ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Biological Systems Uses in Authentication's Application.

Divya Prakash Shrivastava[1]* and Mohamed M Agnan[2].

[1]Faculty of Computer and Information Science, HCT University, Abu Dhabi, UAE.
[2]Faculty of Medical Technology, Al Jabal Al Gharbi University, Gheraian, Libya.

ABSTRACT

Currently, biological Science and computer Sciences there are many areas has been discovered, with many techniques, Algorithms have invented and developed and classifies such related problems. The gene is the fundamental unit of information in the living system containing DNA. During research, researcher identified DNA is capable of universal computation. The present scope of bio informatics encompass the wide area of data acquisition, storage, retrieval analysis and interpretation along with parallel development of application tools in terms of computation infrastructure, software, database and so on. Biometric Authentication, DNA assessing their use in security applications and how they compare against one another for particular security protocols and authentication mechanisms

Keywords: DNA, Authentication, Bioinformatics, Biomolecular. Sequence Alignment.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

In biological research and discovery some computation and calculation are always required, but some time not so well defined while complex and complicated. As the computer science are being developed very fast more and more computational techniques are of greater assistance to help the biologist in their research. When the need or request by a biological research is provided, modeling and solving this kind of problems have been more and more well powered by the computer technology. Computer Scientist found great interest in this area and got well-funded. Now days as more and more efforts are being put in this area grows at very rapid rate and becomes one of the hottest one's the age.

We believed that research at the interface of biology and information technology may lead to important new information system (Algorithms and software) and computer technologies (Hardware). The question is what and how can we learn and understand from the biological systems, and how can we adopt them and adapt them to develop these new computer technologies.

BIOLOGY AND COMPUTATION

Bio computation has been used as catch all term for research at the interface of biology and computation, but that term is used in so many ways and for such different subsets of this intersection as to cause confusions. To help guide discussions, we offer that the general area of bio computation can be divided into four major categories: Biomolecular Computation, Computational Biology, Bioinformatics and Biological Computation. This paper is intended to address the first category: biomolecular Computation

Biomolecular Computation

This category includes efforts to exploid biological macromolecules to implement relatively standard methods of computation. Molecular computing in the unique combination of computer science and molecular biology. It tries to solve the computational problems with the DNA (RNA, protein) materials and rules.

It tries to solve the computation problem with the DNA materials and rules. It illustrates the possibilities of using DNA to solve a class of problems that is difficult or impossible to solve using traditional computing methods. It demonstrate unique aspects of DNA as a data structure. It shows that computing with DNA can work in a massively parallel fashion. In 1994, Adleman first presented the power of DNA computing with sample examples[1].

Molecular Structure

DNA has not been considered as a form of Biometric until recently. It dictates how we look and develop; it serves as a full set of building blocks / 'blueprints' for each person and is as unique as a fingerprint. DNA is short for Deoxyribonucleic Acid [2].

With sufficient amounts of technical equipment and specialist knowledge, it is possible to look at the molecular structure of two people and tell exactly how they differ from one another. Humans share 99.9% of DNA [2]; it is the 0.1% which DNA sequencing / analysis looks at. It is estimated that 1 in every 1,200 to 1,500 bases differ between individuals [3].

DNA is a large molecule, whose shape resembles a 'double helix' which conceptually looks very much like a ladder twisted into a spiral. Each molecule consists of a string of nucleotides; each nucleotide has three components: a sugar molecule, a phosphate molecule and a nitrogenous base [3].

Sequencing Software

DNA sequence are familiar structures. Researchers produce them daily through a process called sequencing, and compare them to get information about an organism, a sequence variant, or a disease.

A process by which the sequence of Nucleotides along a strand of DNA is determined. To carry out the sequencing of the human genome scientists cut the DNA up into short fragments, sequenced these fragments

simultaneously and them assembled the entire genome by using sophisticated computer techniques to match the fragments to each other.

