Mutation: A graphical view

P. Renjini Raveendran[1]* and S. Beena[2]

Abstract

Graph theory is a branch of Mathematics that has its application in almost all areas of studies. Phylogenetic analysis deals with the study of evolutionary history of organisms. The concepts of graph theory can be applied to biology to get new insight into the topics.

Mutation is a change in the nucleotide sequence of a short region of a genome. If genomes evolve by gradual mutations, then the difference in nucleotide sequence between a pair of genomes indicates how recently they share a common ancestor. By comparing genomes, it is possible to work out the evolutionary relationship between them. Molecular phylogenetics deals with it. In the 1980s DNA based phylogenetic began to be used on a large scale. This work deals with the graphical analysis of some topics related to the concept of Mutation. The graphical view of the concepts enables the study of Mutation which in turn help to study various evolution events from Directed Graphs. The study can be extended to other graphs also.

Keywords

Directed Graphs, Eulerian Digraphs.

AMS Subject Classification

05C90.

[1]*Department of Mathematics, All Saints' College, University of Kerala, Thiruvananthapuram, Kerala, India.*
[2]*Principal, MMNSS College, Kottiyam, Kollam-691571, Kerala, India.*
* Corresponding author, [1]renjiniraveendran7@gmail.com

Contents

1. Introduction

A graphical representation always enables a person to make the concepts clear. Graphs that give importance to direction or orientation can be represented with the help of Directed Graph or Digraph.

Mutation is a change in the nucleotide sequence of a short region of a genome. Point Mutations are mutations that re-place one nucleotide with the other. Point mutation can be of transitions or transversions.Transitions are due to replacement of a purine with another purine or replacement of a pyrimidine with another pyrimidine. Transversions are replacement of a purine with a pyrimidine or a pyrimidine with a purine. There are certain diseases caused by Point Mutation. Sickle Cell Anaemia is one such disease.

This work deals with the graphical analysis of biological ideas discussed above. Here the concepts of Directed Graphs were used. This work relates some concepts of Mutation to that of Directed Graphs. The graphical conversion of concepts like Point Mutation (Transitions and Transversions), graphical view of codons in Sickle Cell Anemia and their graphical explanations are given in this work.

2. Preliminaries

1. A directed graph $D = (V,A)$ is an ordered pair, where V is a nonempty set of elements called the vertex set and A is a set of ordered pairs of elements of V called the arc set. If $a = (u,v)$ is an arc in D, then a is said to join u to $v \cdot u$ is called the initial vertex and v is called the terminal vertex of A.

2. Indegree of a vertex v is the number of arcs towards v, denoted by $d^-(v)$

3. Outdegree of a vertex v is the number of arcs away from v, denoted by $d^+(v)$

4. A phylogeny describes the evolutionary history of a set of taxa.

5. Eulerian digraph is a directed graph that contains an Euler Circuit

6. Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two graphs. The union of G_1 and G_2 denoted by $G_1 U G_2$ is a graph G with vertex set $V = V_1 U V_2$ and edge set $E = E_1 U E_2$. If G_1 and G_2 have some vertex in common, union is denoted by $G_1 + G_2$.

7. The Deletion of a proper subset S of vertices of G results in a subgraph of G containing vertices of G not in S and edges of G not incident on a vertex in S. If S consists of a single vertex v, then deletion is denoted by $G - v$.

8. Theorem(Characterization of Euleriandigraphs) (defined by [1]) For a connected digraph D the following statements are equivalent:

   (a) Digraph D is Eulerian.
   (b) For every vertex v in D, $d^-(v) = d^+(v)$
   (c) Digraph D is the union of edge disjoint directed cycles.

3. Graphical explanation of point mutation

Mutation is a change in the nucleotide sequence of a short region of a genome. Point Mutations are mutations that replace one nucleotide with the other.

Let us consider a nucleotide sequence as *ABCDEFGHIJ*. If any nucleotides say *H* is replaced by *N* then the sequence becomes *ABCDEFGNIJ*.

Graphically Point Mutation corresponds to the deletion of the vertex *H* from the graphical structure and then another vertex *N* took that place forms edges with the preceding and the following vertices as below.
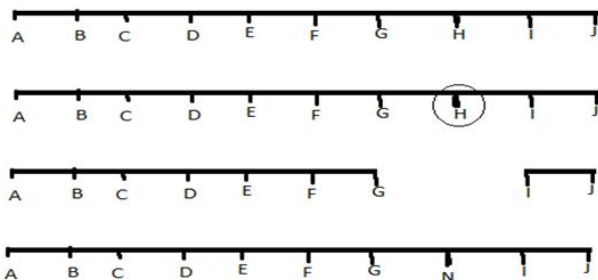

Figure 1

4. Graphical version of transition and transversion

Point mutations are divided into two categories: transition and transversion

(a) Transition:
These are Purine(Adenine(A) and Guanine(G)) to Purine or Pyrimidine(Cytosine(C) and Thymine(T) in DNA) to Pyrimidine changes. That is

$$A \to G, G \to A, C \to T, T \to C$$

Graphically this can be represented as $\mathbf{T}_1$:

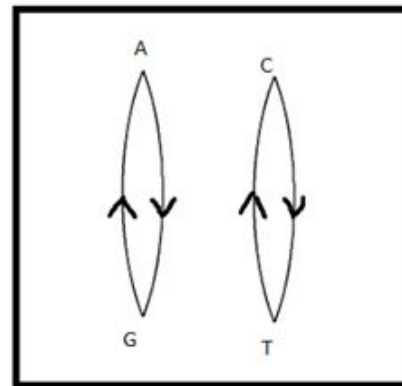Transition can be represented as a Digraph with 4 vertices and 4 arcs.

