Analysis of Human Olfactory Receptor system using Fractal Dimension

Ranjeet Ku Rout\(^a\), Saiyed Umer\(^b\), Smitarani Pati\(^c\), Harveer Singh Pali\(^d\)

\(^a\)Department of Computer Science & Engineering  
National Institute of Technology Srinagar, Hazratbal-190006, J&K, India  
\(^b\)Department of Computer Science & Engineering,  
Aliyah University, Kolkata, India 
\(^c\)Department of Instrumentation and Control Engineering  
Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, Punjab-144011, India  
\(^d\)Department of Mechanical Engineering  
National Institute of Technology Srinagar, Hazratbal-190006, J&K, India  
E-mail: ranjeetkumarout@nitsri.net

Abstract. An extensive diverseness of chemicals having discrete odours can be identified by Humans. On the basis of various deviation of development of proposed model on olfactory receptor of Human database has been proposed. There are very large family of ORs sequences for example rat, Human, chimpanzee and mouse etc. In this paper we work up on the Human OR genes, as very few experimental/quantitative works have been done for different species. It is very difficult to identify a DNA sequence whether belongs to a particular family or not, without any valid biological validation. In our viewpoint, an appropriate understanding of these Human olfactory receptor system DNA sequence s is essential to validate, a particular DNA sequence is in a family or not.

1. Introduction

In the environment prodigious odor molecules are present which are encountered by the Olfactory Receptors (OR) [1, 2]. For detecting these molecules Olfactory receptor neurons are pledged. There are some common molecular features that shared by the entire odor molecules. By using various bio-informatics model the repository system of OR genes are recognized. It is observed that the OR genes of rat and Human is in between by 1200 and 400 respectively [3]. The DNA (Deoxy ribo-nucleic acid) sequence is a long chain of many base pairs present in all living organism of any species. The DNA sequences are ranging from 105 to 109 numbers of base pairs [4]. Any two nucleotide combined together in order to form the base pairs. There are four nucleotide named present in DNA as A(Adenine), T(Thiamine), C(Cytosine) and G(Guanine). The main duty of the genome is to absolutely recognize the imperative biological process and other similarities of the DNA structure for different species observed by the nucleotide sequences, as stated above. Within very few years, there is a serious research is going on in the field of biological, for the study of genome, in which DNA sequences are capable of producing different structure of an organism. The DNA sequences associated with these processes are having a few covert interrelation, symmetries, or law between them. Hence, our most imperative target is to
disclosing the pattern of the protein structures or interrelation among different nucleotide. It is helpful in the analysis of each DNA sequence characterization. Well-known methods are also used to elucidate these secret quantitative features of DNA by the help of fractal dimensions and mathematical morphology [26, 17]. In this paper to get the quantitative results, a model has been proposed by using various aspects. This prospective proposed model has been trained by taking 496 ORs as a training data set from the family of human. As per the content of Foundation for Biomedical Research, most of the lab animals are mice and rats. For several reasons scientists are having assurance on the mice and rats, rat and mics are moderate and can be easily sustain up to of two-three years of life span, which is used for observing associate rodents within abridged life span. To purchasing this in a large amount are quite economical.

To identify all the genes of Human National Institutes of Health and Energy department of USA are managing a known Human Genome Project (HGP) [5, 6] [7, 8]. The main aim of this project is to have grant researchers for doing research and to understand the prototypes of Human genomes. Humans are functioning based on the coded instructions in DNA sequences. Some significant research have done in the area of extracting the existing knowledge of DNA sequences. In the field of mathematical biology, medicine, biotechnology and life sciences various mathematical models like Boolean networks and some mathematical models [24, 25] are used for modeling and extracting information form different networks like gene regulatory network (GRN) and protein protein interaction networks (PPI). Currently, gene therapy is acting as the most likely research area in medical science [7]. It is the best technique which uses genes for different treatments or prevents various disease. This therapy technique is plays a promising roll in the field of treatment for a numerous diseases, till date the technique is some how risky and in order to make an assurance that it will be safe and effective researches are going on. Now a day, this therapy technique is being tested for their treatments [7, 8]. There are many technical challenges associated with the treatment of diseases has to be tackled before considering gene therapy as a practical approach. For doing that, the first step is to have the accentual appprehension of genomes. This will help us in a decisive depiction of a specific DNA. The accentual approach of genes will be a preservative as key signature of a DNA sequence. In this paper, by using Fractal Geometry [10] as a mathematical measurement for olfactory receptors (ORs) of Humans [9, 11]. Clusters have been generated and analyzed through these quantitative parameters.