The DNA is sequenced using a technique known as Electrophoresis. The molecule is separated into its two constituent strands and placed at one end of a gelatine-like gel, prepared in advance using complex procedures. Electrodes are placed at either end of the gel and a current passed through the gel, causing bases in the DNA to move through the gel. This creates 'bands'; the position and strength of each band depends entirely on the size of the base within the DNA. The larger the sample, the more difficult it will be to move through the gel; strands that are the same size will move to the same position in the gel. Only 200-300 bases can be represented on one gel piece; multiple gels are used and the whole process is repeated for the whole DNA strand [4].

The sequence obtained is separated into four distinct batches and analysed using a machine known as an 'Assembler'. A lab technician pours the gel containing the sequence into two glass plates which are two hundredths of a millimetre apart. A chemically modified base is then added to each batch and the chains of bases are fixed upon the addition of the chemically modified base [4].

Once the bases have been colour coded using the dye, the assembler loads the DNA into 96 lanes that are typically 3 or 4 meters long and 30cm in width and then reads the order of bases at the positions that they are present in the gel. When bases move through the gel and appear at the other end, the dye emits a fluorescent colour when scanned by a laser [4].

Sequencing software matches the colour exhibited to its corresponding base and forms a binary representation of the DNA strand. The diagrams below illustrate how the sequencing software assigns bit sequences to each base.

Each sequence generated by the sequencer is known as a "read". Recursive bit sequences are known as "repeat-induced" and are removed by the Sequencer, leaving bit sequences that are unique to the individual.

Assemblers require up to a thousand times more RAM than general purpose computers due to the exponential number of sequence combinations and the huge number of comparisons required to keep track of sequence matches [3].

Feature Extraction

Enrolment requires the involvement of specialist knowledge, such as lab technicians that have a good experience and handling of DNA.

To start with, the sample that is extracted must be at least 5KB in size. This equates to 625 base pairs that are required to encode one stand of DNA (one base is represented as two bits; there are four possible bases that are to be presented) [4]. Lab specialists 'prime' the DNA by placing it into a test tube and splitting it into its two separate counter parts using restriction enzymes. The diagrams below illustrate a DNA molecule and when it has been split into its constituent components.

The DNA must be the correct concentration - if it is too strong, distilled water is used to dilute it. If it is too weak then more DNA is added or replicated using replication software. Temperature is an important factor in maintaining the integrity and structure of the DNA. If the temperature is too high, excess bases may separate from their strands and cloud the solution of DNA, thereby corrupting the sample.

DNA Sequence comparison

Once the sequence have been obtained, the task becomes to compare them. Sequences are compared, among other things, for evolutionary relationships, sequence polymorphism or overlaps in target sequence reconstruction. Sequences compared include shotgun reads, assembled genomes or chromosomes, EST sequences mapped to genomic sequence or protein sequences. Another common task is to find sequences similar to a query sequence in a sequence database.

ISSN: 0975-8585

There are many methods to compare protein sequence. The Dynamic Programming design for protein comparison depend on sequence database size, find optimal alignment between two sequences and use of application etc.

AUTHENTICATION

Individuals must first register their form of identity with the system by means of capturing a raw biometric to be used in the system. This process is called Enrolment and is composed of three distinct phases: Capture, Process and Enrol.

- Capture: A raw biometric is captured by the Biometric sensing device.
- Process: Characteristics that are unique to individuals and distinguish individuals from one another are extracted from the raw Biometric and transformed into a biometric "template".
- Enrol: The processed template is stored in a suitable storage medium such as a database on a disk storage device or on a portable device such as a Smart Card, whereby later comparisons can be made easily.

