Fractal Geometry and Nonlinear Analysis in Medicine and Biology

Research Article ISSN: 2058-9506

A quantitative understanding of human sex chromosomal genes

Sk. Sarif Hassan[1], Pabitra Pal Choudhury[2], Antara Sengupta[2], Binayak Sahu[3], Rojalin Mishra[3], Devendra Kumar Yadav[3], Saswatee Panda[3], Dharamveer Pradhan[3], Shrusti Dash[3] and Gourav Pradhan[3]

[1]Department of Mathematics, University of Petroleum and Energy Studies, Dehradun, India

[2]MCKV Institute of Engineering, Liluah, Howrah, West Bengal, India

[3]Visiting Student at Institute of Mathematics & Applications, Bhubaneswar 751003, Odisha, India

Abstract

In the last few decades, the human allosomes are engrossed in an intensive attention among researchers. The allosomes are now already been sequenced and found there are about 2000 and 78 genes in human X and Y chromosomes respectively. The hemizygosity of the human X chromosome in males exposes recessive disease alleles, and this phenomenon has prompted decades of intensive study of X-linked disorders. By contrast, the small size of the human Y chromosome, and its prominent long-arm heterochromatic region suggested absence of function beyond sex determination. But the present problem is to accomplish whether a given sequence of nucleotides i.e. a DNA is a Human X or Y chromosomal genes or not, without any biological experimental support. In our perspective, a proper quantitative understanding of these genes is required to justify or nullify whether a given sequence is a Human X or Y chromosomal gene. In this paper, some of the X and Y chromosomal genes have been quantified in genomic and proteomic level through Fractal Geometric and Mathematical Morphometric analysis. Using the proposed quantitative model, one can easily make probable justification or deterministic nullification whether a given sequence of nucleotides is a probable Human X or Y chromosomal gene or not, without seeking any biological experiment. Of course, a further biological experiment is essential to validate it as the probable Human X or Y chromosomal gene homologue. This study would enable Biologists to understand these genes in more quantitative manner instead of their qualitative features.

Introduction

In the last ten years, Genomics has revolutionized the study of evolution. Evolution changes the sequence of DNA molecules, and comparing DNA sequences allow us to reconstruct evolutionary events from the past. The availability of DNA sequences from multiple vertebrates has confirmed that the process of sex chromosome evolution as envisioned by theorists has played out multiple times in the evolution of vertebrate sex chromosomes. However, complete, high-quality sequences of sex chromosomes have led to discoveries that were unanticipated by existing theory. The next stage of genomic research will begin to derive meaningful knowledge from these genes. A quantitative genomic understanding will have a major impact in the fields of medicine, biotechnology, and the life sciences [1,2].

One of the most frontier contemporary challenges is to make a revolution in medical science by introducing *Genetic Therapy* [3]. *Gene therapy* is an experimental technique that uses genes to treat or prevent disease. This method of therapy would countenance us to treat a disorder by inserting a gene into a patient's cells instead of using drugs or surgery. The most commonly practiced approaches of gene therapy include…

- Replacing a mutated gene that causes disease with a healthy copy of the gene.
- Inactivating, or "knocking out," a mutated gene that is functioning improperly.
- Introducing a new gene into the body to help fight a disease.

Although gene therapy is a promising treatment option for a number of diseases (including inherited disorders, some types of cancer, and certain viral infections), the technique remains risky and is still under study to make sure that it will be safe and effective. Gene therapy is currently being tested for the treatment of diseases that have no other cures [3-6]. Prior to gene therapy as a practical approach for treating diseases, we must overcome many technical challenges. One of the nontrivial challenges is to get quantitative insight of genes. This would help us in precise characterization of a particular DNA. This quantitative study of genes will be an add-on as the genetic signature of a gene.

In the present study, a mathematical quantification of human X and Y chromosomal genes [7-13] has been done by using *Fractal Geometry* [14-16]. So on using this proposed quantitative model, one can easily make probable justification (or deterministic nullification) whether a given sequence of nucleotides is a probable Human X/Y chromosomal genes or its homologue or not, without seeking any biological experiment. This study would help researchers in understanding these genes in differentiating from each other through the very nucleotide syntactical presentation.

Model decomposition and representation

(A) DNA 4-colored representation: Let a DNA sequence is in the form of four-letter (ATGC) nucleotides sequence (Figure 1A). Such

Correspondence to: Sk Sarif Hassan, Department of Mathematics, University of Petroleum and Energy Studies, Dehradun, India, E-mail: s.hassan@ddn.upes.ac.in

Key words: *human X and Y chromosomes, genes, fractal dimension, hurst exponent, protein analysis*

Received: October 25, 2016; **Accepted:** December 16, 2016; **Published:** December 19, 2016

sequence shown in Figure 1A is converted as a function (Figure 1B) depicting colors Red, Blue, Green, and Yellow respectively for A, T, G, and C [17,18]. This allows $f(x,y)$ having maximum of 4 colors, i.e. $0 \le f(x,y) \le 3$.

(B) Binary representation: A DNA as a one dimensional nucleotide sequence, and is represented as a map such that $T(A) = 00;\ T(T) = 11,\ T(C) = 01$ *and* $T(G) = 10$. This mapping yields a DNA sequence in a binary string format. A portion of such a binary string is shown below of some fixed size (twice of the DNA sequence length). We call this representation of DNA as *2-adic string* of DNA [18].

