

Quantitative Description of Genomic Evolution of Olfactory Receptors

Sk. S. Hassan, P. Pal Choudhury, B. S. Daya Sagar, Senior Member, IEEE

S. Chakraborty, R. Guha and A. Goswami

Abstract—We investigate how evolutionary network is associated among Human, Chimpanzee and Mouse with regards to their genomic information. We provide a quantitative description of genomic evolution through the fractal and mathematical morphology. We have considered olfactory receptors for our case study. These olfactory receptors do function in different species with the subtle differences in between the structures of DNA sequences. Those subtle differences could be exposed through intricate details of Fractal and Mathematical Morphology.

Index Terms— Olfactory Receptor, Fractal dimension, Morphological Skeleton, Bifurcation Dimension.

I. INTRODUCTION

HUMANS recognize a gigantic variety of chemicals as having distinctive odors. Odor perception initiates in the nose, where odorants are detected by a large family of olfactory receptors (ORs) [1]. ORs which constitute the largest gene superfamily in the human genome are the basis for the sense of smell. There appear to be about $30,000 \pm 40,000$ protein-coding genes in the human genome [2]. However, the genes are more complex, with more alternative splicing generating a larger number of protein products. To gain insight into the mechanisms underlying odor perception, and amount of complexities and quantitative differences in different genes of different species, we adumbrated here a quantitative descriptions of three OR sequences taken from

Manuscript received March 12, 2011.

Sk. S. Hassan is with Indian Statistical Institute, Kolkata, 700108 India (corresponding author to provide phone: +919609174055; e-mail: sarimif@isical.ac.in).

P. Pal Choudhury is with Indian Statistical Institute, Kolkata, 700108 India (e-mail: pabitra@isical.ac.in).

B. S. Daya Sagar, is with Indian Statistical Institute, Bangalore, India. (email: bsdsagar@isibang.ac.in).

S. Chakraborty is with Indian Statistical Institute, Kolkata, 700108 India (e-mail: shantanav89@gmail.com).

R. Guha is with Indian Statistical Institute, Kolkata, 700108 India (e-mail: ranita1990@gmail.com).

A. Goswami is with Indian Statistical Institute, Kolkata, 700108 India (e-mail: agoswami@isical.ac.in).

Human, Mouse and Chimpanzee for case study. There are many works done experimentally in different research labs across the globe but to the best of our knowledge, there is not so much of work done to decipher the quantitative content of genome. We believe the geometry and morphology of the DNA structure are very imperative aspect in studying their functions. So here we follow some of best known techniques to decipher those quantitative aspects of DNA through Fractals and Mathematical Morphology which are immensely used to study many problems in different branches of science and technology including the domain of Biology. In this paper, particularly we have captured the evolutionary connections among ORs with the help of their textural quantitative descriptions.

II. SOME BASICS AND FUNDAMENTALS

In this article we are about to use some standard techniques from Mathematical Morphology and Fractal. So let us warm up about some of the definitions from Mathematical morphology and Fractals.

A. Basics on Mathematical Morphology

Mathematical Morphology [3, 4] is a mathematics field specially meant for quantification shapes that are predominant in all natural sciences. The Mathematical morphology is based on axiomatic set theory and more relevantly lattice theory. Of late, the field has gained tremendous popularity out of its obvious uses in the field of image analysis which provides a quantitative description of geometrical structures. Morphology can provide boundaries of images, their skeletons, convex hulls, watershed for segmentation and many more [5, 6]. The primary aim is to extract important features from image data, from which a quantitative significant understanding of the topology of the image can be drawn. In this article, we apply certain morphological transformations essentially to generate distribution sequences of poly-strings that are present in DNA sequence templates. To perform these analyses, two fundamental morphological transformations employed include morphological erosion (to shrink) and morphological dilation (to expand) as explained in equations (1) and (2). For further descriptions reader may refer to [7, 8, 9, and 10]. Also we

have emphasized some other quantitative parameters as reviewed from a well written article [11].

B. Basics on Fractal

The word *Fractal* is derived from the Latin adjective *fractus*. The corresponding Latin verb *frangere* means ‘to break’ to create irregular fragments. In 1975, B. Mandelbrot coined the subject Fractal. The precise definition of “Fractal” according to Benoit Mandelbrot is, a set for which the Hausdroff Besicovitch dimension strictly exceeds the topological dimension [12, 13, 14].