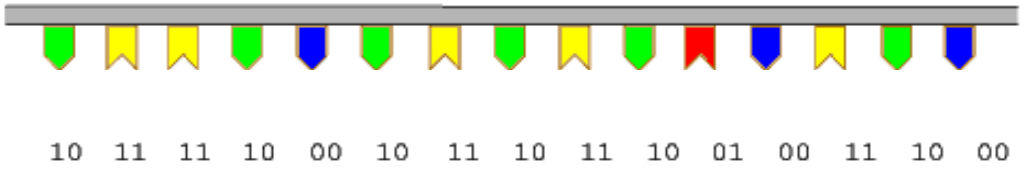

Figure 1: Binary representation of the DNA strand

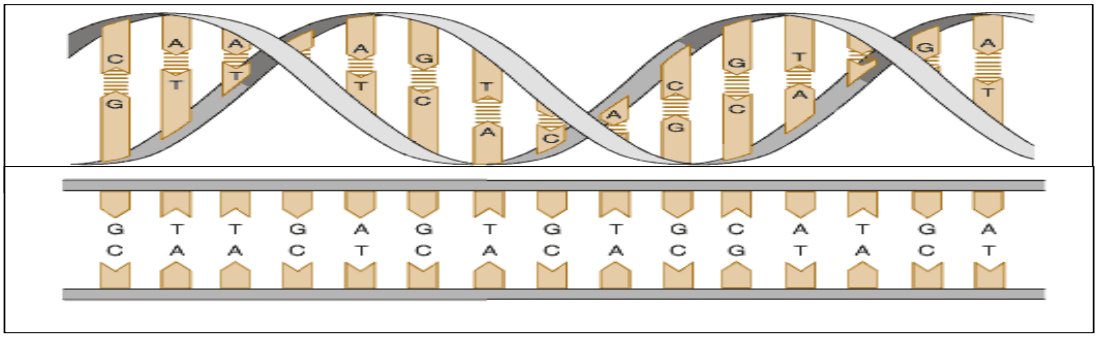

Figure 2: DNA molecule and its constituent components.

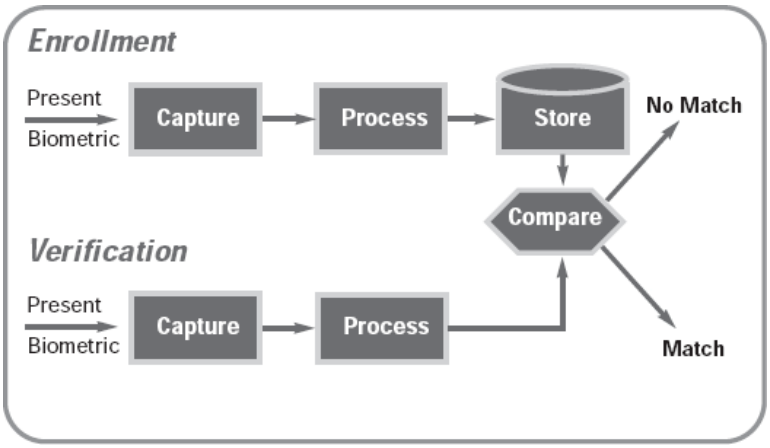

Figure 3: The process of Enrollment and Authentication

Once Enrolment is complete, the system can authenticate individuals by means of using the stored template. Authentication is the process whereby a new biometric sample is captured by the individual who is authenticating with the system and compared to the registered (enrolled) biometric template. There are two forms of Authentication: Verification and Identification.

The success of a system in performing verification is measured using the metrics below. Successful systems will have high True Positive and True Negative values, a poor system will have high False Positive and False Negative values. Each metric is defined as follows:

- TP: correctly allow access to an authorised user
- TN: correctly deny access to an unauthorised user
- FP: incorrectly allow access to an unauthorised user (FAR)
- FN: incorrectly deny access to an authorised user (FRR)

A diagram illustrating the process of Enrollment and Authentication is shown in figure 3.

The Matching Process

Matching involves repeating the enrolment process to obtain a test sequence to compare against the stored template. Bit sequences are compared between the test sequence and the template: if all bit sequences match then the individual is authenticated; if the bit sequences do not match the individual is rejected.

The best systems at present take ten minutes to perform sequencing and matching for DNA samples, but this is not seen as 'real-time'; other forms of verification such as fingerprint can instantly verify individuals. DeCypher [6] is currently the world's fastest sequencer and assembler.

Current Applications

Currently, there are databanks and reserves in the United States, such as the GenBank and MegaBases at Johns Hopkins University that are used to determine the identity of individuals for forensic purposes.

Current uses of DNA as a biometric include a genetic fingerprint obtained from hair, blood and cell samples, used to form part of the next generation of ID cards in China. At present the Chinese government is investing 182 million in the expansion of a national database which will contain in excess of two million profiles [7].

