Mathematical Sciences Letters
An International Journal

http://dx.doi.org/10.18576/msl/110105

Fractals- Cantor Sets in DNA

S. Pavithra and G. Jayalalitha[*]

Department of Mathematics, VELS Institute of Science, Technology and Advanced Studies, Pallavarm, Chennai-600117, India

Received: 12 Oct. 2021, Revised: 12 Nov. 2021, Accepted: 12 Dec. 2021

Abstract: The Chronicle of some recent progress to pertain the Fractals concepts in Biology and Medicine will set one's sight on. DNA Sequence is a complex biological mechanism in different Genes. The domain of DNA Mapping has been said that DNA Sequence is Homeomorphisms. The rupture of the DNA Sequence is mainly due to protein base mismatch is considered as Random Cantor Set by Fractal behavior. *Complexity is the number of base pairs of unique or non-repeating nucleobases in a given segment of DNA Sequence or component of the Genome*. If a few of the DNA Sequence is repeated its Length (L) will be different in the same Genes. Different lengths of Rupture DNA Sequence will give different Dimension. The Dimension of DNA Sequence of Different Genome is considered as a Random Cantor Set and it satisfies the property of statistical self-similarity of the Fractals. DNA Sequence is a Fractal. Obliging to prognosis in Various Medical fields.

Keywords: DNA, Fractal, Cantor Set and Function, Gene, Homeomorphisms.

1 Introduction

1.1 Fractals

The history of fractals dates back to 1975 when fractals were discovered by Benoit Mandelbrot. He explained them as being geometric shapes that when divided into parts, each part would be a smaller term Fractals which was derived from the Latin word fractus. The Latin word fractus means broken or fractured. A fractal is a never-ending pattern. Fractals are infinitely complex patterns that are self-similar across different scales. They are created by repeating a simple process over and over in an ongoing feedback loop. Geometrically, they exist in between our familiar dimensions. Fractal patterns are extremely familiar since nature is full of Fractals(Figure 3) [7]. The Study of Fractals gives us a quantitative language to describe the myriad of self-similar shapes found in the natural world. Fractals need not be natural objects; they can be human-made (Figure 1) and can also unfold in time in addition to space. Fractals Method has been successfully used to study many problems in Physics, Mathematics, Engineering, and Finance and even in Biology [9]. For instance: Trees, Rivers, Coastline, Mountains, Clouds, Seashells, the population of cities, the distribution of the number of links into web pages all can be usefully viewed as fractals. Fractal objects at first blush seem intricate and complex (Figure 2). A few simple tools and ideas for analyzing fractals prove to be surprisingly powerful and flexible [11,12].

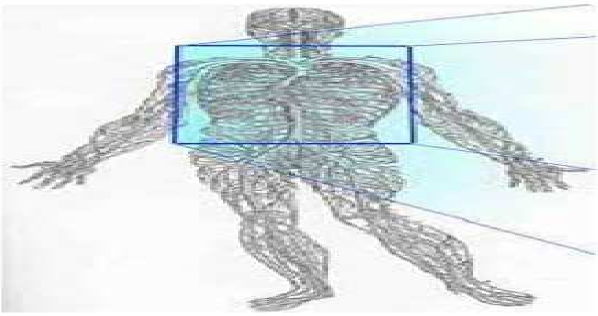


Fig. 1: Fractal body

1.2 Deoxyribonucleic Acid (DNA)

Deoxyribonucleic acid or DNA is a molecule that contains the instruction an organism needs to develop,

---

[*] Corresponding author e-mail: g.jayalalithamaths.sbs@velsuniv.ac.in

Fig. 2: Complex and Mathematical Fractals

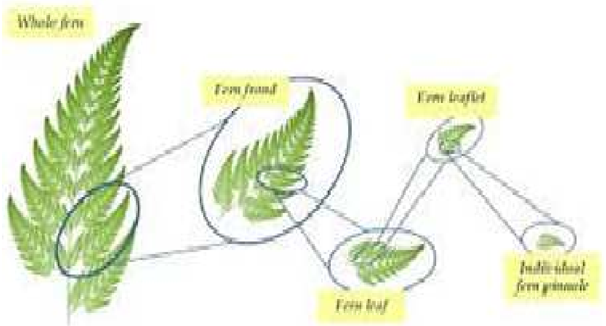


Fig. 3: Fractals in Nature

live and reproduce. These instructions are found inside every cell and are passed down from parents to children.