001110000100101000111110000000000001100001000001111000100010011110011110101111111000001001111100010111111010101001001111111110000010110010011110000000011000011110011001101100100110000000000000111111110111100110011111111110100011111110111111001111111101000000000011110011000000000111110101011101100000100001001111011111010011011001…..(some more 0, 1 are there in the string).

(C) 4-Adic representation: Also we consider a DNA as a string of four variables 0, 1, 2 and 3 (as shown below) corresponding to A T C and G respectively. We name this string as *4-adic string* of DNA [17,18].

023010330223000002003002201010220221122200102230322211103122230032031230002002202021310200000122232202022222101222122022221000002202000022333232003010221220022120030000232230222231220212210232222202210022003301222233200010222312332322002322000030303223333000233023310233331212333003012112030200010122200303312…

(D) Threshold decomposition: We have decomposed the four-colored image $f(x,y)$ into four binary images $f^{i}(x,y)=z$ (Figures 2A-D) for a DNA through the threshold decomposition function defined as:

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

Those decomposed binary images for one human X and Y are denoted as are $f_{SRY}{}^{A}, f_{SRY}{}^{T}, f_{SRY}{}^{G} and\ f_{SRY}{}^{C}$ shown in the following:

(E) Skeleton decomposition: Morphological skeletons Figures 5 & 6 a-d for threshold decomposed binary images of SRY shown in (Figures 5 & 6 a-d) is obtained according to (3).

S(X)=[(X$ nB)(X $nB)0B] …… (3) where *B* is a structuring element that is symmetric about the origin, and $nB = B \oplus B \oplus B \ldots \oplus B(n\ times)$.

Density and intricacy of the skeleton for those decomposed binary images depend upon the frequency of occurrence of nucleotide chosen as threshold and their spatial distribution. The intricacy of the skeleton is proportional to the heterogeneity in the spatial distribution of the skeleton.

(F) Protein plot representations: A number of fundamental properties namely percentage of Accessible Residues (AR), Buried Residues (BR), Alpha Helix (Chou & Fasman) (ALCF), Amino Acid Composition (ACC), Beta Sheet (Chou & Fasman) (BSCF), Beta Turn (Chou & Fasman) (BTCF), Coil (Deleage & Roux) (CDR), Hydrophobicity (Aboderin) (HA), Molecular Weight (MW), and Polarity (P) of the human X and Y chromosomal genes have been considered (Figure 3).

All protein plots are generated from the gene sequences using Matlab (*bioinformatics toolbox*) (Figure 4). Then box-counting dimension for each of the protein plot have been calculated through BENOIT™.

In the next section let us elaborate the methods applied to DNA string to extract the quantitative details.

Methods

The quantitative details of X and Y chromosomal genes have been studied in the light of fractal dimension. The very basic of one such fractal dimension method is *Box Counting Dimension* which is illustrated below.

Box-counting method: The most practical and commonly used method of calculation of fractal dimension is Box-counting dimension. This is mainly because it is easy to calculate mathematically and because it is easily estimated empirically. We note that the number of line segments of length δ that are needed to cover a line of length l is $\frac{1}{\delta}$, that the number of squares with side length δ that are needed to cover a square with area A is $\frac{A}{\delta^{2}}$, and that the number of cubes with side

GTTGAGGGGGTGTTGAGGGCGGAGAAATGCAAGTTTCATTACAAAAGTTAACGTAACAAAGAATCTGGTAGAAGTGAGTTTTGGATAGTAAAATAAGTTTCGAACTC
TGGCACCTTTCAATTTTGTCGCACTCTCCTTGTTTTTGACAATGCAATCATATGCTTCTGCTATGTTAAGCGTATTCAACAGCGATGATTACAGTCCAGCTGTGCAA
GAGAATATTCCCGCTCTCCGGAGAAGCTCTTCCTTCCTTTGCACTGAAAGCTGTAACTCTAAGTATCAGTGTGAAACGGGAGAAAACAGTAAAGGCAACGTCCAGGA
TAGAGTGAAGCGACCCATGAACGCATTCATCGTGTGGTCTCGCGATCAGAGGCGCAAGATGGCTCTAGAGAATCCCAGAATGCGAAACTCAGAGATCAGCAAGCAGC
TGGGATACCAGTGGAAAATGCTTACTGAAGCCGAAAAATGGCCATTCTTCCAGGAGGCACAGAAATTACAGGCCATGCACAGAGAGAAATACCCGAATTATAAGTAT
CGACCTCGTCGGAAGGCGAAGATGCTGCCGAAGAATTGCAGTTTGCTTCCCGCAGATCCCGCTTCGGTACTCTGCAGCGAAGTGCAACTGGACAACAGGTTGTACAG
GGATGACTGTACGAAAGCCACACACTCAAGAATGGAGCACCAGCTAGGCCACTTACCGCCCATCAACGCAGCCAGCTCACCGCAGCAACGGGACCGCTACAGCCACT
GGACAAAGCTGTAGGACAATCGGGTAACATTGGCTACAAAGACCTACCTAGATGCTCCTTTTTACGATAACTTACAGCCCTCACTTTCTTATGTTTAGTTTCAATAT
TGTTTTCTTTTCTCTGGCTAATAAAGGCCTTATTCATTTCA

Figure 1A. A DNA string of four variables A T C and G of SRY.

