Understanding Functional Protein-Protein Interactions of ABCB11 and ADA in Human and Mouse

Antara Sengupta
Dept. of MCA
MCKV Institute of Engineering
Howrah, India
Antara.sngpt@gmail.com

Sk. Sarif Hassan
Dept. of Mathematics
University of Petroleum and Energy Studies
Dehradun, India
sarif.hassan@icts.res.in

Pabitra Pal Choudhury
Applied Statistics Unit
Indian Statistical Institute
Kolkata, India
pabitrapalchoudhury@gmail.com

Abstract—Proteins are macromolecules which hardly act alone; they need to make interactions with some other proteins to do so. Numerous factors are there which can regulate the interactions between proteins [4]. Here in this present study we aim to understand Protein-Protein Interactions (PPIs) of two proteins ABCB11 and ADA from quantitative point of view. One of our major aims also is to study the factors that regulate the PPIs and thus to distinguish these PPIs with proper quantification across the two species Homo Sapiens and Mus Musculus respectively to know how one protein interacts with different set of proteins in different species.

Index Terms—Quantitative Understanding, Protein properties, Molecular properties, Protein-Protein Interactions (PPIs), Fractal Dimension.

INTRODUCTION

Proteins are the workhouses where most of the biological processes in a cell take place. Proteins are macromolecules and need associations of some other proteins to participate in various essential molecular processes within a cell and there is a talk about Protein-Protein Interactions (PPIs) [7]. Numerous factors are there which can regulate the PPIs. Abnormal PPIs can lead to develop the basis of diseases like cancers. Not only that but drugs also target PPIs for interfacial inhibition. As an example Brefeldin A (BFA) [1] acts as inhibitor and attacks macromolecular complexes when the complex is in a transition state(structurally and energetically unbalanced) and goes for drug binding[3]. Certain biological properties are there in gene as well as protein level which play significant role in transcriptions, protein structure formations, and gene expressions and can have significant role in protein network formation.

Moreover, proteins fold spontaneously into complicated three-dimensional structures which are essential for several biological activities. Mainly the driving energy for this folding process comes from the hydrophobic effect, Van der Waals forces, and salt bridges at specific binding domains on each protein [6]. The strength of the binding depends on the size of the binding domain. Leucine zipper is a common surface domain that can provide stable protein-protein interactions. It consists of α-helices on each protein that bind to each other in a parallel fashion through the hydrophobic bonding of regularly-spaced leucine residues on each α-helix that project between the adjacent helix peptide chains. Because of the tight molecular packing, leucine zippers provide stable binding for multi-protein complexes [7].

Quantitative understanding of genes/proteins refers to its unique characterizations or a numeric vector corresponding to each sequence which would act as their signature by finding some quantitative attributes. The quantification essentially captures how the nucleotides or amino acids are arranged in the sequence [9]. Here in this paper it is tried to make quantitative analysis of the genes ABCB11 and ADA and the genes which are functionally attached with them across the two species Homo Sapiens and Mus Musculus. The analysis has been made with genes and Proteins at their sequence levels and as well as the physical properties of those protein to study the factors that regulate the PPIs and thus to distinguish these PPIs with proper quantification across the two species to know how one protein interacts with different set of proteins in different species.
A. Model Representation
We code any given DNA sequence to numeric sequence for further analysis. There are various ways of doing so. Here we have coded as A=1, T=2, G=3 and C=4 and thus a DNA sequence would be transformed to a sequence of corresponding numeric sequence.

B. Data Set Specification
In this paper as the PPIs of the proteins ABCB11 and ADA for the species Sapiens and Mus Musculus are taken from String Database which is shown below. It is to be noted that the associations of the proteins are purely functional associations. It can be observed that the proteins ABCB11 and ADA both have some proteins in the networks which are common for both the species, whereas some proteins are specifically participating in the specified species not in other.