A nucleic acid that carries the genetic information in cells and some viruses, consisting of two long chains of nucleotides twisted into a double helix and joined by hydrogen bonds. The two DNA strands are also known as polynucleotides since they are composed of similar monomeric units called nucleotides (Figure 5). Each nucleotides composed of one four nitrogen-containing nucleobases (adenine (A), thymine(T), cytosine(C), guanine(G)) (Figure 5). The nitrogenous bases of the two separate polynucleotide strands are bound together according to base-pairing rules (A with T and C with G), with hydrogen bonds to make double-stranded DNA. The complementary nitrogenous bases are divided into two groups' pyrimidines and purines. In pyrimidines ate thymine and cytosine, the purines are adenine and guanine [1]. The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar rings, which are known as the $3'$ and $5'$ carbons, the prime symbol is used to distinguish these carbon atoms from those of the base to which the deoxyribose forms a glycosidic bond. Any DNA strand therefore normally has one end at which there is a phosphoryl attached to the $5'$ carbon of a ribose (the $5'$ phosphoryl) and another end at which there is a free hydroxyl attached to the $3'$ carbon of a ribose (the $3'$ hydroxyl)[13]. The orientation of the $3'$ and $5'$ carbons along the sugar-phosphate backbone confers directionality (sometimes called polarity) to each DNA strand [2].

2 Preliminaries

2.1 Cantor Function

The Cantor function $C : [0,1] \to [0,1]$, let X be in [0, 1] and obtain the function C(X).The functions extend to the whole interval [0, 1] by making it constant on these removed intervals. The function obtained by extension is called cantor's function[9].

2.2 Closed Walk

A walk is considered to be closed if the starting vertex is the same as the ending vertex, $v_0 = v_k$.

2.3 Closed Set

A set is closed if and only if it contains all of its limits.

Example: The unit interval [0,1] is closed in the metric space of real numbers, and the set $[0,1] \cap Q$ of rational numbers between 0 and 1 (inclusive) is closed in the space of rational numbers, but $[0,1] \cap Q$ is not closed in the real numbers.

2.4 Convergence

Let (X_n) be a sequence of real numbers. The sequence (X_n) is said to converge to a real number as if for all $\varepsilon > 0$, there exist n in N such that $|x-a| < \varepsilon$ for all n∈N

2.5 Absolutely Convergence

A sequence $\sum X_n$ is called absolutely convergent if $\sum |X_n|$ is convergent.

2.6 Homeomorphisms

A function $f : X \to Y$ between two topological spaces is a homeomorphism

if it has the following properties

- Bijection (one-to-one and onto)
- Continuous function
- Inverse Exist

3 Cantor Set

Cantor's set is the set C left after this procedure of deleting the open middle third subinterval is performed infinitely many times. Repeating this process finally cantor dust is formed. Cantor dust is a multi-dimensional version of the cantor set. It can be formed by taking a finite Cartesian product of the Cantor set, making it a cantor space. Like the Cantor set Cantor dust has zero measure. Cantor's set has no isolated points. In any neighborhood of points in Cantor's set, there is another point from the cantor's set. In topology, a set that is compact and has no isolated points is called a perfect set.

3.1 Theorem

Prove the following theorem

i The cantor set is non-empty.
ii The cantor set is closed and nowhere dense.
iii The cantor set C is perfect and disconnected.

(i) Proof:

Cantor Set is defined as the intersection of a decreasing sequence of sets each of which is defined as the union of a finite number of closed intervals. Each trisection of I_n to form I_{n+1} leaves exactly two points. For example, from algorithm removing $(1/3, 2/3)$ from [0, 1] leaves the points P_0=1/3 and P_1=2/3. Since the Cantor set is the infinite intersection of each I_n, C contains the endpoints of each such subinterval and is non-empty. It is infinite and shown in Figure 4.

3.2 Algorithm

I_0: Begin with the interval [0, 1].
I_1: Divide [0, 1] into 3 subintervals and delete the open middle subinterval $(1/3, 2/3)$.
I_2: Divided each of the 2 resulting intervals above into 3 subintervals and delete the open middle subintervals$(1/9, 2/9)$ and $(7/9, 8/9)$. Continue this procedure indefinitely. At each step, delete the open middle third subinterval of each interval obtained in the previous step.