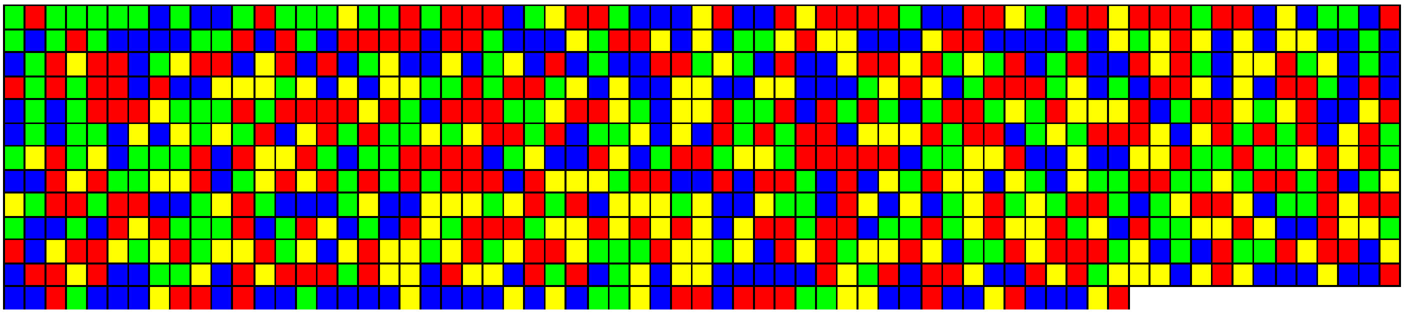


Figure 1B. Function generated by proper colour coding for ATGC of SRY.

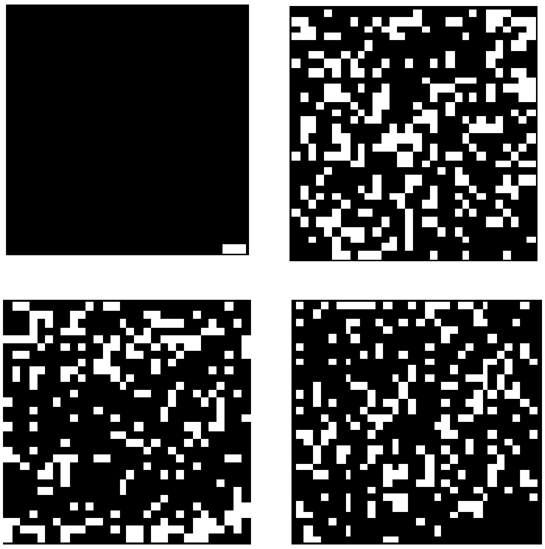


Figure 2. Threshold decomposed binary images of OR10AB1P (Black and white denotes complimentary space and one of the ATGC). (A) A (B) T (C) G, and (D) C.

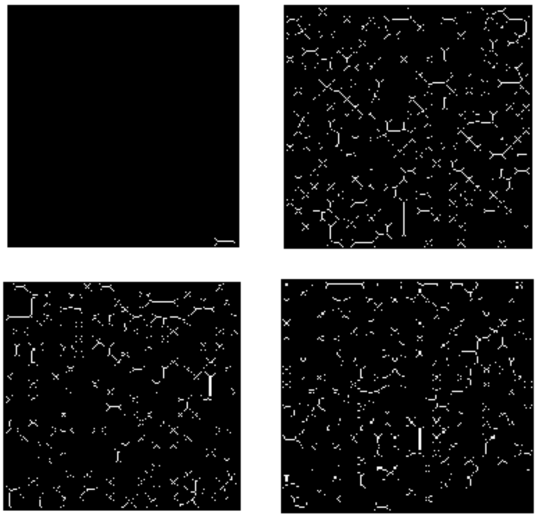


Figure 3. Skeletons of $f_{SRY}{}^{A}$, $f_{SRY}{}^{T}$, $f_{SRY}{}^{G}$ and $f_{SRY}{}^{C}$.

length δ that are needed to cover a cube with volume V is $\frac{v}{\delta^3}$.

So in general, the box-counting dimension (or just ``box dimension") of a set S subset of $\mathbb{R}^n$ as follows:

For any $\varepsilon > 0$, let $N_\varepsilon(S)$ be the minimum number of n-dimensional cubes of side-length ε needed to cover S. If there is a number D so that

$$N_\varepsilon(S) = 1/\varepsilon^D$$

The D is called the **Box-Counting Dimension** of S.

It is to be Noted that the Box-Counting Dimension is D if and only if there is some positive constant m so that

$$\lim_{\varepsilon\to 0}\frac{N_\varepsilon(S)}{1/\varepsilon^D} = m$$

The above equation gives $D = \lim_{\varepsilon\to 0}\frac{\log m - \log N_\varepsilon(S)}{\log\varepsilon} = -\lim_{\varepsilon\to 0}\frac{\log N_\varepsilon(S)}{\log\varepsilon}$.

Note that the log m term drops out, because it is constant while the denominator becomes infinite as $\varepsilon \to 0$. Also, since $0 < \varepsilon < 1$, $\log\varepsilon$ is negative, so D is positive.

But in practice, this method computes the number of cells required to entirely cover an object, with grids of cells of varying size. Practically, this is performed by superimposing regular grids over an object and by counting the number of occupied cells. The logarithm of $N(r)$, the number of occupied cells, versus the logarithm of $1/r$, where r is the size of one cell, gives a line whose gradient corresponds to the box dimension [14].

FD of DNA walks of the genes

The DNA walk is defined as a sum of the progression $\sum D_n, n = 1,2,.....,N \;\&\; D_n \in \{1,2,3,4\}$ which is the cumulative sum on the DNA string representation $\left\{D_1, D_1 + D_2,......., \sum_{m=1}^{n-1} D_m,........, \sum_{m=1}^{N} D_m\right\}$ [19].