TABLE I ABCB11 and ADA PPIs.Tick defines presence of the protein the species as mentioned.

| Proteins in the Network of | Homo | Mus |
|----------------------------|------|-----|
| ABCB11                     | ✓    | ✓   |
| ABCC11                     | ✓    | ✓   |
| ATP5B1                     | ✓    | ✓   |
| BAAT                       | ✓    | ✓   |
| Nf0b2                      | ✓    | ✓   |
| NR5A2                      | ✓    | ✓   |
| NRIH4                      | ✓    | ✓   |
| Slco10a1                   | ✓    | ✓   |
| Slco1a1                    | ✓    | ✓   |
| SLCO1A2                    | ✓    | ✓   |
| SLCO1A4                    | ✓    | ✓   |
| SLCO1A6                    | ✓    | ✓   |
| SLCO1B1                    | ✓    | ✓   |
| SLCO1B2                    | ✓    | ✓   |
| SLCO1B3                    | ✓    | ✓   |

II. METHODOLOGY

Factors that regulate the PPIs are known qualitatively. To model the regulatory factors mathematically it is necessary to characterize the PPIs in DNA and Protein sequence levels which are performed as follows:-

A. Quantitative understanding of DNA sequence

Quantitative understanding of a gene in molecular level can be done by applying several mathematical parameters [13][9] which can derive some data and can verify some fact. In this paper, to build quantitative model it is tried to find out the underlying geometries of DNA structure and the hidden geometrical rules, that is to capture spatial ordering of bases across the DNA and Protein so that we can make a mapping between those geometrical rules with biological activities of those proteins.

Quantitative understanding of DNA sequence is taken place through following phases:-

1) Generating Indicator Matrix and Calculating Fractal Dimension of the Indicator Matrices

In Mathematics the term ‘Indicator’ is a numerical measure of a quality or characteristic of some aspect of a program; evidence that something is occurring, that progress is being made. The notion of indicator matrix and its characterization through fractal dimension was proposed by Carlo Cattani [13][2]. DNA sequences have four basic components (A = adenine, C = cytosine, G = guanine, T = thymine) which is defined as four alphabets A, C, G, T respectively. Let us consider $D \equiv \{A, C, G, T\}$ be the set of nucleotides and $x \in D$ where, $x$ is any alphabet of D. A DNA sequence is the finite symbolic string $S = \{x_1, x_2, \ldots, x_N\}$ such that $x_h \in \{A, C, G, T\}$, for $h = 1, 2, \ldots, N$ and $x_h \in D$ be the value $x$ at the position $h$. According to C. Cattani the Indicator Matrix can be characterized through fractal dimension as follows,

$$ f: S \times S \to \{0, 1\} \text{ such that } f(x_h, x_k) \equiv \begin{cases} 
1 & \text{if } x_h = x_k \\
0 & \text{if } x_h \neq x_k 
\end{cases} $$

where $x_h, x_k \in S$

So now it is possible to easily describe a NxN sparse binary matrix from an indicator matrix of N length, which may be written as,

$$ \text{M}_{hk} = f(x_h, x_k), x_k \in S, h, k = 1, 2, 3, 4 $$

But it is not possible to differentiate zeros formed by distinct base pairs. So slight modification is needed into it, which is given below [7],

$$ f: S \times S \to \{1, 2, 3, 4\} \text{ such that } f(x_h, x_k) \equiv \begin{cases} 
1 & \text{if } x_h = x_k \\
2 & \text{if } x_h \neq x_k, \text{ and } x_k \neq x_i \\
3 & \text{if } x_h = x_k, x_i \neq x_k \text{ and } x_i \neq x_k \\
4 & \text{if } x_h \neq x_k, x_i \neq x_k \\
\end{cases} $$

So the matrix $\text{M}_{hk}$ will be decomposed into four binary matrices namely, A1, A2, A3, A4 and corresponding four binary images will be made for each DNA sequence.

The fractal dimensions of those indicator matrices have been calculated using “Box-Counting” method [17]. Box-Counting is a method of collection of data for analyzing complex pattern like dataset, images etc. and breaking them into smaller part or
pieces and analyze them individually. Fractal dimensions for each indicator matrix are derived using BENOIT software.

Indicator matrices for ABCB11 and ADA and the genes in their networks are made for both the species. Fractal Dimensions (FDs) of four indicator matrices are also calculated for each species which are shown in Supplementary file 1 and Supplementary file 2 respectively.

### TABLE II. Interval of Fractal Dimensions and the names of genes which have similar characteristics to ABCB11 and ADA respectively

| Species       | Gene Name                                   | FD of DNA Walk | Genes having similar characteristics |
|---------------|---------------------------------------------|----------------|---------------------------------------|
| Homo Sapiens  | ABCB11                                      | (1.78626, 1.90321) | BAAT                                  |
| Mus Musculus  | and genes in network                        | (1.7942, 1.88461) | Scl01a2                               |
| Homo Sapiens  | ADA                                         | (1.76487, 1.94016) | PNP                                   |
| Mus Musculus  | and genes in network                        | (1.7824, 1.88461) | Nt5e                                  |

1. The gene ‘SLCO1A2’ and ‘alb’ have similar characteristics of Abcb11 of Homo Sapiens and Mus Musculus respectively for FDs of all four indicator matrices.
2. The gene NT5C and Nt5e3 have similar characteristics of ADA of Homo Sapiens and Mus Musculus respectively for FDs of all four indicator matrices.