(ii) Proof:

Let $C = \cap_{K\geq 1}^{\infty} E_k$, Where [0,1] = $E_0 \subset E_1 \subset$... is a decreasing sequence of closed sets, E_k a union of 2^k disjoint closed kth- level basic intervals.

The lengths of the intervals are random. The 'Random cantor set' C is statistically self- similar, in that the distribution of the set $C \cap I$ is the same as that of C.

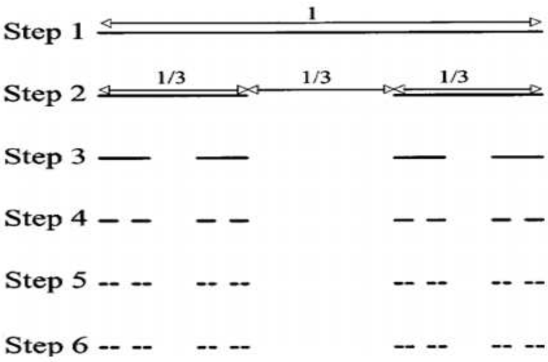


Fig. 4: Cantor set

$$C = \cap_{n=1}^{\infty} \cap_{K=0}^{3^{n-1}-1} \left([0, (3k+1)/3^n] \cup [(3k+2)/3^n]\right)$$

From theorem 3.1(i), C is the intersection of closed sets, which implies that C is itself closed. Furthermore, by the theorem 3.1(i), the cantor set contains no intervals of non-zero length, and so int$(C) \neq \emptyset$.

(iii) Proof:

Fix any $\varepsilon > 0$ and point $P \in C$. Let n∈N is sufficiently large such that $1/3^n < \in$. Then, P is guaranteed to be in one of the intervals (I_n for some n∈N) that makes up C, each length $1/3^n$. The endpoints of the cantor set in this interval are infinite in number and all contained in the open interval (P-ε, P+ε).so, P is a cluster point of C, $M_{\in}$(P) containing an infinite number of points. Since we are considering any $P \in C$, C is perfect.

Furthermore, this interval I_n is closed in R and so in the Cantor set C as well. Since $I_n^C = C/I_n$ consists of a countable number of closed intervals it is itself closed. It can then represents C as the disjoint union of two open sets (C $\cap$ I_n) and (C $\cap I_n^C$), the results begin that the Cantor set C is disconnected [5].

4 Cantor Set In DNA

A random self-similar object of fractal dimension less than 1 embedded on a one-dimensional (1-D) line, such as the cantor set cites ref:1,ref:8. It is considered in this study random Cantor set since a DNA sequence appears in the form of random chains. Visually, Cantor sets contain mostly space and some filled sites in all length scales[14, 15].

If, for example, describing the "G" reaction, there would be a mixture of the following products of synthesis [6]:

5'-GAATGTCCTTTCTCTAAGTCCTAAG
3'GGAGACTTACAGGAAAGAGATTCAGGATTCAGGAGGCCTACCATGAAGATCAAG5'

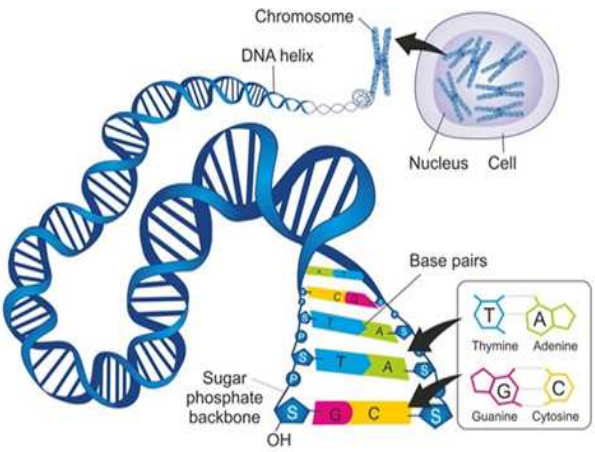


Fig. 5: DNA Sequence

5’-GAATGTCCTTTCTCTAAGTCCTAAGTCCTCCG
3’GGAGACTTACAGGAAAGAGATTCAGGATTCAGGAGGCCTACCATGAAGATCA
5’-GAATGTCCTTTCTCTAAGTCCTAAGTCCTCCGG
3’GGAGACTTACAGGAAAGAGATTCAGGATTCAGGAGGCCTACCATGAAGATCA
5’-GAATGTCCTTTCTCTAAGTCCTAAGTCCTCCGGATG
3’GGAGACTTACAGGAAAGAGATTCAGGATTCAGGAGGCCTACCATGAAGATCA

