Sweet, D. J., and C. H. W. Sweet. "DNA analysis of dental pulp to link incinerated remains of homicide victim to crime scene." *Journal of forensic science* 40.2 (1995): 310-314.

[7]. Ribaux, O., and P. Margot. "Inference structures for crime analysis and intelligence: the example of burglary using forensic science data." *Forensic Science International* 100.3 (1999): 193-210.

[8]. Allen, Marie, et al. "Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities." *Journal of Forensic Science* 43.3 (1998): 453-464.

[9]. Holland, Mitchell M., and Thomas J. Parsons. "Mitochondrial DNA sequence analysis-validation and use for forensic casework." *Forensic science review* 11 (1999): 21-50.

[10]. Wilson, Mark R., et al. "Guidelines for the use of mitochondrial DNA sequencing in forensic science." *Crime Lab Digest* 20.4 (1993): 68-77.