Middle-East Journal of Scientific Research 24 (S1): 131-135, 2016
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2016.24.S1.28

An Efficient Approach and its VLSI Implementation for the Study of Dna Sequences

[1]T. Anuradha and [2]DR. T.M. Inbamalar

[1]P.G. Scholar, Rmk Engineering College, Kavaraipettai, India
[2]M.E., Ph.D Associate Professor, Rmk Engineering College, Kavaraipettai, India

Abstract: Deoxyribo Nucleic Acid (DNA) sequencingdeals with the alignment of DNA strings with their nucleotides and their subsequences. Today, large genome projects such as the Human Genome Project, are yielding information about coding sequences using computational technologies. Bioinformatics is the development and application of mathematical, statistical and information technology methods which benefits the human beings. Over the past decade, Bioinformatics, the science of discovering, locating and characterizing genes has become ubiquitous and increasing exponentially.The genetic features and its characteristics are studied by DNA sequencing Technologies. Gen Bank is NIH genetic sequence database collection of all publicly available DNA sequences.This paper introduces an algorithm for DNA string matching. Its main application is intended to be the detection of similarity of strings in DNA chains.The DNA sequences are converted into binary sequences and passed through bandpassfilter. Cross-correlation is performed which measures the similarity between series. From the output obtained helps to find the similarity of DNA sequences.In Bioinformatics due to more data and more time for computation of that data, DNA sequencing becomes a difficult task. But in this method, the time taken for computation is less and less complexity.

Key words: Bioinformatics · Deoxyribo Nuleic Acid · Digital signal Processing · VLSI

INTRODUCTION

DNA sequencing is the technique used to determine the sequence of bases adenine (A), cytosine (C), guanine (G) and thymine (T) in a DNA molecule. The order of these bases defines what we are now and in the future [1]. It carries the information which the cell needs to assemble the DNA and proteins. It may be used to determine the sequence of individual genes, larger genetic regions and full chromosomes or entire genomes. The order of individual nucleotides which exists in DNA molecule is provided by the sequencing. This is varied by the genetic information of different beings such as animals, plants, bacteria or any other source. In many areas such as medicine [2], forensics and biological sciences these are very useful. There are methods defined for the DNA sequencing. Some of them [3] are Maxam and Gilbert method, Fredrick Sanger method. Generally DNA sequencing consists of four main steps. They are PCR (Polymerase Chain Reaction) sequencing Reaction, Electrophoresis, Computer Scanning. But the modern Sequencing uses Sanger technique. The DNA Sequencing is used for Protein Sequencing. It resembles the replication of DNA. The other types of sequences are Dye-Terminator sequencing [4], Nanopore sequencing and Shotgun sequencing.

The process of aligning two or more Genetic sequences with each other in order to determine evolutionary changes and similarities is called as Multiple Alignment. Dynamic programming is the approach for alignment of two sequences. This approach is not effective for more large sequences. This method is applicable for multiple sequences [5] also. There are heuristic methods like the local multiple alignment using the Sum of Pairs scoring function to fasten dynamic programming. Scoring function determines the comparison sequence result for the DNA sequence which contains the amino acids like A,T,C,G. Maximum score value meant for the closed relationship for the compared sequence with the base sequence.

Corresponding Author: T. Anuradha, P.G. Scholar, Rmk Engineering College, Kavaraipettai.

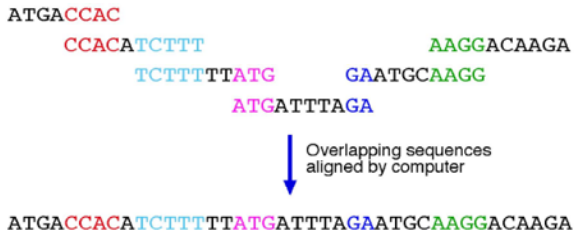


Fig. 1.1: DNA Sequences

The DNA sequences involved a location-specific primer extension strategy was established. DNA polymerase catalysis and specific nucleotide figure prominently in current sequencing schemes, were used to sequence [6] the cohesive ends of lambda phage DNA. Through this determination of DNA sequence using synthetic location-specific primers takes place.primer-extension strategy was adopted by Frederick Sanger to develop large number of DNAsequencing methods. Sequencing methods, like chemical degradation" are developed. Using a method wandering-spotanalysis 24 base pairs contained sequences are studied by Gilbert and Maxam in 1973.DNAtechnology and its methods in sequencing advancements seperates DNA samples from one source to other.