NTT DATA have produced two devices that incorporate DNA as a biometric: the Commodity ID System and the Smart Card with Registered Seal. Both products use a secret key for use in personal recognition systems [8].

The Commodity ID System looks at providing authenticity for legal documents that are easy to forge at the moment and products that are easy to imitate in all industries. The system approaches this problem through a combination of two steps:

- Manufacturers of products shall possess a unique DNA sequence for each product and generate a DNA-ID code which can be placed onto the product as a barcode.
- Authenticity of the product can be tested in all future instances by scanning the barcode and testing to see whether it's DNA matches that recorded by the manufacturer.

For legal documents, DNA can be dissolved and mixed with ink that is used to print the document in question and a match query executed to see whether the ink used matches that of the template. However, this condition only holds if no one has stolen the ink and replicated it already, thereby rendering this process useless.

A Smart Card with Registered Seal [8] is to be released in accordance with the Electronic Signature Law in China, whereby an electronic registered seal function will prevent identity theft through the use of public and private encrypted keys based on the person's DNA-ID. The person registers their public key derived from their

DNA with the certification agency; digital signing is then performed using the private key and authentication takes place on the card using the public key and private key. To make the card more secure, a barcode containing DNA is affixed onto the card, similar to the barcode used in the Commodity ID System.

Future Applications

Some of the shortcomings of biometric authentication have been discussed; a solution to making a more reliable authentication system would be to combine different biometric systems. For instance, providing security at an airport requires a high level of security especially at customs where passport checks take place; this would require centralised databases of users and their biometric templates.

Facial recognition and DNA authentication could take place. This is a future scenario, where it would be possible to perform DNA sequencing in real time. 3D facial recognition could be used to identify the traveller and a DNA sample could be tested for a match with the template stored in the user's passport which would then verify that the user is really who they claim to be. In cases where DNA sequencing is not possible, the traveller's fingerprint can be compared to the template stored in the passport.

There is ongoing research into how best to apply biometrics to airport security due to the US governments demand that all travellers wishing to enter the US must have some form of biometric information in their passport.

CONCLUSION

Advantages

DNA is easy to represent in binary format; four digits are used to encode each base, allowing the matching process to be done quickly and efficiently.

It is seen as the most complex biometric due to the complexity of the sequencing and matching processes. People tend to think that DNA is the most secure technology and this would be true if it were only to be used for high security applications.

It is extremely hard to forge or imitate as it proves difficult to extract good samples that are likely to match the template stored for the individual in the system.

DNA has been used by government sources to confirm the identity of particular individuals. To authenticate such an individual on a large scale and prove their identity to the whole world requires a reliable biometric.

DNA is digital, increasing the accuracy and allowing true recall to be gained for the process of authentication.

Disadvantages

Enrollment is highly intrusive; it requires a sample of blood from the individual that is to be registered with the system and more than one sample may need to be taken if the process needs to be repeated many times.

To perform Enrollment requires specialist knowledge and equipment; both are expensive to acquire. As a result of this, the only situation in which DNA would be cost-effective is in high security applications, such as defense and national security.

Recent advancements in science such as cloning and the near-completion of the Human Genome sequencing project means that the concept of 'uniqueness' for DNA could be obliterated. Identical twins share the same DNA; hence this biometric would not be able to distinguish any points for authentication for identical twins.

ISSN: 0975-8585

REFERENCES

[1] Leonard M. Adelman. Molecular Computation Of Solutions To Combinatorial Problems Science, 1994.
[2] http://www.genetrack.bc.ca/sibship.php.
[3] http://www.genomenewsnetwork.org/whats_a_genome/Chp1_1_1.shtml."
[4] http://seqcore.brcf.med.umich.edu/doc/educ/dnapr/sequencing.html.
[5] http://www.bseinquiry.gov.uk/report/ volume2/fig1_1.htm.
[6] http://www.timelogic.com/decypher_intro.html.
[7] http://www.silicon.com/management/government/0,39024677,10005834,00.htm.
[8] http://www.nttdata.co.jp/en/media/2001/071700.html