Also we define $a_n \overset{\text{def}}{=} \sum_{i=1}^{n} f(A,x_i)$, $g_n \overset{\text{def}}{=} \sum_{i=1}^{n} f(G,x_i)$, $c_n \overset{\text{def}}{=} \sum_{i=1}^{n} f(C,x_i)$ & $t_n \overset{\text{def}}{=} \sum_{i=1}^{n} f(U,x_i)$. It has been resulted by plotting (P_n, Q_n) as we have defined two functions:

$$P_n \overset{\text{def}}{=} \sin a_n^2 - \sin g_n^2 \text{ and } Q_n \overset{\text{def}}{=} \sin t_n^2 - \sin c_n^2$$

Here we compute the Fractal dimension of all DNA walk for the *4-adic string* of all sex genes. The plot of the DNA walk for the SRY

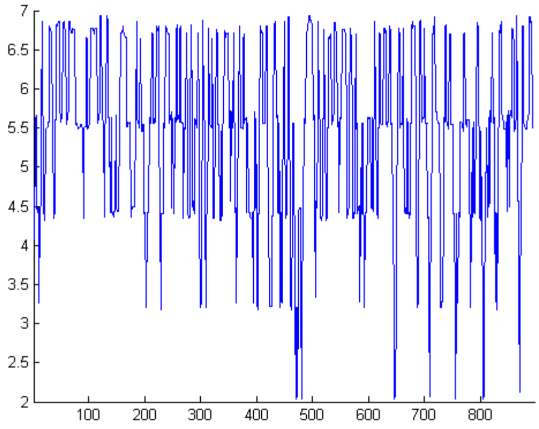


Figure 4. Accessible residues (AR-Protein Plot) of SRY.

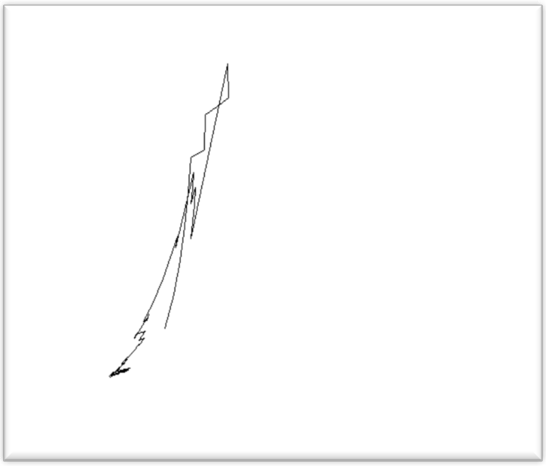
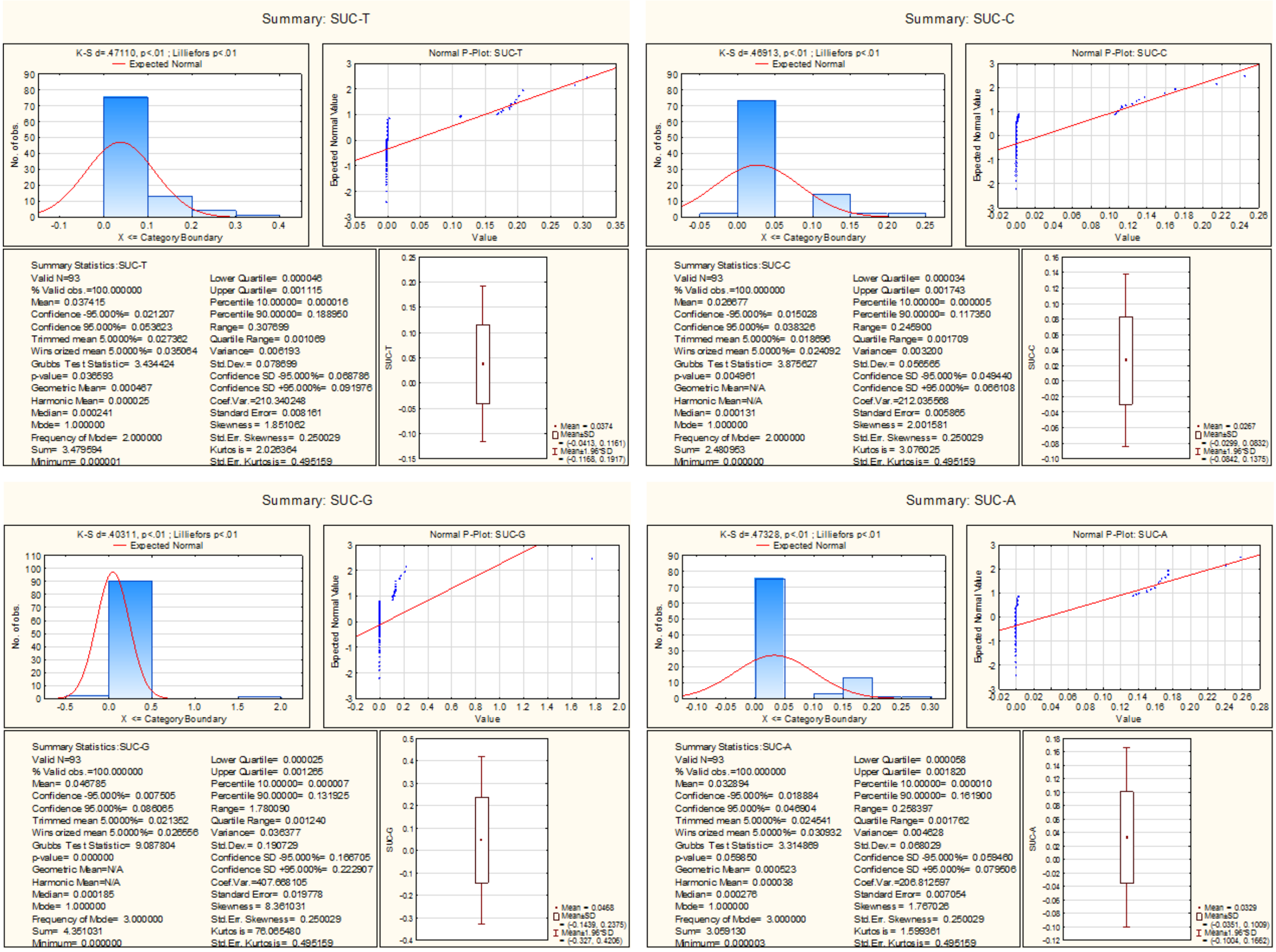