Mandelbrot founded his insights in the idea of self similarity, requiring that a true fractal “fracture” or break apart into smaller pieces that resemble the whole. This is a special case of the idea that there should be a dynamical system underlying the geometry of the set. This is partly why the idea of fractal has become so popular throughout science; it is a fundamental aim of science to seek to understand the underlying dynamical properties of any natural phenomena. It has now become apparent that relatively simple dynamics, more precisely dynamical system can produce the fantastically intricate shapes and behavior that occur throughout nature. To understand those complex dynamics we need to know some of the fractal parameters. One of the fundamental fractal parameters is ‘Fractal Dimension’. There are several methods like box counting method, perimeter area dimension method and so on to compute fractal dimension of an object. In this paper we follow box counting method and is computed through well known software called BENOIT™.

The fractal dimension alone does not give an idea of what “fractals” are really about. So there was a real need of defining some other fractal parameters to clutch other features. One of the important parameter is ‘Succolarity’ which is really meant for the continuous density of the image / fractal. The primary notion of succolarity was given by Mandelbrot and later R. H. C. de Melo and A. Conci described the method to compute the succolarity of an image/object [15].

III. RESULTS AND DISCUSSION

We estimated four quantities indices namely poly string mean, standard deviation, Hurst exponent, fractal dimension based on morphological quantification and succolarity index for ORs of three species namely Human, Mouse and Chimpanzee. Without loss of generality, let us consider the olfactory receptors (ORs) OR1D2, CONTIG3463.6-1888, GA_x5J8B7W3YLM-7052533-7051808 of Human, Chimpanzee, and Mouse respectively for our case study. It is noted that first we have selected the OR OR1D2 from HORDE database and it was blasted in the NCBI database to get highly similar OR sequences in Chimpanzee, and Mouse and we found the CONTIG3463.6-1888, GA_x5J8B7W3YLM-7052533-7051808 of Human, Chimpanzee, and Mouse respectively.

A. Evolutionary Connection of ORs of Mouse and Chimpanzee with Human ORs

It should be noted that we have classified all the human ORs based on classification methodology on the poly-string mean and standard deviation as proposed in [15]. Using the same we have classified OR1D2 (Human), GA_x5J8B7W3YLM-7052533-7051808 (Mouse) and CONTIG3463.6-1888(Chimpanzee) and the results are shown in table-I. Also, we have considered a DNA as a one dimensional nucleotide sequence and let us define a map $T(A) = 00$; $T(T) = 11$, $T(C) = 01$ and $T(G) = 10$. So corresponding to a DNA sequence we now have a binary string of some fixed size (twice of the DNA sequence length). Let us discuss in brief what Hurst exponent is.

Hurst Exponent: The concept of Hurst Exponent was introduced by Harold Edwin Hurst and later in Fractal Geometry, B. Mandelbrot had modified it as a parameter of relative tendency of a time series to either strongly regress to the mean or ‘cluster’ in a direction [16]. In statistical terms, it is sometime referred to long range correlation of a one dimensional time series.

Let us consider a string $x = \{x_i\}_{i=1 \text{ to } n}$, and then we can define readily the following entities regarding the sequence as follows:

$$m_{x,n} = \frac{1}{n} \sum_{i=1}^n x_i \quad X(i, n) = \sum_{j=1}^i \{x_j - m_{x,n}\}$$

$$R(n) = \max X(i, n) - \min X(i, n) : 1 \leq i \leq n$$

$$S(n) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - m_{x,n})^2}$$

Then Hurst Exponent (H) is defined as $\left(\frac{n}{2}\right)^H = \frac{R(n)}{S(n)}$

In our case ‘n’ denote the length of the binary strings corresponding to each DNA sequence and x_i are the binary digits of the strings. The Hurst exponent varies between 0 and 1, with higher values indicating a smoother trend and less roughness. We then calculate the Hurst exponent (table- I) for the binary string.

