

Identifying Deletion Mutation Using Matrix

M. Yamuna¹ and A.Elakkiya²

^{1,2} School of Advanced Sciences, VIT University, Vellore, India.
myamuna@vit.ac.in; elakkiyaappu@gmail.com



ABSTRACT: Mutations, is the change in the blue print of life and that blue print is our DNA in which all phenotypic expressions are coded. The Biotechnologist and researches in this field had today discovered that mutation does not only change the appearance but are involved in behavior of men and women's. Mutations may vary in size from single DNA base to a large fragment of a chromosome. In this paper we propose a method of identification deletion mutation using matrix representation.

KEYWORDS: Mutation, DNA Sequence, Deletion Mutation, Matrix.

1. INTRODUCTION

A mutation, which may arise during replication and/or recombination, is a permanent change in the nucleotide sequence of DNA. Damaged DNA can be mutated either by substitution, deletion or insertion of base pairs. Mutations, for the most part, are harmless except when they lead to cell death or tumor formation. Because of the lethal potential of DNA mutations cells have evolved mechanisms for repairing damaged DNA.

Changes in DNA caused by mutation can cause errors in protein sequence, creating partially or completely non-functional proteins. To function correctly, each cell depends on thousands of proteins to function in the right places at the right times. When a mutation alters a protein that plays a critical role in the body, a medical condition can result [1].

In [2], Gregory A Ryslik et al have given a novel methodology that increases the power to identify mutational clusters by taking protein tertiary structure. Amin Zia et.al proposed computational methods to rank insertion-deletion mutations in the coding as well as non-coding regions and nonsense mutations [3]. In [4] David T. Jones et.al generated mutation data matrices from large numbers of protein sequences by means of an approximate peptide-based sequence comparison algorithm, the set sequences are clustered at the 85% identity level. The closest relating pairs of sequences are aligned, and observed amino acid exchanges tallied in a matrix. In [5] M .Yamuna et.al proposed a method of mutation identification using graph theory properties.

2. MATERIAL AND METHODS

2.1 DNA Sequencing

DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule. It includes any method or technology that is used to determine the order of the four bases - adenine, guanine, cytosine, and thymine - in a strand of DNA [6]. Figure 1 [7] provides an example for DNA sequence.

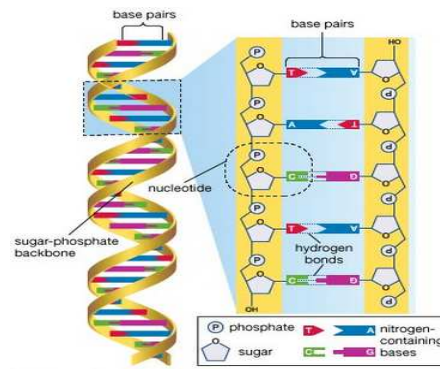


Fig. 1. Example for DNA sequence

2.2 Mutation

Mutation is a permanent alteration of the nucleotide sequence of the genome of an organism, virus, or extra chromosomal DNA or other genetic elements [8]. Figure 2 [9] provides an example for mutation.

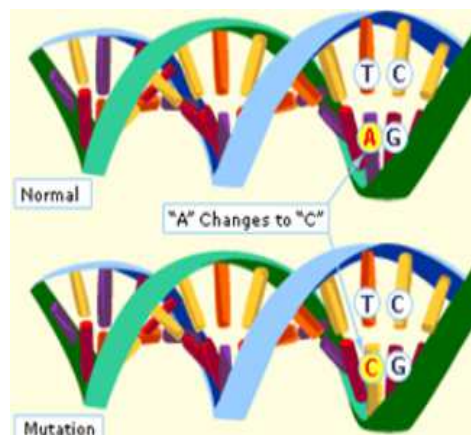


Fig. 2. Example for mutation

There are many different ways that DNA can be changed, resulting in different types of mutation. They are Substitution, Insertion, and Deletion. Figure 3 [10] provides an example for types of mutation.