The rest part of the article is arranged as: In Section 2, data specification and Proposed method for DNA sequence analysis have been discussed. In Section 3, classification based on k-mean algorithm has been discussed. In section 4, the proposed method has been discussed based on the result. Summary of this article is drawn in Section 5.

2. Mathematical preliminaries of fractal geometry

In this section two different mathematical parameter fractal dimension and Box counting method for calculating fractal dimension have been discussed.

2.1. Fractals

To describe Nature completely with intricate details Euclidean geometry is not enough [12], [13], [14], [15] as it is very complex and also many irregular shapes provide a much better representation of many natural phenomena. Fractal geometry generally used as a framework to study different irregular shape. There are many use of fractal and multi-fractal geometry mainly in the field of Physics, Economics, Biology, Medicine and Computer Science; among many other areas [16]. The current applications are, among others, image compression, computer graphics and special effects in movies, music generation and pattern classification and so on. The conversational models are used by the biological researchers. Still scientists are identifying that the characterization of many natural designs can be characterized by using fractal geometry.
The general characterization of biological systems and processes are done by many sub structural levels, which has the common pattern similarity repeated in a constantly decreasing cascade[18].

DNA sequences are also self-similar. Some biologists having their view point that the evolutionary relationships in animals can be settled by the DNA Fractal properties. Possibly the fractal geometry will be used by the biologist in future as to develop complete models of the patterns and processes observed in nature. In this dissertation, Fractal geometry has been used as an useful tool for the analysis of tertiary protein structure i.e. in deciphering the inner content of the three dimensional structure of proteins[19].

2.2. Box Counting Dimension

The computation of the number of cells need for covering the whole object, by the help of cell grids which are varying in size. Basically, this process is carried out by super imposing the traditional grids on an object and calculating the number of cells covered by the object. In this logarithm of \( N(r) \), the number of occupied cells, versus the logarithm of \( \frac{1}{r} \), where \( r \) is the size of one cell, gives a line whose gradient corresponds to the box dimension [14], [18], [20]. For detection of the FD of \( Sk_n \), the FD by the help of Box-Counting method is defined (1).

\[
Dim_{box}(Sk_n) = \lim_{r \to 0} \frac{\log N(r)}{\log \frac{1}{r}}
\]

3. Date Used and Specification

Olfactory receptor DNA sequence and protein sequences of different species has been reported in [21]. it is also known as ORDB or OR system database. (https://senselab.med.yale.edu/ordb/info/mouseorsequence). In this research work all DNA sequences are accumulated from ORDB. The complete nucleotide sequence of OR10K2 is represented in Figure 1.

Figure 1. Shows a DNA sequence of Olfactory receptor system (OR10K2)

4. The proposed Method for representation of primary protein sequences

In this section, a novel method has been proposed to figure out the various parameters with the help fractal dimension [?, ?, 10]. Which is used to generate some properties among all OR DNA sequences in the quantitative manner. The generation of each quantitative parameter index for OR sequences are well vindicated as follows:
4.1. Graphical representation of DNA sequence

In this section, we discussed about the representation of DNA sequence into a graph. In the first quadrants of the graph pyrimidines (C and T) and in second quadrant purines (A and G) are represented respectively. The four nucleotide vectors are shown in below.

\[
\begin{align*}
A & \rightarrow (-1, -1, -1) \\
C & \rightarrow (-1, +1, +1) \\
G & \rightarrow (+1, -1, +1) \\
T & \rightarrow (+1, +1, -1)
\end{align*}
\]

Here, we emphasize on the general arrangement of the four nucleotide in the Cartesian coordinate points. The graphical representation of the DNA sequence of Human Olfactory system is shown in figure 2. The three dimensional graph construction procedure is demonstrated in Algorithm 1.