Figure 5. DNA Walk (P_n, Q_n) for SRY.



```
Summary Statistics:SUC-T
Valid N=93                              Lower Quartile= 0.000046
% Valid obs.=100.000000                 Upper Quartile= 0.001115
Mean= 0.037415                          Percentile 10.00000= 0.000016
Confidence -95.000%= 0.021207           Percentile 90.00000= 0.188950
Confidence 95.000%= 0.053623            Range= 0.307699
Trimmed mean 5.0000%= 0.027362          Quartile Range= 0.001069
Winsorized mean 5.0000%= 0.035064       Variance= 0.006193
Grubbs Test Statistic= 3.434424         Std.Dev.= 0.078699
p-value= 0.036593                       Confidence SD -95.000%= 0.068786
Geometric Mean= 0.000467                Confidence SD +95.000%= 0.091976
Harmonic Mean= 0.000025                 Coef.Var.=210.340248
Median= 0.000241                        Standard Error= 0.008161
Mode= 1.000000                          Skewness= 1.851062
Frequency of Mode= 2.000000             Std.Err. Skewness= 0.250029
Sum= 3.479594                           Kurtosis= 2.026364
Minimum= 0.000001                       Std.Err. Kurtosis= 0.495159
```

```
Summary Statistics:SUC-C
Valid N=93                              Lower Quartile= 0.000034
% Valid obs.=100.000000                 Upper Quartile= 0.001743
Mean= 0.026677                          Percentile 10.00000= 0.000005
Confidence -95.000%= 0.015028           Percentile 90.00000= 0.117350
Confidence 95.000%= 0.038326            Range= 0.245900
Trimmed mean 5.0000%= 0.018696          Quartile Range= 0.001709
Winsorized mean 5.0000%= 0.024092       Variance= 0.003200
Grubbs Test Statistic= 3.875627         Std.Dev.= 0.056565
p-value= 0.004961                       Confidence SD -95.000%= 0.049440
Geometric Mean=N/A                      Confidence SD +95.000%= 0.066108
Harmonic Mean=N/A                       Coef.Var.=212.035568
Median= 0.000131                        Standard Error= 0.005865
Mode= 1.000000                          Skewness= 2.001581
Frequency of Mode= 2.000000             Std.Err. Skewness= 0.250029
Sum= 2.480953                           Kurtosis= 3.076025
Minimum= 0.000000                       Std.Err. Kurtosis= 0.495159
```

```
Summary Statistics:SUC-G
Valid N=93                              Lower Quartile= 0.000025
% Valid obs.=100.000000                 Upper Quartile= 0.001265
Mean= 0.046785                          Percentile 10.00000= 0.000007
Confidence -95.000%= 0.007505           Percentile 90.00000= 0.131925
Confidence 95.000%= 0.086065            Range= 1.780090
Trimmed mean 5.0000%= 0.021352          Quartile Range= 0.001240
Winsorized mean 5.0000%= 0.026556       Variance= 0.036377
Grubbs Test Statistic= 9.087804         Std.Dev.= 0.190729
p-value= 0.000000                       Confidence SD -95.000%= 0.166705
Geometric Mean=N/A                      Confidence SD +95.000%= 0.222907
Harmonic Mean=N/A                       Coef.Var.=407.668105
Median= 0.000185                        Standard Error= 0.019778
Mode= 1.000000                          Skewness= 8.361031
Frequency of Mode= 3.000000             Std.Err. Skewness= 0.250029
Sum= 4.351031                           Kurtosis= 76.065480
Minimum= 0.000000                       Std.Err. Kurtosis= 0.495159
```

```
Summary Statistics:SUC-A
Valid N=93                              Lower Quartile= 0.000058
% Valid obs.=100.000000                 Upper Quartile= 0.001820
Mean= 0.032894                          Percentile 10.00000= 0.000010
Confidence -95.000%= 0.018884           Percentile 90.00000= 0.161900
Confidence 95.000%= 0.046904            Range= 0.258397
Trimmed mean 5.0000%= 0.024541          Quartile Range= 0.001762
Winsorized mean 5.0000%= 0.030932       Variance= 0.004628
Grubbs Test Statistic= 3.314869         Std.Dev.= 0.068029
p-value= 0.059850                       Confidence SD -95.000%= 0.059460
Geometric Mean= 0.000523                Confidence SD +95.000%= 0.079506
Harmonic Mean= 0.000038                 Coef.Var.=206.812597
Median= 0.000276                        Standard Error= 0.007054
Mode= 1.000000                          Skewness= 1.767026
Frequency of Mode= 3.000000             Std.Err. Skewness= 0.250029
Sum= 3.059130                           Kurtosis= 1.599361
Minimum= 0.000003                       Std.Err. Kurtosis= 0.495159
```

Figure 6. Descriptive Statistics of Succolarities of genomic texture A, T, C and G.