5’-GAATGTCCTTTCTCTAAGTCCTAAGTCCTCCGGATGG
3’GGAGACTTACAGGAAAGAGATTCAGGATTCAGGAGGCCTACCATGAAGATCA

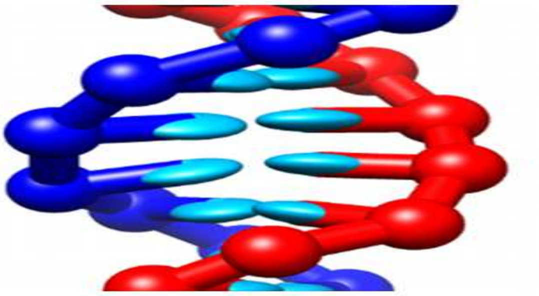
5’-GAATGTCCTTTCTCTAAGTCCTAAGTCCTCCGGATGGTACTTCTAG
3’GGAGACTTACAGGAAAGAGATTCAGGATTCAGGAGGCCTACCATGAAGATCAAG

For different lengths of DNA Sequence different Genome sizes and dimensions is obtained are shown in Table 1[11].

4.1 *Computation of Different Dimension of Genome to Binary Number*

For Example:

Converting the dimension of DNA from base 10 to base 2. i.e., convert 0.375_{10} to base 2
Solution:
Let N be a complexity and a_i, r_i be normal constants
2N = 2 x 0.375 = 0.75
$a_1 = 0$, $r_1 = 0.75 - |0.75| = 0.75$
$2r_1 = 2 \text{ x } 0.75 = 1.5$
$a_2 = 1$, $r_2 = 1.5 - |1.5| = 0.5$
$2r_2 = 2 \text{ x } 0.5 = 1$
$a_3 = 1$, $r_3 = 1\text{-}1 = 0$
Since, $r_3 = 0$. therefore, $0.375_{10} = 0.011_2$



beginfigure[h]

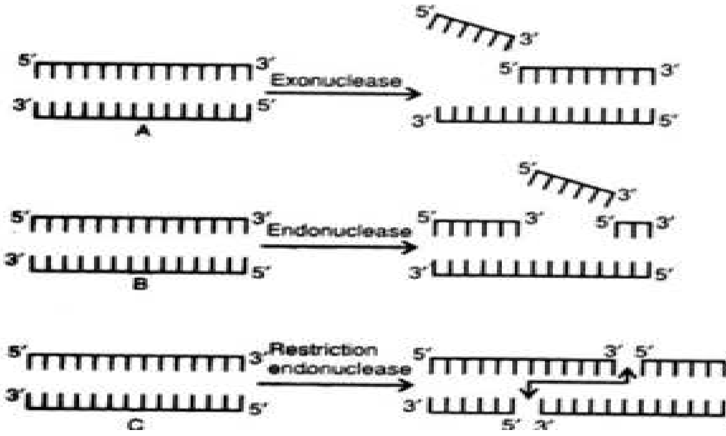


Fig. 6: Ruptured DNA Sequences in Different Genome

Proceeding with different Dimension following iterations is obtained shown in Table 3. In Different iteration for Different Genome Size at Different Dimension, the Random Cantor set lies between [0, 1].

Table 1: Different Type of Genes

ORGANISM TYPE	ORGANISM	GENOME SIZE (BASE PAIRS)
Virus	Bacteriophage	3569
Virus	HIV	9749
Plant	Paris japonica	150,000,000,000
Fungus-yeast	Saccharomyces cerevisae	12,100,000
Insect	Fruit fly Honey bee	175,000,000 236,000,000
Mammal	Homo sapiens	3,289,000,000
Fish	Types of puffer fish	385,000,000