### 2) Fractal Dimension of DNA Walk

As defined in the [9][13], we can define DNA walk as,

$$a_n = \sum_{i=1}^{m} f(A, x_i, e_i) = \sum_{i=1}^{m} f(T, x_i, e_i) = \sum_{i=1}^{m} f(C, x_i, e_i) = \sum_{i=1}^{m} f(G, x_i, e_i).$$

It has been resulted by plotting \((W_n, V_n)\) as two functions where \(W_n \equiv \sin \alpha_n - \sin \beta_n\) and \(V_n \equiv \sin \gamma_n - \sin \delta_n\) are defined.

The Fractal Dimensions of all the genes are deduced using Box Counting method.

### TABLE III. Interval of Fractal Dimensions of DNA Walk and the names of genes which have similar characteristics to ABCB11 and ADA respectively

| Species       | Gene Name | FD of DNA Walk | Genes having similar Characteristics |
|---------------|-----------|----------------|---------------------------------------|
| Homo Sapiens  | ABCB11 and genes in network | (1.78626, 1.90321) | BAAT                                  |
| Mus Musculus  | and genes in network | (1.7942, 1.88461) | Scl01a2                               |
| Homo Sapiens  | ADA and genes in network | (1.76487, 1.94016) | PNP                                   |
| Mus Musculus  | and genes in network | (1.7824, 1.88461) | Nt5e                                  |

### 3) Poly-string Mean and Standard Deviation:-

A gene is a string constituting of different permutations of the base pairs A, C, T and G where repetition of a base pair is allowed. It is possible to classify the sequences based on the ordering of poly-string mean of A, C, T, and G in the string [12][13].

$$\text{Mean}_{\text{p}}(x) = \frac{1}{m} \sum_{i=1}^{m} x_i$$

Where, \(x_i \in \{A, T, C, G\}\), \(i=1,2,3,\ldots,m\) and \(m\) is the length of longest poly-string over the string.

The poly-string mean and SD are calculated for all the genes of the PPI of ABCB11 and ADA respectively and the detail result is given in the supplementary files 1 and 2.

### TABLE IV. The poly-string mean and SD for all the genes of the PPI of ABCB11 and ADA

| Species       | Gene Name | Classifications based on poly-string mean | Classifications based on poly-string SD |
|---------------|-----------|------------------------------------------|----------------------------------------|
| Homo Sapiens  | ABCB11    | 5                                        | 5                                      |
| Mus Musculus  | and genes in network | 7                                        | 8                                      |
| Homo Sapiens  | ADA and genes in network | 6                                        | 8                                      |
| Mus Musculus  | and genes in network | 7                                        | 7                                      |

There are 256 possible poly-strings of length four using four nucleotides A, T, C, G. It can be observed that the genes in PPIs of ABCB11 for Homo Sapiens and Mus Musculus have been classified into 5 and 7 classes respectively according to poly-string mean whereas, the genes in PPIs of ADA have been classified into 5 and 8 classes respectively.

In the case of poly-string SD the genes in the PPIs of ABCB11 for Homo Sapiens and Mus Musculus have been classified into 6 and 7 classes respectively. The genes in the PPIs of ADA for
Homo Sapiens and Mus Musculus have been classified into 8 and 7 classes respectively.

B. Quantification of Bio-physical Properties

Followings are the properties of DNA sequences which are calculated by using Oligo Calc. These properties include the physical properties of oligonucleotides:

1) Physical Constants
   i. Length
   ii. Molecular weight
   iii. GC Contents

2) Melting Temperature (TM) Calculations
   i. Basic
   ii. Salt Adjusted

3) Thermodynamic Contents Conditions: 1M NaCl at 25°C at pH 7
   i. Rlnk
   ii. DeltaH
   iii. DeltaG
   iv. DeltaS

The Bio-physical Properties are calculated for all the genes of the PPI of ABCB11 and ADA respectively are derived and the detail result is given in the supplementary files 1 and 2.