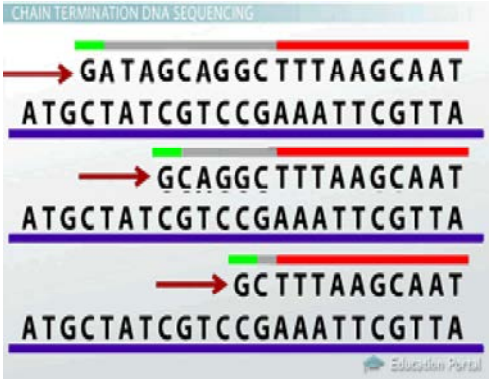


Fig. 1.2: SANGER method sequencing

For the field of complex trait genetics Next-generation sequencing has many technological leaps forward and realized numerous possibilities [7]. The issues which are hardly to solve are multiple production, design orientation with analysis and interpretation issues. The Progress in development and optimization started to materialize for appropriate delivery of data and powerful analysis [8].

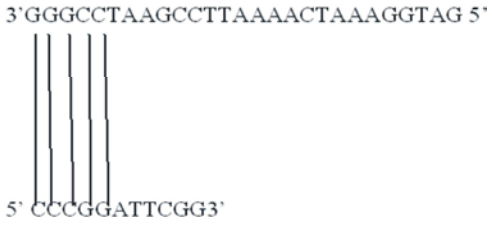
Fig. 1.3: Next generation sequencing method.

Matching of DNA sequences is defined through the DNA profiles. To determine the characteristics of DNA this is the forensic technique. DNA is collected from the blood samples and cut into small pieces by restriction enzyme [9]. For better idea of data processing requirements and role of automation plays, we have to consider map-based sequencing methodology used by the Human Genome Project.At first extraction of DNA from the subject organism. Then partition of genome into large segments and each segment [10] cloned into stable form like a Bacterial Artificial Chromosome.

DNA Template to be sequenced:

3’ GGGCCTAAGCCTTAAACTGGTTATTTG 5’

The sequencing primer annealed to the template to Start DNA sequencing:



The contributions of this work include the efficient and reliable technique for the matching of biological sequences. The input DNA sequences are converted into numerical binary sequences. Then the sequences are passed through the band pass filter. The sequences are correlated and denoising with DWT takes place which results the output. By this we can find the similarities and matching of the sequences In Section II the closely related prior work is discussed and Section III describes proposed techniques used Section IV deals with implementation and simulation Results. Finally, conclusionis mentioned in Section V.

Background: Several methods exist for the computation and development of sequencing methods. Various aspects in matching are briefly discussed with respect to the contributions made by various authors. Dr. Seema Verma et al proposed high speed VLSI Comparator hardware Design for performing the matching of DNA sequencing. This is based on High Level Synthesis. The drawback is complexity and consumes time to execute. The Design of pipelined comparator framework for rapid matching of the biological samples. The GAUT II which is a high level synthesis tool allows hardware to execute the process. Complete genome of Haemophilus influenza Bacterium which is a free living organism was first proposed in 1995 by Venter, Hamilton Smith and colleagues. The chromosome with 1,830,137 bases marked the first published use of whole-genome shotgun sequencing, eliminating initial mapping needs. To produce a draft sequence of the human genomethe shotgun sequencing methods are used in 2001. Decipheration of complete DNA sequence of the Epstein-Barr virusby medical research council consists of 172,282 nucleotides. There was a turning point marked in DNA sequencing by the completion of sequence because it was not prior of genetic profile of the virus. A non-radioactive method for transferring the DNA sequences reaction mixtures onto matrix which was immobilizing during electrophoresis was developed by Pohl and co-workers in the early 1980s. Followed by the commercialization of the DNA sequencer called "Direct-Blotting-Electrophoresis-System used in EU genome-sequencing program framework, which determines the complete DNA sequence of the yeast. The successes are starting to emerge in the literature. The linking sequence polymorphism key for the full variation, frequency and effective spectrum to polygenic phenotypes is set to transform. Thus the complex trait genetics research is carried out.