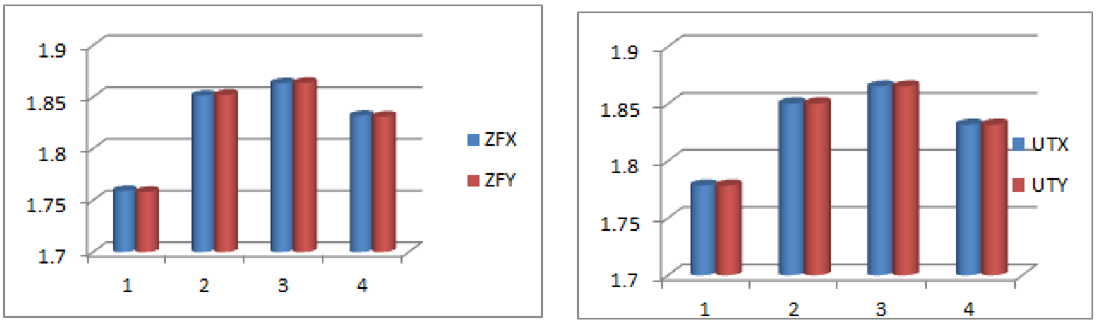


Figure 7. Histograms of Fractal dimension of threshold decompositions.

string is shown in Figure 5.

The box-counting dimension for the DNA walk of SRY is 1.94611. In the similar manner we have computed all the box counting dimension of all genes.

Hurst exponent of the DNA sequences

Hurst exponent is referred as the "index of dependence," and is the relative tendency of a time series either to regress strongly to the mean or to cluster in a direction. It is a measure of long range correlation of one-dimensional time series [19,20].

Let us consider a string $H=\{h_i\}, i=1,2,......,n$

$$m_{X,n}=\frac{1}{n}\sum_{i=1}^{n}h_i$$

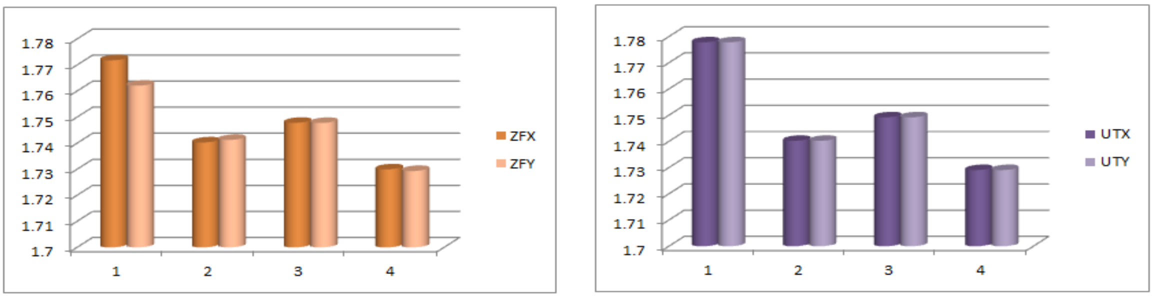


Figure 8. Histograms of Fractal dimension of threshold decompositions.