Table 2: Different Length of Ruptured DNA Sequence

LENGTH	DNA SEQUENCE
120	ACACAGCTATGCGTGAAAAGGGCTAGCCCGGCTTCGATACAGA CACCGGAAGAATCTGACGTCGTCTCTAGCTGCGGGTCATCGG CGACTCTCTCTTATCCTGATTCTCAGTCAGTCC
220	GCTTCCGAGATTTTGTCCGGCGAAAAGCACGACCGAGTGCCG CTGCGTCAAGAAATCCCACGCTACTCGCCTTATTGCAAGAAC GCGTGTACTGTATGACATTTGGAGGCTCAATCTTCTGGCACA GTAGGGCCAGTATCCAACCGTGCCATGGCGTGGCGCTAGTTT CTAATGCTACTCGACTATTCAAAGCACTGTTTACAATTGTCGAT
260	GGACTCTTGCTCGAGAGTTGCAGAGCCGTACTGTTTAGTCAAC AAAAATCTATCATTTATTGTTGCAATGAAAGGAGAAGTAGCAT ATCGATTTTCCGATATACTATTACTAAGCACCTTTAGTAATTG ATCGCGCACTTGGGGGTCGAACGATCACGCGAACTACAGA GCTTCCGCTACCATAATCCGCACCGGGCCGTTTCCCTTGC GCAGATAGCCCGAGCTCGGTAAGCCATCACGGGGTTCTTTT GATAACAAAC

Table 3: Different Dimension of Cantor of Different DNA Sequence

DIMENSION Of DIFFERENT GENOME(D)	$2N=2(D)$	1st ITERATION	2nd ITERATION	3rd ITERATION
0.631	1.262	0.262	0.524	0.048
0.83	1.66	0.66	0.32	0.64
0.68	1.36	0.36	0.72	0.44
0.97	1.94	0.94	0.88	0.76
0.98	1.96	0.96	0.92	0.84
0.99	1.98	0.98	0.96	0.92

By 2.6 DNA sequence satisfies bijection, continuous function and Inverse exist hence (Figure 7) DNA Mapping is homeomorphisms. DNA Sequence has a Cantor function [3]. Due to Mapping for different Genome different length of Sequence is obtained.

4.2 Application of DNA Medical Field

DNA fingerprints can help in diagnosing (Figure 10) inherited disorders like cystic fibrosis, hemophilia, Huntington's disease, Familial Alzheimer's Disease, Thalassemia, Sickle cell anemia and many others in

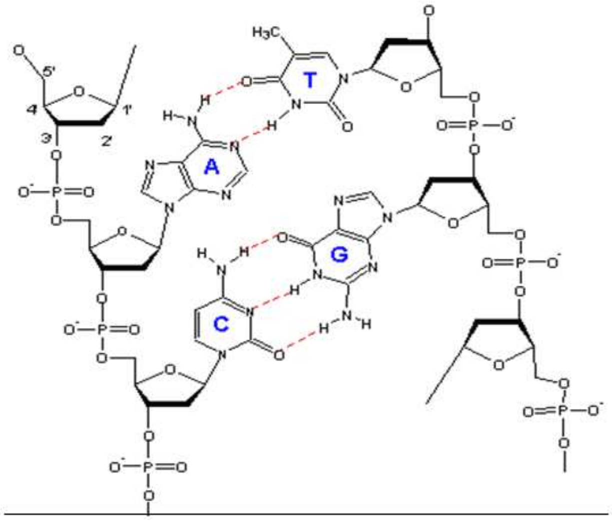


Fig. 7: DNA Mapping

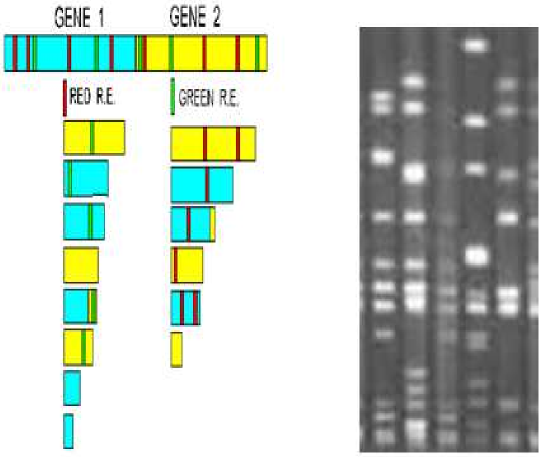


Fig. 8: Two genes with palindrome sites for two different restriction enzymes of different person’s

Fig. 9: Diagnosing Fingerprints in DNA Sequence

prenatal and new-born babies [4,12]. The mutation in the particular gene is that these genetic disorders change the RFLP pattern (Figure 9).