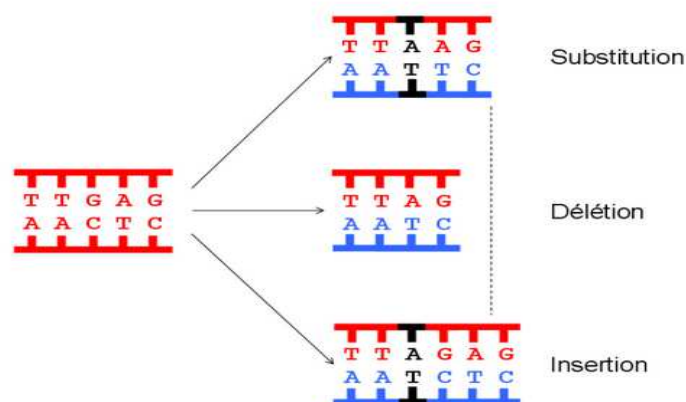


Fig. 3. Types of mutation

2.3 Deletion Mutation

In genetics, a deletion (also called gene deletion, deficiency, or deletion mutation) is a mutation (a genetic aberration) in which a part of a chromosome or a sequence of DNA is lost during DNA replication. Any number of nucleotides can be deleted, from a single base to an entire piece of chromosome. The smallest single base deletion mutations are believed occur by a single base flipping in the template DNA, followed by template DNA strand slippage, within the DNA polymerase active site [11]. Figure 4 [12] provides an example for types of mutation.

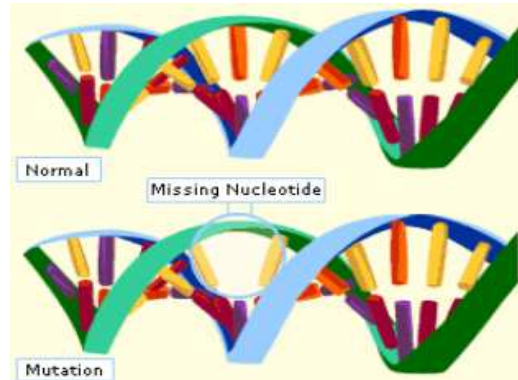


Fig. 4. Types of mutation

3. PROPOSED METHOD

Matrix Construction

Consider any random DNA sequence. The bases in the DNA sequence are considered to be adjacent to next bases in that sequence.

For example consider a random sequence TGACTG. We say that T adjacent to G, G adjacent to A, A adjacent to C, C adjacent to T, T adjacent to G. We define the matrix with 4 rows and 4 columns. Each row and column represents the four bases, we choose the order as ATGC.

DNA Matrix

For any given DNA sequence, we define a DNA sequence matrix D as follows,

$$D = [d_{ij}]_{4 \times 4} = \begin{cases} 1, & \text{if base } i \text{ adjacent to base } j \\ 0, & \text{otherwise} \end{cases}$$

For example for the sequence ATCCAT,

$$D = \begin{matrix} & \begin{matrix} A & T & G & C \end{matrix} \\ \begin{matrix} A \\ T \\ G \\ C \end{matrix} & \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

If two adjacent bases appear more than once in a sequence then we add one more numerical value to the corresponding entry.

Deletion Mutation Verification using DNA Matrix

Let S_1 and S_2 be the two sequences to be verified for deletion mutation. Create the DNA matrix for both the sequences. Label the matrices as D_1 and D_2 . If these two sequences have deletion mutation then we know that some bases are deleted in any of the two sequences. Here we are considering the deletion mutation where exactly one base is deleted. If K is the length of the sequence S_1 , then $K - 1$ is the length of sequence S_2 . When the sequences are compared they will mismatch exactly at one position, this means that in matrix D_1 and D_2 all the entries will match each other in the corresponding position except

for this mismatching place. Let the mismatching part in S_1 be $x_1y_1x_2$, where $x_1 \neq y_1$ and sequence in S_2 be x_1x_2 respectively that is y_1 is deleted in the sequence S_2 . So in the matrix D_1 , we will have additional entries in the positions x_1y_1 and y_1x_2 . Also in the matrix D_2 , we have an additional entry in the position x_1x_2 . This means that entry x_iy_i in D_1 not equal to x_iy_i in D_2 , $i = 1, 2$. From this discussion we conclude that if two sequences have deletion mutation then element wise comparison of D_1 and D_2 results in mismatching entries at exactly at three positions.