**Illustration:** A DNA sequence is consist of four nucleotide G, A, T, and C is defined as \( S = 'ATCGTGTATGCATGCATGCTGATTGC' \). Such sequence shown in string \( S \) is converted as in 3D graph using the co-ordinates as \(-1\) and \(1\) in the four quadrant as shown in the figure 2. The olfactory receptor DNA sequence of \( OR_{4F29}, OR_{4F16}, OR_{4F05}, OR_{4G11P} \) and \( OR_{4GP} \) are shown in the figure 3 and the corresponding three dimensional representations are shown in Figure 3.

**Algorithm 1 :** Three dimensional graph representation of Human Olfactory Receptor System procedure.

**Input:** Olfactory Receptor DNA sequence of Human.

**Output:** The reconstructed Three Dimensional image of the DNA sequence.

1. Let \( X, Y, \) and \( Z \) are three array used to store the co-ordinates of a DNA sequence \( S \).
2. for \( i = 1 \) to \( \text{length}(S) \) do
3.   if \( S[i] = 'A' \) || \( S[i] = 'C' \) then
4.       \( X[i] = -1 \)
5.   end if
6.   if \( S[i] = 'G' || S[i] = 'T' \) then
7.       \( X[i] = 1 \)
8.   end if
9.   if \( S[i] = 'A' \) || \( S[i] = 'G' \) then
10.  \( Y[i] = -1 \)
11. end if
12. if \( S[i] = 'C' || S[i] = 'T' \) then
13.  \( Y[i] = 1 \)
14. end if
15. if \( S[i] = 'A' || S[i] = 'T' \) then
16.  \( Z[i] = -1 \)
17. end if
18. if \( S[i] = 'C' || S[i] = 'G' \) then
19.  \( Z[i] = 1 \)
20. end if
21. end for

5. Result interpretation and Discussion

In this section, the Fractal Dimension of all the graph generated through the proposed method of all DNA sequences (600) has been generated. This results using the explained method for
Figure 2. Shows a three dimension representation of the DNA sequence S

Figure 3. Shows the corresponding three dimension graph of the DNA sequences shown in figure ??
Figure 4. Shows the fractal dimension of all the graph of the DNA sequences of human olfactory sequence.

Figure 5. Shows obtained cluster of fractal dimension of all the graph of the DNA sequences of human olfactory sequence.

all olfactory receptor system sequence of human are depicted. The three dimensional graph has been depicted for all the Olfactory Receptor DNA sequences of Human. Then the fractal dimension for each of the 3D graph has been computed. The FD index values for all the human ORs sequence are ranging from 1.1569 to 1.569 which is shown below in the figure 4. The obtained Fractal Dimension values of Human OR for the parameter are used for classification using k-mean clustering technique. Six different clusters have been formed. The cluster obtained has been shown in figure 5.

6. Conclusions
In this present work, the whole Human olfactory receptor system DNA sequence has been taken into consideration. In the other hand, a novel 3D graphical representation method has been proposed for quantification of DNA sequence on the basis of fractal dimension of the 3D graphs and a set of cluster has been made based on fractal dimension. The FD values for all the human ORs sequence are ranging from 1.1569 to 1.569.
References

[1] Malnic, B., Godfrey, P. A., & Buck, L. B. (2004). The human olfactory receptor gene family. Proceedings of the National Academy of Sciences, 101(8), 2584-2589.

[2] Crasto, C., Singer, M. S., & Shepherd, G. M. (2001). The olfactory receptor family album. Genome biology, 2(10), reviews1027-1.