<i>Olfactory Receptors</i>	<i>Class According to Poly-String Mean/SD</i>	<i>Hurst Exponent (H)</i>	<i>Maps to OR</i> <i>W.r.to H</i>
OR1D2	CGTA/CGAT	0.598911	OR1D2
GA_x5J8B7W3YLM-7052533-7051808	GCTA/GCAT	0.645594	OR4D2
CONTIG3463.6-1888	CGAT/ACTG	0.539152	OR3A3

Table-I: Evolutionary Connection of ORs with Human

The Mouse OR (GA_x5J8B7W3YLM-7052533-7051808) maps to a human OR OR4D2 based on classification and closest Hurst exponent. But it is to be noted that GA_x5J8B7W3YLM-7052533-7051808 is more similar to OR1D2. But as far as Hurst exponent is concerned (amount of long range correlation in the sequence) the mouse OR maps to OR4D2. In this connection, it is our strong conviction that, OR4D2 and OR1D2 are structurally similar in sequence despite the fact that they belong to different families as per HORDE qualitative classification. Also we discovered that mouse and human ORs are significantly similar in structure and in function.

The Chimpanzee OR (CONTIG3463.6-1888) maps to a human OR OR3A3 according to the classification (shown in table -I). Although OR3A3 and OR1D2 belong to different families but with respect to evolution in connection with Chimpanzee OR CONTIG3463.6-1888, they are structurally almost same as per quantification shown above.

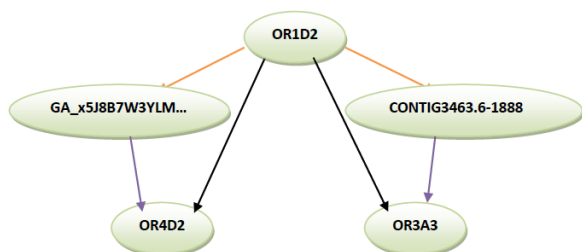


Fig-I: Evolutionary connect ion among Human, Mouse and Chimpanzee ORs

OR1D2, GA_x5J8B7W3YLM-7052533-7051808 and CONTIG3463.6-1888 are also most similar to OR4D2 and OR3A3 as shown above. They are evolutionarily connected and hence through biological evolution CONTIG3463.6-1888 and GA_x5J8B7W3YLM-7052533-7051808 are updated as OR3A3 and OR4D2 respectively.

B. Fractal and Morphological Quantification of ORs

The fractal dimensions of DNA nucleotide sequences---OR1D2 CONTIG3463.6-1888, GA_x5J8B7W3YLM-7052533-7051808 of Human, Chimpanzee, and Mouse respectively----generated by plotting the sequences in two axes (fig.-II) and that define a mapping as follows $f: \{X, Y\} \rightarrow \{0, 1\}$ as

$$f(X, Y) = 0 \text{ if } Y \neq X$$

$$= 1 \text{ otherwise}$$

yields respectively 1.77687, 1.81916 and 1.82963.

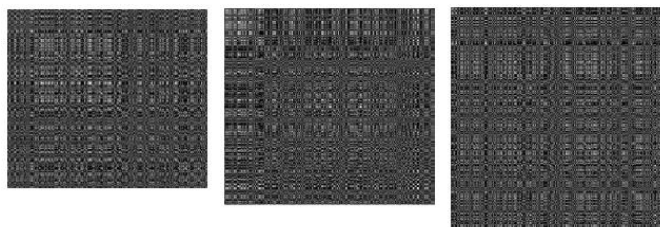


Fig-II: Binary Image matrix

Here we observe that fractal dimensions of ORs of Chimpanzee and Mouse are significantly similar. Through genomic evolution they got updated into OR1D2 in human the fractal dimension of is reduced by a small amount 0.04 i.e. through genomic evolution amount of complexity or disorderliness got decremented.

C. Results on Succolarity

For three olfactory receptor DNA sequences of Human, Mouse and Chimpanzee succolarity index are calculated. A DNA sequence can be thought of as a texture of four disjoint templates of A, T, C and G. For their four different templates of each DNA sequence the succolarity for each of those three sequences are shown in fig.-III.