Figure 2. Transition

(b) Transversion
These are purine to pyrimidine or Pyrimidine to purine changes.

$$A \to C, A \to T, G \to C, G \longrightarrow T, C \to A, C \to G, T \to A, T \to G$$

Graphically Transversion can be represented as $\mathbf{T}_2$.

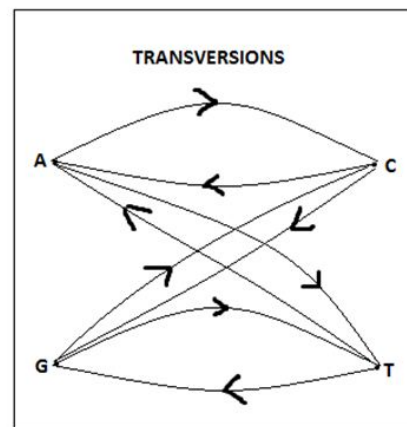This is a directed graph with 4 vertices and 8 arcs.

Figure 3

5. Combining the graphs of transition and transversion

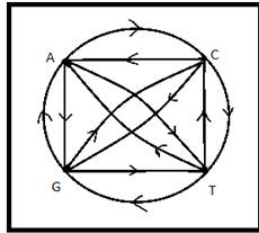The above two directed graphs can be combined to form a single Digraph *P* as follows:

Figure 4

5.1 Some important deductions

1. *P* is the union of $\mathbf{T}_1$ and $\mathbf{T}_2$

2. For each vertex A, C, T and G, indegree and outdegree are same for each vertex *v*, that is $d^-(v) = d^+(v) = 3$

3. *P* is Eulerian

4. *P* can be represented as the union of edge disjoint directed cycles. *P* is the union of these 6 edge disjoint directed cycles.
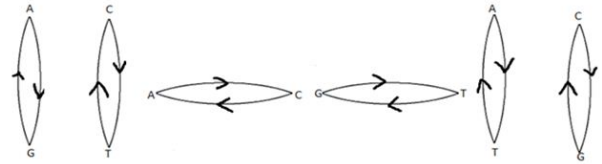

Figure 5

6. Sickle cell anaemia- a disease caused by point mutation

Sickle-cell anaemia is caused by a point mutation in the β-globin chain of haemoglobin which causes the hydrophilic amino acid glutamic acid to be replaced by the hydrophobic amino acid valine at the sixth position.

Sequence for Normal Haemoglobin.

AUG	GUG	CAC	CUG	ACU	CCU	GAG	GAG	AAG	UCU	GCC	GUU	ACU
START	Val	His	Leu	Thr	Pro	Glu	Glu	Lys	Ser	Ala	Val	Thr

Sequence for Sickle Cell Haemoglobin

AUG	GUG	CAC	CUG	ACU	CCU	GUG	GAG	AAG	UCU	GCC	GUU	ACU
START	Val	His	Leu	Thr	Pro	Val	Glu	Lys	Ser	Ala	Val	Thr

The only change is $\mathbf{GAG} \rightarrow \mathbf{GUG}$ in the 6th position. Here in this work each of the codon are represented by means of a directed graph. Each of the codon on three letters can be represented as directed graphs. So the normal and diseased sequence of codons takes the graphical structure as below.
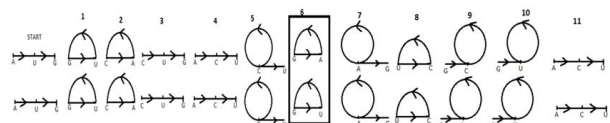
First row denotes the graphical view of normal sequence of codon and the second row denotes the diseased sequence of codons. Graphically considering normal and the diseased codons, there will be

1. 2 changes in arcs (In normal the arcs are GA and AG. In diseased codon the arcs are GU and UG. All other arcs remain the same.)

2. 1 change in vertex (The vertex *A* in normal codon is replaced by *U* in diseased codon)

Figure 6

Here in this work,the graphical view of Point mutation, Transitions and Transversion etc, were expressed in terms of Directed Graphs. Then analyse the graphical properties of those obtained graphs. Then express the codon change in Sickle cell anaemia in terms of Directed Graphs and explain the change in codon graphically.

7. Conclusion

The graphical analyses of biological process are possible by converting the biological concepts to that of graph theory.

References

[1] Dr G. Suresh Singh, Graph Theory, PHI Private Limited, New Delhi 2010.

[2] Narsing Deo, Graph Theory with Applications to Engineering and Computer Science, PHI Learning, 1974.

[3] Daniel H Huson, Regula Rupp, Celine Scornavacca, Phylogenetic Networks: Concepts, Algorithms and Applications, Cambridge University Press, 2010.

[4] T. A. Brown, Genomes, 2nd Edition, BIOS Scientific Publishers Limited.