1. Overall observation for the species of Homo sapiens speaks that although the quantitative difference of the physical properties of the genes ABCB11 and SLCOA2 are small but GC count of Physical constants and Melting Temperature Calculations are same of ABCB11 and SLCO1B1 are same.

2. From the result it is observed for the species of Homo sapiens that majorly the quantitative differences of physical constants and Thermodynamic Contents Conditions of the genes ADA and NT5C2 are smaller but in the case of GC count and Melting Temperature (TM) Calculations the genes ADA and ADORA1 are having very less differences.

3. From the result it is observed for the species of Mus Musculus that quantitative differences of physical constants and Thermodynamic Contents Conditions of the genes ADA and NT5C2 are smallest and gene abcb11 have similarity of physical properties with Nr5a2.

4. The result as a whole enlightens that, physical properties of both of the species can’t be equal.

C. Quantification of Protein Structural Properties

Here, we extend our quantitative analysis to the protein structural level and which is given below.

The protein properties which are considered here are viz. Alpha Helix (Chou Fasman), Beta Strand (Chou Fasman), Parallel Beta Strand, Anti Parallel Beta Strand, Molecular Beta Strand, Total Beta Strand, Coil, Hydrophobicity (Aboderin) and Polarity Grantham. The plotting of protein properties has been made using MATLAB (Bioinformatics Tools) and the fractal dimension has been deducted using BENOITTM software.

D. Quantitative comparisons of DNAs in terms of 2D graphical representations based on triplets

We know that Codon consists of three adjacent nucleotides which codes for a particular amino acid. Thus 64 codons are there which codes for 20 amino acids and 3 codons are there which cause termination of protein synthesis. Now each codons have their corresponding triplets of DNA which we can get from the DNA

From the result we can get the range of FDs within which the Protein Properties of the proteins ABCB11 and ADA and the proteins in their networks are lying for the 2 species.

The Protein Properties of ABCB11 are more similar with the properties of the proteins ATP5B1 for the species of Homo Sapiens but in case of Mus Musculus it has similarity with ‘alb’. Whereas, it is worth enough to say that for both the species the Protein Properties of ADA are more similar with the properties of the protein ADK except the properties of Molecular Beta Strand and Parallel Beta Strand, which are more similar with PNP. But as a whole the quantitative understanding of the protein properties reveal that the protein PNP can have a continuous mapping with the gene ADA.

### Table IV. Ranges of protein properties of each species

| Protein Name       | Homo Sapiens | Mus Musculus | Homo Sapiens | Mus Musculus |
|--------------------|--------------|--------------|--------------|--------------|
|                    | ABCB11 & proteins in network | ADA & proteins in network |
| Alpha Helix        | (1.93681, 1.94316) | (1.93031, 1.94357) | (1.93681, 1.94316) | (1.93957, 1.94348) |
| Antiparallel beta strand | (1.9361, 1.94326) | (1.93087, 1.94293) | (1.9361, 1.94326) | (1.93944, 1.94329) |
| Beta Sheet         | (1.93688, 1.94354) | (1.93061, 1.94296) | (1.93688, 1.94354) | (1.93942, 1.94352) |
| Coil               | (1.93972, 1.94426) | (1.93398, 1.94385) | (1.93972, 1.94426) | (1.94115, 1.94422) |
| Hydrophobicity     | (1.93548, 1.94321) | (1.92922, 1.94312) | (1.93548, 1.94321) | (1.93924, 1.94322) |
| Molecular Beta Strand | (1.93906, 1.94411) | (1.93416, 1.94378) | (1.93906, 1.94411) | (1.93989, 1.9443) |
| Parallel Beta Strand | (1.93717, 1.94366) | (1.92921, 1.94295) | (1.93717, 1.94366) | (1.93832, 1.94368) |
| Polarity Grantham  | (1.93577, 1.94368) | (1.93072, 1.94332) | (1.93577, 1.94368) | (1.94009, 1.94406) |
| Total Beta Strand  | (1.9368, 1.94338) | (1.94271, 1.94273) | (1.9368, 1.94338) | (1.93908, 1.94345) |
sequences. If we define a mapping \( \Psi \) to map each triplet to different weight, for any two pairs of triplets \( (X_1, Y_1) \) and \( (X_2, Y_2) \), where \( X_1, Y_1, X_2, Y_2 \) are all triplets, if the corresponding codons of \( X_1 \) and \( Y_1 \) code for the same amino acid but corresponding codons of \( X_2 \) and \( Y_2 \) code for different amino acid, then that satisfy the following rule:

\[
| \Psi (X_1) - \Psi (Y_1) | < | \Psi (X_2) - \Psi (Y_2) |
\]