5 Results

At different dimensions the different genome sizes of DNA Sequences will get ruptured. This Ruptured DNA Sequence is considered as Fractals [10]. It can be analyzed by Cantor Set. By 2.2 DNA Sequence is closed walk. The Cantor set contains as many points as the interval from which it is taken, yet it contains no interval of nonzero length. The irrational numbers have the same property, but the Cantor set has the additional property of being closed, so it is not even dense in any interval, unlike the irrational numbers which are dense in every interval. By theorem 3(iii) Cantor set is convergent and disconnected. Different length of a DNA sequence is convergent because of the random Cantor set. The rupture of DNA sequence implies an absolute convergence [3,6,9]. Table 2 for different lengths of DNA Sequence in different Genome Sizes has formed at different Random Cantor set (Figure 4). It has different Dimension values such as 0.631, 0.83, 0.68, 0.97, 0.98, 0.99 etc., as shown in the Table 3. By the method 4.1 Different Dimension of Genome of different Length (L) which gives a binary number that also a Cantor Set. This is due to complexity (N). In Table 3 for different Dimension values at different iteration, it gives a non-integer value which is lies between [0, 1]. The Cantor set is the prototype of the fractal. It is self-similar because it is equal to two copies of itself if each copy is shrunk by a scaling factor of 3 (Figure 4). The Cantor set invariant up to Homeomorphisms. By using 2.6 and Figure 6, Ruptured DNA Sequence is a mapping and hence it satisfies Homeomorphisms. By theorem 3 DNA Sequence is non-empty, perfect, and closed. This DNA Sequence has a Cantor Set and fractal.

6 Conclusion

This document naturally does not bring an end to the use of fractals in the field of DNA Sequence. It only proves the applicability of some new fractal parameters that we have defined. DNA process is a never-ending process. Due to protein base mismatches DNA Sequence ruptures. So DNA mapping is formed. Since DNA sequence is itself a Cantor set it can be concluded that DNA Sequence is a Fractal. By using various methods in fractals can help us to identify the dimension of cells from there it can diagnose diseases in humans. The practical application of DNA fingerprinting is numerous. DNA fingerprints can help diagnose inherited disorders.

References

[1] Alberts. B., et.al., *Molecular Biology of the cell Garland publishing INC*,New York, 1994.
[2] Almirantis. Y., *Journal of Theoretical Biology*, Vol.196, 1999.
[3] Bittner. H.R., Wlczek. P., Sernetz. M., *Characterization of fractal biological objects by image analysis,* Acta. Stereo, Vol.8 (1), 1989.
[4] Cross. S.S., *The application of fractal geometric analysis to microscopic images. Micron*, Vol.25 1994.
[5] Kenneth Falconer., *Fractal geometry University of St Andrews*, 2013.
[6] Lewin. B., *Genes VI Oxford University press*, Oxford, 1999.
[7] Michael F. Barnsley, *Fractals Everywhere*, 2012.
[8] Provata. A., and Almirants. Y., *Fractal Cantor Patterns in the Sequence Structure of DNA journal*, 1999.
[9] Vicsek. T., *Fractal Growth phenomena World Scientific*, Singapore, 1989.
[10] Mandelbrot. B.B., *The Fractal Geometry of Nature, Freeman*, San Francisco, CA, 1982.
[11] Carlo Cattani., *fractals and hidden symmetries in DNA*, Mathematical Problems in Engineering, Vol.2010, pp31, 2010.
[12] Clifford. A. Pickover., *The pattern Book Fractals, Art and Nature*, World Scientific publication, singapore, pp 30, 2000.
[13] Michael Frame., Amelia Urry., *Fractals World: Grown, Built and Imagined*, Yale University press, U.K.Office, pp 29, 58 (2016).
[14] Aliahamad Behzad., Dinesh. k., Poosapandi Arjunan., Sridhar., *Fractals: Application in Biological Signalling and Image Processing*, A science publisher books, Australia, pp 7 and 19, 2017.
[15] Sadana., Ajit., *DNA Biosensor*, Elsevier Publication, UK, 2018.