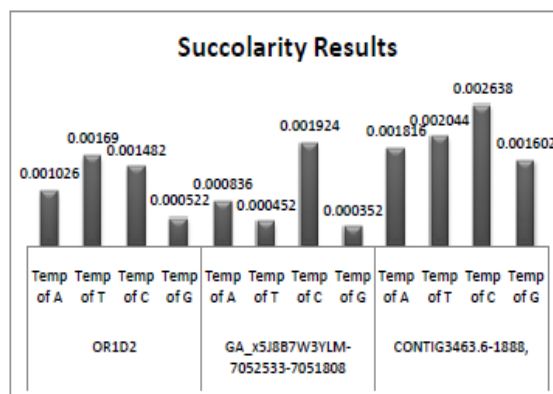


Figure-III: Succolarity of ORs

The succolarity of all the textures of A, T, C, and G are almost same for Mouse and Chimpanzee ORs but in case of Human OR OR1D2 are less than that of other two ORs. It is seen that over genomic evolution the succolarity (amount of continuous density) in sequence structure in Human OR is smaller than the other similar sequences of Mouse and Chimpanzee.

D. Results on Bifurcation Dimension of Skeleton

Let $f(x, y)$ denote a DNA sequence which is with two bits per pixel. This allows $f(x, y)$ to have maximum of 4 colors, i.e. $0 \leq f(x, y) \leq 3$. Here A, T, G, C are denoted respectively by red, blue, green yellow as shown below for CONTIG3463.6-1888.

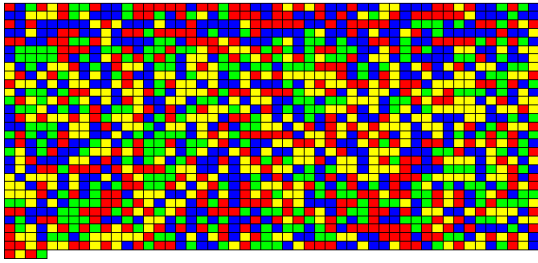


Figure-IV: Colored template of CONTIG3463.6-1888

We then decomposed the four colored image $f(x, y)$ into four binary images following the threshold decomposition function as defined below:

$$f^i(x, y) = 1 \quad z = i : i = 0, 1, 2 \text{ and } 3.$$

$$= 0 \text{ otherwise}$$

Corresponding to each of the binary image namely as shown in figure-V, we have obtained their respective skeletons of which the one for CONTIG3463.6-1888 which are shown below in figure-VI.

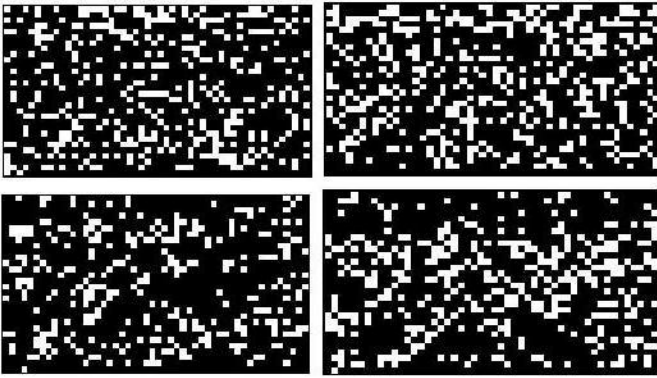


Fig.-V: Binary images of OR of Chimpanzee (Black & white denote background and foreground)

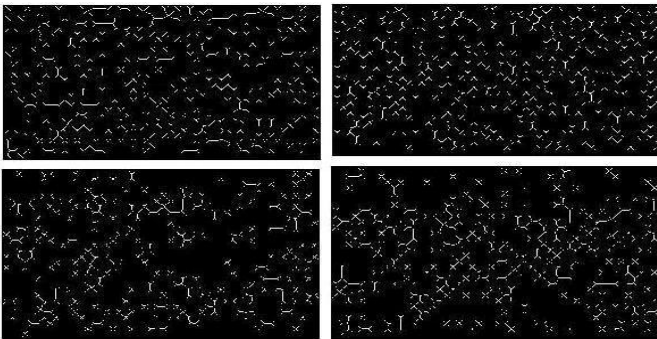


Fig.-VI: Skeletons of OR Chimpanzee

Intricacy of the skeleton for those decomposed binary images depends upon the spatial distribution. The higher intricacy of the skeleton proportional to higher the heterogeneity in the spatial distribution of the skeleton and vice versa. Hence, using morphometry based the technique [3, 4] *Bifurcation*

Dimension for these skeletons are computed and found similarity between the species as shown in fig.-VII.