Proposed Work: The correlation and its methods are used which performs the matching of the biological sequences. The similarity measure of two series is defined in signal processing. Thus this is also called as sliding Dot-Product or Sliding Inner Product. In this work the input sequences are firstly converted into the binary numerical sequences. Then it is passed through the filter. The sequences obtained are Cross-Correlated and Discrete Wavelet Transformation is performed. The output is obtained. The output that obtained is compared and matching is done. Described through the block diagram as in Fig. 3.1.

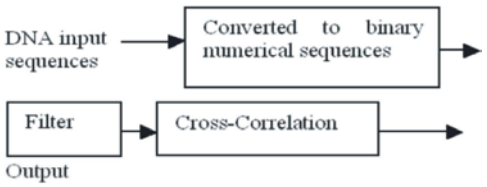


Fig. 3.1: Block diagram

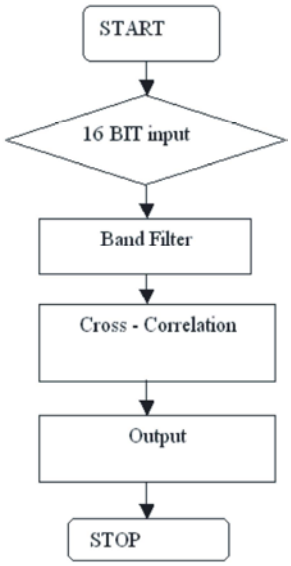


Fig. 3.2: Flowchart of Simulation

The input DNA sequences are given. This is then encoded and turned to 16 bit binary data as the input.Then the sequence is passed through the filter to avoid the unwanted errors. The sequences are correlated and DWT is implemented.The implementation is through the flowchart representation as in Fig. 3.2. Techniques that are used for the computation of matching of DNA are the Encoding and Filtering. For the removal of unwanted noises the filtering takes place through the band pass filter. To decrease the time taken for the processing and computational complexity reduction we are going for these methods.

Discrete Wavelet Transform (DWT) Algorithm computes the wavelet coefficients and the similarities between the sequences through correlation. These Correlations are used in Recognition of pattern, Single particle analysis, Crypt analysis and neurophysiology. The main function of DWT is to perform the task in both numerical and functional analysis. Thus considered as efficient one. Compared to FFT it has both frequency and location information. The DNA sequence matchingare seen by simulations in Section IV.

RESULTS

The proposed algorithm has been implemented using the Modelsim software. ModelSim software here used to perform simulations with Logical verification, to ensure expected results and Behavioral verification, to verify logical and timing for the input sequences. has the high-performance, low-voltage applications in communications and computing systems. The simulation of DNA sequence matching through