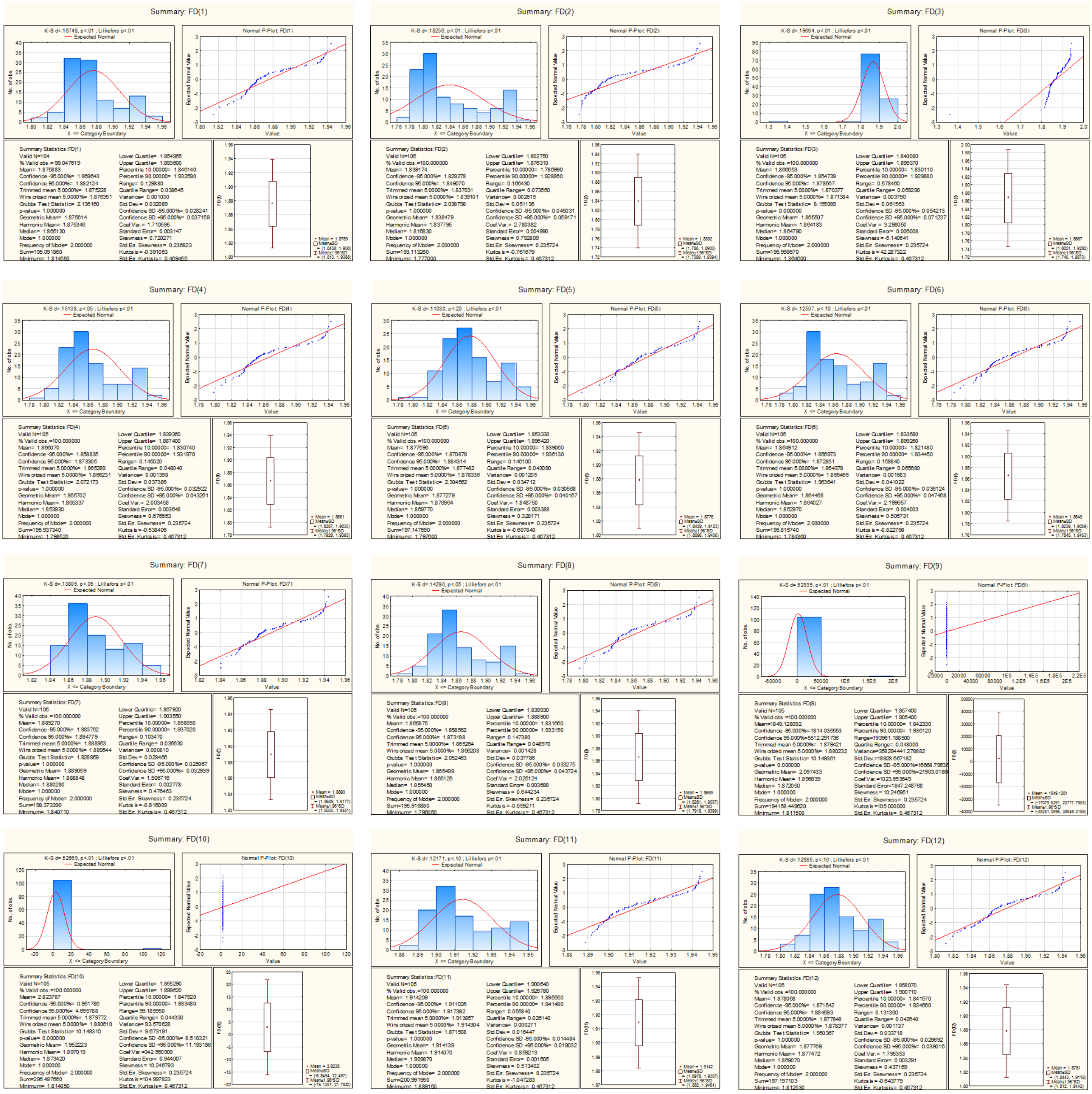


Figure 9. Descriptive Statistics of fractal dimension of Protein Plots.

$$H(i,X)=\sum_{j=1}^{i}\{h_j-m_{x,n}\}$$

$$R(n)=\max H(i,n)-\min H(i,n)1\le i\le n$$

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

The Hurst exponent H is defined as: $\left(\frac{n}{2}\right)^H=\frac{R(n)}{S(n)}$, where n is the length of the string. The range for which the Hurst exponent, H indicates negative, positive auto-correlation are 0<H<0.5 and 0.5<H<1 respectively. A value of H=0.5 indicates a true random walk, where it is equally likely that a decrease or an increase will follow from any particular value [20].

Here we consider 2-adic strings of DNA for computation of Hurst exponent.

The Hurst exponents of the *2-adic string* of SRY are 0.69106. This is how we have computed Hurst exponent for all the human genes [18].

Succolarity

The degree of percolation of an image (how much a given fluid can flow through this image) can be measured through Succolarity, a fractal parameter [21,22].

The succolarity of a binary image is defined as

$$\sigma(BS(k),dir)=\frac{\sum_{k=1}^{n}OP(BS(k))\times PR(BS(k),pc)}{\sum_{k=1}^{n}PR(BS(k),pc)}$$

where '*dir*' denotes direction; $BS(n)$ where n is the number of possible divisions of a binary image in boxes. The occupation percentage (OP) is defined as, for each box size, k, the sum of the multiplications of the $OP(BS(k))$, where k is a number from 1 to n, by the pressure $PR(BS(k),pc)$, where *pc* is the position on x or y of the centroid of the box on the scale of pressure) applied to the box are calculated. Therefore for any binary decomposed images of $f(x,y)$, the succolarity can be obtained.

Here we compute succolarity of the decomposed images for DNA as shown in the previous section. The succolarity of the four decomposed images $f_{SRY}{}^{A}$, $f_{SRY}{}^{T}$, $f_{SRY}{}^{G}$ *and* $f_{SRY}{}^{C}$ of SRY are 0.000351, 0.000782, 0.000272 and 0.000267 respectively.

Similarly, we have computed the succolarity of the decomposed images for all sex genes.

Statistical autocorrelations

It is one of several descriptors, describing how far the values lie from the mean (expected value).

For a given sequence $\{Y_1,Y_2.....Y_N\}$,

$\sigma^2 \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N}Y_i^2-(\frac{1}{N}\sum_{i=1}^{N}Y_i)^2$ and the variance at distance N-k is given as

$$\sigma^2 \overset{\text{def}}{=} \frac{1}{N-k}\sum_{i=1}^{N-k}Y_i^2-(\frac{1}{N-k}\sum_{i=1}^{N-k}Y_i)^2\ [11].$$

It is easily computable that the variance for the string Figure 1(C) is 1.29.

Mean and SD ordering of gene sequences

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. We can classify the miRNA sequences based on the ordering of poly-string mean of A, C, T, and G in the string. Given a string *X*, we calculate the mean of poly-strings consisting only of A, C, T and G separately [15,16].

Mean $Nu=2(Nu_1+Nu_2+Nu_3+Nu_4.....+Nu_m)/m.(m+1)$ where $Nu_i\in\{A,U,C,G\},i=1,2,....,m$ and m is the length of the longest poly-string over the string.

According to the non-decreasing order of mean, we have classified all the genes into different classes. The mean order of sequence *1(A)* is AUGC i.e. mean of poly-string of A is less than the same of U and so on.

Results and discussions

Let us now elaborate in detail, the result obtained for all sex genes using the above stated methods.

Fractal dimension of DNA walk

For all the 92 human sex chromosomal genes along with their homologues in other species, the fractal dimension of the DNA walks lies in the interval (1.896, 1.948) as shown in Tables 1 and Figure 11.