Example

Let $S_1 = \text{GATCCCTAGT}$ and $S_2 = \text{GATCCCAGT}$.

Step 1 Construct the matrix D_1 and D_2 as we discussed above,

$$D_1 = \begin{matrix} & \begin{matrix} A & T & G & C \end{matrix} \\ \begin{matrix} A \\ T \\ G \\ C \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix} \end{matrix}$$

$$D_2 = \begin{matrix} & \begin{matrix} A & T & G & C \end{matrix} \\ \begin{matrix} A \\ T \\ G \\ C \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 2 \end{bmatrix} \end{matrix}$$

In D_1 and D_2 , mismatching entries are exactly at three positions. So we conclude that two sequences have deletion mutation.

4. CONCLUSION

Thousands of proteins need to do their job in the right places at the right time for any cell to function properly. Sometimes, gene mutation prevents one or more of these proteins from working properly. By change in gene instruction, in making protein, a mutation causes the protein to malfunction. This leads to severe medical conditions.

A permanent change in the sequence of DNA that makes up a gene is known as mutation. All mutations have the potential to be very damaging, but most are benign. Mutations are an important part of evolution, allowing us to develop adaptations and diversity.

In this method the two mutated DNA sequences can be identified using matrices. It is difficult to find out the deleted bases in such a long DNA sequences, by using this method we can identify the deletion mutation through mismatching entries at exactly three positions in the matrices.

References

- [1] DNA mutation and Repair. [Online] Available: <http://www.nbs.csudh.edu/chemistry/faculty/nsturm/CHEMXL153/DNAMutationRepair.htm>
- [2] Gregory A Ryslik, Yuwei Cheng, Kei-Hoi Cheung, Yorgo Modis, Hongyu Zhao. *MC Bioinformatics* 2014, 15:86. [Online] Available: <http://www.biomedcentral.com/1471-2105/15/86>, 2014.
- [3] Amin Zia, Alan M. Moses, Ranking insertion, deletion and nonsense mutations based on their effect on genetic information, 2011. [Online] Available: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-299>, July 22, 2011.
- [4] David T. Jones, William R. Taylor, Janet M. Thornton, The rapid generation of mutation data matrices from protein sequences, 1991. [Online] Available : <http://bioinformatics.oxfordjournals.org/content/8/3/275.abstract>, December 6, 1991 .
- [5] M. Yamuna and K. Karthika, "Point mutation determination using graph theory", *Der Pharmacia Lettre*, vol. 7, no. 7, pp. 58-66, 2015.
- [6] DNA Sequencing. [Online] Available: https://www.google.co.in/?gws_rd=ssl#q=dna+sequencing.
- [7] DNA Sequencing Example. [Online] Available: <http://media-2.web.britannica.com/eb-media/64/389x450x47664-004-7088EE3D.jpg.page.speed.ic.hlxrPIfLPM.jpg>.
- [8] Mutation. [Online] Available: https://www.google.co.in/?gws_rd=ssl#q=mutation+.
- [9] Mutation Example. [Online] Available: <https://classconnection.s3.amazonaws.com/826/flashcards/512826/png/picture11323327410672.png>
- [10] Types of Mutation. [Online] Available: http://2.bp.blogspot.com/-dywXBfCewIE/Uyq_EpsS72I/AAAAAAAAAG8/iLotTN5BuFc/s1600/mutation0.png
- [11] Deletion Mutation. [Online] Available: <http://zube.brinkster.net/SBI4U14/Genetics/Heredity/Reading/DeletionMutationGEN.gif>.
- [12] Deletion Mutation Example. [Online] Available: https://en.wikipedia.org/wiki/Deletion_%28genetics%29#/media/File:Deletion_vectorized.svg.