[3] Niimura, Y. (2009). Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents. Human genomics, 4(2), 107.

[4] Elloumi, M., & Zomaya, A. Y. (2011). Algorithms in computational molecular biology: techniques, approaches and applications (Vol. 21). John Wiley & Sons.

[5] Fortun, M. A. (2001). The human genome project: Past, present, and future anterior. In Science, History and Social Activism (pp. 339-362). Springer, Dordrecht.

[6] Collins, F. S., Morgan, M., & Patrinos, A. (2003). The Human Genome Project: lessons from large-scale biology. Science, 300(5617), 286-290.

[7] Collins, F. S., Morgan, M., & Patrinos, A. (2003). The Human Genome Project: lessons from large-scale biology. Science, 300(5617), 286-290.

[8] Collins, F. S., & McKusick, V. A. (2001). Implications of the Human Genome Project for medical science. Jama, 285(5), 540-544.

[9] Friedmann, T., & Roblin, R. (1972). Gene therapy for human genetic disease?. Science, 175(4025), 949-955.

[10] Develi, K., & Babadagli, T. (1998). Quantification of natural fracture surfaces using fractal geometry. Mathematical Geology, 30(8), 971-998.

[11] Zhang, X., & Firestein, S. (2002). The olfactory receptor gene superfamily of the mouse. Nature neuroscience, 5(2), 124.

[12] Falconer, K. (2004). Fractal geometry: mathematical foundations and applications. John Wiley & Sons.

[13] Barnsley, M. F. (2014). Fractals everywhere. Academic press.

[14] Avnir, D., Biham, O., Lidar, D., & Malcai, O. (1998). Is the geometry of nature fractal?. Science, 279(5347), 39-40.

[15] Edgar, G. (2007). Measure, topology, and fractal geometry. Springer Science & Business Media.

[16] Biswas, M. K., Ghose, T., Guha, S., & Biswas, P. K. (1998). Fractal dimension estimation for texture images: a parallel approach. Pattern Recognition Letters, 19(3-4), 309-313.

[17] Pati, S., & KumarPahuja, S. (2018, January). Geodesic Distance in Comparison Between Tertiary Protein Structures. In 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 856-859). IEEE.

[18] Hassan, S. S., Nayak, B. K., & Choudhury, P. P. Studies on Fractal Formation and Its Morphological Analysis with Some Novel Applications.

[19] Mandelbrot, B. B. (1983). The fractal geometry of nature (Vol. 173, p. 51). New York: WH freeman.

[20] Develi, K., & Babadagli, T. (1998). Quantification of natural fracture surfaces using fractal geometry. Mathematical Geology, 30(8), 971-998.

[21] Hassan, S. S., Choudhury, P. P., Daya Sagar, B. S., Chakraborty, S., Guha, R., & Goswami, A. (2015). Quantitative description of genomic evolution of olfactory receptors. Asian-European Journal of Mathematics, 8(03), 1550043.

[22] Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. IEEE Transactions on Pattern Analysis & Machine Intelligence, 7, 881-892.

[23] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern recognition letters, 31(8), 651-666.

[24] Rout, R. K., Maity, S. P., Choudhury, P. P., Das, J. K., Hassan, S. K., & Pandey, H. M. (2019). “Analysis of Boolean functions based on interaction graphs and their influence in system biology.” Neural Computing and Applications 1-19

[25] Rout, R. K., Hassan, S. K., Sindhawi, S., Umer S. and Pandey, H. M. (2019) "Intelligent Classification and Analysis of Essential Genes Species Using Quantitative Methods." ACM Transactions on Multimedia Computing, Communications, and Applications.

[26] Rout, R. K., Pal Choudhury, P., Maity, S. P., Daya Sagar, B. S., & Hassan, S. S. (2018). Fractal and mathematical morphology in intricate comparison between tertiary protein structures. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 6(2), 192-203.