The mapping has been illustrated below in the tables V and VI:

**TABLE V. Codons and their corresponding Amino Acids.**

| Codon | Corresponding Amino Acid | Codon | Corresponding Amino Acid |
|-------|--------------------------|-------|--------------------------|
| GCU, GCC, GCA, GCG | Alanine | UGG | Tryptophan |
| CUC, CUC, CUG, UUA/UUG | Leucine | CAU, CAC | Histidine |
| AGU, AGA, AGG | Arginine | UAU, UAC | Tyrosine |
| AAA, AAG | Lysine | AUG | Methionine |
| GAU, GAC | Aspartic acid | GUU, GUC, GUA, GUG | Valine |
| AUG | Methionine | AAA, UAG, UGA | Stop Codons |
| UUU, UUC | Phenylalanine | UCU, UCC, UCA, UCG, AGU, GC | Serine |
| UGU, UGC | Cysteine | CAA, CAG | Glutamine |
| CCU, CCC, CCA, CCG | Proline | ACU, ACC, ACA, ACG | Threonine |
| GGU, GGC, GGA, GGG | Cysteine | | Glycine |

**TABLE VI. Triplets and their corresponding weight \( (\Psi) \).**

| Codon | Corresponding Triplet | Weight \( (\Psi) \) | Codon | Corresponding Triplet | Weight \( (\Psi) \) |
|-------|-----------------------|------------------|-------|-----------------------|------------------|
| GCU | GCT | 1.1 | CUU | CTT | 11.1 |
| GCC | GCC | 1.2 | CUC | CTC | 11.2 |
| GCA | GCA | 1.3 | CUU | CTA | 11.3 |
| GCG | GCG | 1.4 | CUG | CTG | 11.4 |
| CGU | CGT | 2.1 | UUA | TTA | 11.5 |
| CGC | CGC | 2.2 | UUG | TTG | 11.6 |
| CGA | CGA | 2.3 | AAA | AAA | 12.3 |
| CGG | CGG | 2.4 | AAG | AAG | 12.4 |
| AGA | AGA | 2.5 | UUU | TTT | 13.1 |
| AGG | AGG | 2.6 | UUC | TTC | 13.2 |
| GAU | GAT | 3.3 | CCU | CCT | 14.1 |
| GAC | GAC | 3.4 | CCC | CCC | 14.2 |
| AAA | AAT | 4.1 | CCA | CCA | 14.3 |
| AAC | AAC | 4.2 | CCG | CCG | 14.4 |
| UGU | UGT | 5.1 | UCU | TCT | 15.1 |
| UGC | TGC | 5.2 | UCC | TCC | 15.2 |
| GAA | GAA | 6.1 | UCA | TCA | 15.3 |
| GAG | GAG | 6.2 | UCG | TCG | 15.4 |
| CAA | CAA | 7.1 | AGU | AGT | 15.5 |
| CAG | CAG | 7.2 | AGC | AGC | 15.6 |
| GGU | GGT | 8.1 | ACU | ACT | 16.1 |
| GCC | GCC | 8.2 | ACC | ACC | 16.2 |
| GGA | GGA | 8.3 | ACA | ACA | 16.3 |
| GGG | GGG | 8.4 | ACG | ACG | 16.4 |
| CAU | CAT | 9.1 | UGG | TGG | 17.3 |
| CAC | CAC | 9.2 | UAU | TAT | 18.1 |
| AUU | ATT | 10.1 | UAC | TAC | 18.2 |
| AUC | ATC | 10.2 | GUU | GTT | 19.1 |
| AUA | ATA | 10.3 | GUC | GTC | 19.2 |
| UAA | TAA | 21.1 | GUU | GTA | 19.3 |
| UAG | TAG | 21.2 | GUG | GTG | 19.4 |
| UGA | TGA | 21.3 | AUG | ATG | 20.1 |
Let $G=g_1, g_2, g_3, \ldots, g_n$ be a DNA primary sequence, where $g_i=\{A, T, C, G\}$ for any $i=1, 2, \ldots, n$, so that, $G=t_1, t_2, \ldots, t_m$ be the corresponding triplet, where $m=n/3$. Then to plot $G$ into a plot set, we can define mapping $\Theta$ as follows [18],

$$\Theta(G)=\{(1, \Psi(t_1)), (2, \Psi(t_2)), \ldots, (m, \Psi(t_m))\} \quad (2)$$