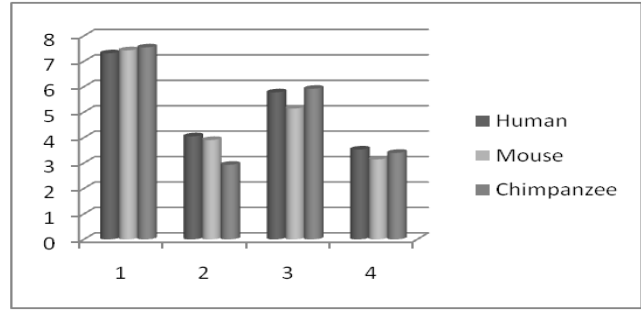


Fig.-VII: Histogram of Bifurcation Dimensions.

In fig.-VII, it is apparent that they do not follow a strict order. We believe this parameter provide a distinction between the functions of ORs.

IV. CONCLUSION AND FUTURE ENDEAVORS

In this paper we have shown an evolutionary connection among Human, Mouse and Chimpanzee ORs. These sequences have very close sequential similarity but they do differ in different species due to their intricate details of the structures in the DNA sequence. Those intricate details are illustrated here. In our further studies, we provide a quantitative classification based on Fractals and Mathematical Morphology with some more details about all the ORs of Human, Chimpanzee and Mouse.

ACKNOWLEDGMENT

The authors are grateful to Prof. Jean Serra, Emeritus Professor, ESIEE-Engineering, University Paris-Est, Former Director of Centre for Mathematical Morphology, France for his kind advice and suggestions and also express their gratitude to Mr. Pratap Vardhan for his technical help in writing programs.

REFERENCES

- [1] B. Malnic, PA Godfrey; The olfactory receptor gene family; PNAS, February 24, 2004 vol. 101 no. 8 2584-2589 2004.
- [2] P Kitts, EV Koonin, I Korf, D Kulp, D Lancet, Initial sequencing and analysis of the human genome, Nature, 409, 860-921, 2001.
- [3] G Matheron and J Serra, *History of Mathematical Morphology*, http://cmm.ensmp.fr/~serra/pdf/birth_of_mm.pdf
- [4] B. S. Daya Sagar, 1996, *Fractal relations of a morphological skeleton*, Chaos, Solitons & Fractals, v. 7, no. 11, p. 1871-1879.
- [5] B. S. Daya Sagar, et al 1998, *Morphometric relations of fractal-skeletal based channel network model*, Discrete Dynamics in Nature and Society, v. 2, no. 2, p. 77-92.
- [6] P. Radhakrishnan, B. S. Daya Sagar, and Teo Lay Lian, (2004) *Estimation of fractal dimension through morphological decomposition*, Chaos, Solitons & Fractals, v. 21, no. 3, p. 563-572.
- [7] B. S. Daya Sagar and Tay Lea Tien, (2004) *Allometric power-law relationships in a Hortonian fractal digital elevation model*, Geophysical Research Letters, (American Geophysical Union), v. 31, no. 6, L06501.
- [8] Shih and Mitchell Threshold Decomposition of Gray -Scale Morphology into Binary Morphology, IEEE transactions on pattern analysis and machine intelligence Vol. II, No. 1.
- [9] P Radhakrishnan, B. S. Daya Sagar - *Morphological decomposition of sandstone pore-space: fractal power-laws* Chaos, Solitons & Fractals.
- [10] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, Inc. Orlando, FL, USA ©1983 ISBN:0126372403.

- [11] Carlo Cattani, *Fractals and Hidden Symmetries in DNA*, Mathematical Problems in Engineering Volume 2010, Article ID 507056.
- [12] Choudhury, P.P.; Carry Value Transformation (CVT): Its Application in Fractal formation, Advance Computing Conference, 2009. IACC 2009. IEEE.
- [13] Choudhury, P.P. et al *Act of CVT and EVT In the Formation of Number-Theoretic Fractals*, IJCC, Vol 9, No. March 2011, USA.
- [14] B. B. Mandelbrot, *The fractal geometry of nature*. New York, 1982.
- [15] R. H. C. de Melo and A. Conci, Succolarity: Defining a Method to calculate this Fractal Measure, ieeexplore.ieee.org/iel5/4595514/4604343/04604424.pdf.
- [16] Yu Zu-Guo et al *Fractals in DNA sequence analysis*, Chinese Physics, Vol 11, No. 12, 2002, 1313-1318.