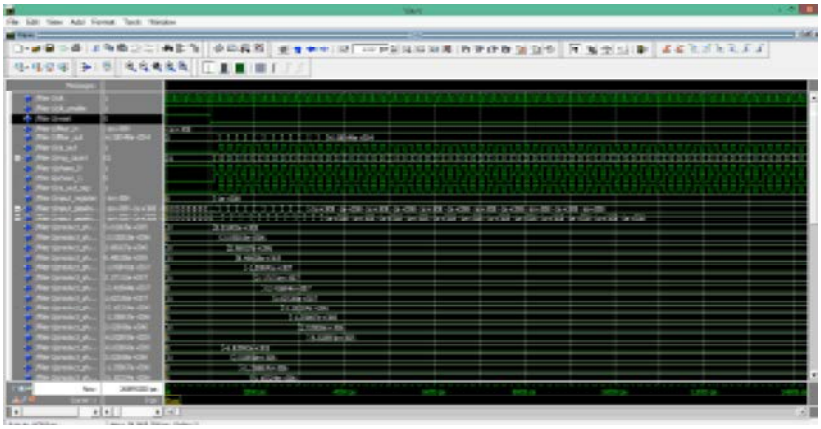


Fig. 4.1: FILTER Module

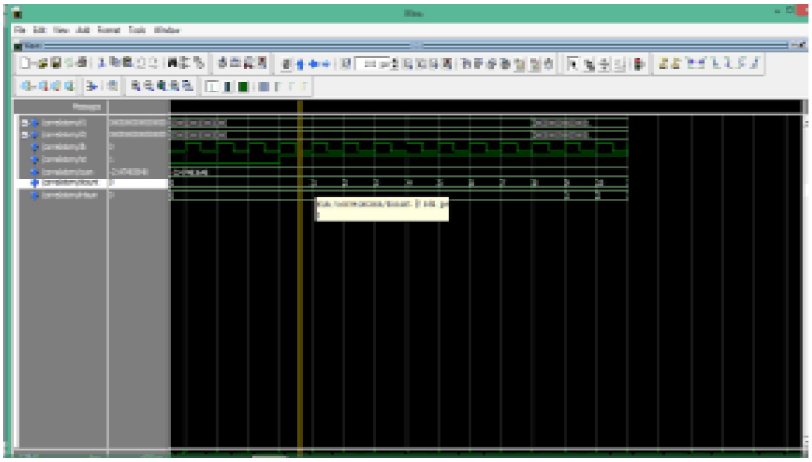


Fig. 4.2: CROSS-CORRELATION Module

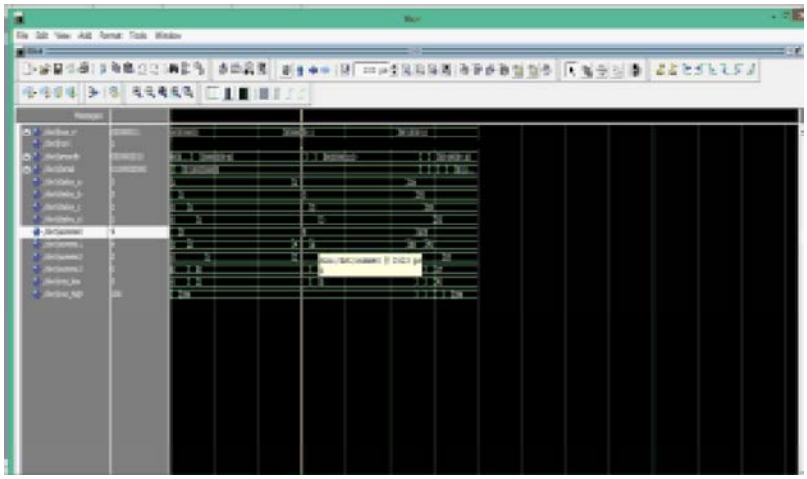


Fig. 4.3: DWT Module

correlation shown in Fig 4.2 by the DWT Module representing the output. From these we can observe the similarities and matching of the sequences in Fig 4.3.

CONCLUSION

Main aim of the paper is to study the properties of DNA sequences efficiently. Throgh this algorithm and through this analysis done on the DNA sequences, detection time is reduced and this provides solution to the existing slow DNA sequences analysis processes. Moreover, through the proposed architecture the similarity measures between DNA sequences can be determined with greater efficiency. Also, the proposed architecture can be implemented for real time DNA sequence chains. Hence, it can be useful for fast comparison and similarity detection of DNA sequences.

REFERENCES

1. Butler, T.Z., M. Pavlenokov and I. Derrington, 2008. 'Single-molecule DNA detection with an engineered MspA protein nanopore' Proc. Natl. Acad. Sci., 106(9).
2. Chentaal, M., 2014. 'Accelerating the Next Generation Long Read Mapping with the FPGA-Based System' IEEE/ACM Transactions on Computational Biology and Bioinformatics, 11(5): 840-852.
3. Dr. Seema Verma, Sanjeev Kumar and Dibyayan Das Sharma, 2011. 'Hardware Design of a High Speed VLSI Comparator Framework for DNA Sequence Matching Using High Level Synthesis ' 2(4).
4. Gen Bank, Genetic sequence database. Available: www.ncbi.nlm.nih.gov/genbank.
5. http://en.wikipedia.org/wiki/DNA sequencing.
6. Hu, Y. and P. Georgiou, 2014. 'Robust ISFET pH-Measuring Front-End for Chemical Reaction Monitoring', 8(2): 177-185.
7. Mardis, E.R., 2011. 'Next-generation DNA sequencing methods.' Annual review of genomics and human genetics, 9: 387-402.
8. Merriman, B. and J.M. Rothberg, 2012. 'Progress in ion torrent semiconductor chip based sequencing.' Electrophoresis, 33(23): 3397-417.
9. Purnell, R., 2008. 'Nucleotide Identification and Orientation Discrimination of DNA Homopolymers Immobilized in Protein Nanopore.
10. Sebastiao, N., 2012. 'Integrated hardware architecture for efficient computation of the n-Best Bio-sequence local alignments in embedded platforms' IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 20(7): 1262-1275.