PPI Networks of ABCB11 and ADA for both the species reflects that some proteins are there, which are present in the networks of both species, like in case of ABCB11 (Abcb11, Alb, Baat, Nr0b2, NRIH4, Slco10a1) are common, whereas in case of ADA (ADA, ADK, DCK, NT5C, NT5C2, NT5E, PNP) are common. In this paper all DNA sequences of those genes are represented through 2D graph as graphical representation of DNA sequence provides viewing, sorting and even comparing various gene structures in simpler way. Thus quantitative comparisons of those DNAs are made possible in terms of 2D graphical representations.

One example of quantitative comparisons between DNAs of ABCB11 of Homo Sapiens and Mus Musculus in terms of 2D graph is shown below in figure 1:-

![Quantitative comparisons of ABCB11 of Homo Sapiens and Mus Musculus](image)

**Fig 1:** Quantitative comparisons of ABCB11 of *Homo sapiens* and *Mus musculus* of first 30 triplets

### III. CONCLUSION

In this paper, quantitative understanding and analysis of Functional Protein-Protein Interactions of ABCB11 and ADA in Human and Mouse has been taken place. Characterization and comparison of functionally associated PPIs through the light of different parameters have shown a statistical view of those factors which ultimately are responsible for network constructions of those two proteins for each species. Moreover, as PPI network of same protein may defer from species to species, the paper reflects a quantitative comparative view of PPI networks of same proteins but for different two species.

### IV. APPENDIX

A. Supplementary file1 – *HOMO SAPIENS* DB.xls

B. Supplementary file2 – *MUS MUSCULUS* xls

### REFERENCES

[1] Robineau S., Chabre M., Antonny B. Proc. Natl Acad. Sci. USA 2000:97:9913-9918.

[2] Yu Zu-Guo, (2002) "Fractals in DNA sequence analysis", Chinese Physics, 11 (12), 1313-1318.

[3] Pommier Y., Cherfils J. Trends Pharmacol. Sci. 2005, 26: 138-145.

[4] Javier De Las Rivas, Celia Fontanillo, "Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks" PLoS Comput Biol. 2010 Jun; 6(6): e1000807.

[5] http://www.nlm.nih.gov/medlineplus/breastcancer.htm

[6] Hydrophobic Interactions in Proteins Brian W Matthews, University of Oregon, Eugene, Oregon, USA.

[7] https://www.lifetechnologies.com/in/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-protein-protein-interaction-analysis.html

[8] http://string-db.org

[9] Antara Sengupta, Sk. Sarif Hassan and Pabitra Pal Choudhury. Article: Quantitative Understanding of Breast Cancer Genes. ICA Proceedings on National Conference cum Workshop on Bioinformatics and Computational Biology NCWBCB(1):15-17, May 2014.

[10] Mandelbrot, "The Fractal Geometry of Nature." W.H. Freeman and Company, ISBN 0-7167-1186-9.

[11] Avnir (1998) “Is the geometry of Nature fractal”.

[12] K Develi, T Babadagli (1998) "Quantification of natural fracture surfaces using fractal geometry" Math.

[13] Carlo Cattani, “Fractals and Hidden Symmetries in DNA” Mathematical Problems in Engineering Volume 2010, Article ID 507056

[14] Sk. S. Hassan, P. Pal Choudhury and A. Goswami, (2012), "Underlying Mathematics in Diversification of Human Olfactory Receptors in Different Loci ". Under disciplinary Sciences-Life Sciences December 2013, Volume 5, Issue 4, pp 270-273

[15] R. H. C. de Melo and A. Conci, (2008) “Succolarity: Defining a Method to calculate this Fractal Measure,” ISBN: 978-80-227-2856-0 291- 294.

[16] PJ Deschavanne, A Giron, J Vilain, G Fagot (1999) "Genomic signature: characterization and classification of species assessed by chaos game representation of sequences" Mol. Bio. Evo. 16 (10), 1391-1399.

[17] Sk. Sarif Hassan, Pabitra Pal Choudhury and Aritra Bose, (2011), "A Quantitative Model for Human Olfactory Receptors ", Nature Precedings, npre20126967-2.