From Tables 1 and Figure 11 it is clear that the box counting dimension is cantered at 1.94. The human sex genes and their corresponding homologues have almost same box counting dimensions. Although the ordering and length of the gene sequences are different from each other but they share the same box counting dimension of DNA walk.

Hurst exponent of 2-adic DNA strings

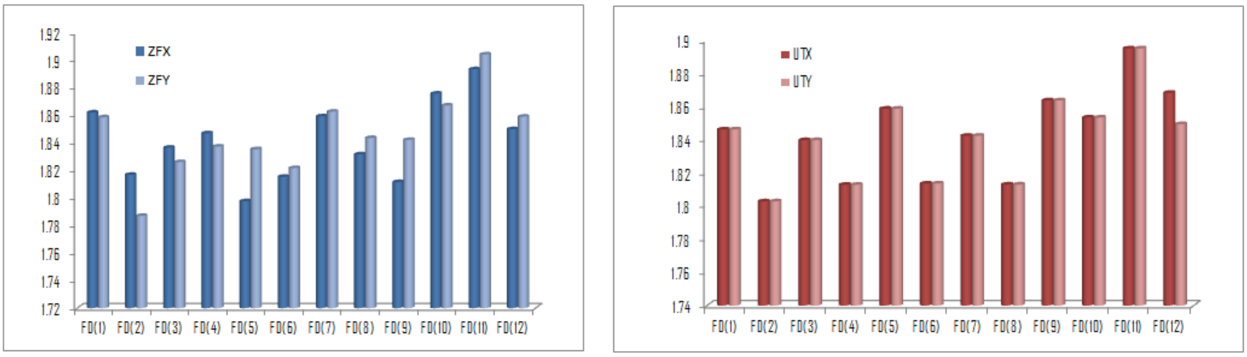


Figure 10. Fractal dimension of Protein Plots of ZFX, ZFY and UTX, UTY.

Table 1. Genes and Their FD of DNA Walk.

Genes	FD of DNA Walk
DHRSX (*Bostaurus*)	1.94577
DHRSX (*Homo sapiens*)	1.94584
CD99 (*Bostaurus*)	1.94527
CD99 (*Homo sapiens*)	1.92993
CD99 (*Pan troglodytes*)	1.94588
CD9912 *(Musmusculus)*	1.94584
ZBED1 *(Bostaurus)*	1.94577
ZBED1 *(Homo sapiens)*	1.92829
PRKX (*Homo sapiens*)	1.94595
PRKX (*Pan troglodytes*)	1.94573
PRKY (*Homo sapiens*)	1.94595
NLGN4Y (*Homo sapiens*)	1.90746
NLGN4Y (*Pan troglodytes*)	1.94615
VCX2 (*Homo sapiens*)	1.89608
VCX2 (*Pan troglodytes*)	1.94592
TBL1Y (*Homo sapiens*)	1.94556
TBL1Y (*Pan troglodytes*)	1.94579
SC25A6 (*Homo sapiens*)	1.9298
SC25A6 (*Musmuscullus*)	1.9458
ASMTL (*Gallus gallusdomesticus*)	1.9457
ASMTL (*Homo sapiens*)	1.92995
ASMTL (*Canis lupus*)	1.9458
ASMT (*Homo sapiens*)	1.9258
ASMT (*Gallus gallusdomesticus)*	1.9459
IL3RA (*Pan troglodytes*)	1.9458
IL3RA (*Musmuscullus*)	1.9458
HSFX1 (*Homo sapiens*)	1.92828
HSFX1 (*Pan troglodytes*)	1.9454
HSFY2 (*Homo sapiens*)	1.92982
IL9R (*Homo sapiens*)	1.92982
IL9R (*Pan troglodytes*)	1.94576
IL9R (*Maccamullata*)	1.94571
SYBL1 (*Homo sapiens*)	1.92956
SYBL1 (*Maccamullata*)	1.9457
SPRY3 (*Homo sapiens*)	1.92964
SPRY3 (*Maccamullata*)	1.93826

We have calculated the *Hurst exponent* of all 2-adic strings of DNA for all the human sex genes including their homologues. Hurst exponent lies in the interval (0.5559, 0.98187) and therefore the binary strings of all genes are positively auto-correlated (Figure 12).

The Hurst exponents for all human sex genes including their corresponding homologue have same as reflected in (Figure 12).

Succolarity indices

Succolarity measures how much a given fluid can flow through an image, considering as obstacles the set of pixels with a defined color (e.g. white) on 2D images analysis. In other words, it is a measure of continuous density of a 2D pattern.

The succolarity of A for all sex genes lies in the interval (0.000003, 0.2584) and so it is evident that the texture of A for each sex gene is having less density (Figure 1).

The succolarity indices of the genomic textures of T of all the human sex genes with their corresponding homologues are spread over the interval (0.000001, 0.3077) (1). It is seen that the succolarity i.e. the continuous density of the genomic texture of T is very low as it is seen in case of the genomic texture A. The succolarities for all these 93 genes are centred at 0.03 and there is no much deviation among the succolarity indices.

In case of the genomic texture G, The succolarity indices of all sex genes lie in the interval (0, 1.78) as shown in Figure 6. The succolarity indices of the genomic texture of C for all genes are computed. It is observed that the indices range from 0 to 0.2459.

It is seen that the human sex genes and their corresponding homologues share almost same succolarity. Also the succolarity indexes of human sex genes are higher than their corresponding homologues of other species. For an example, the succolarity of A of human sex gene SOY is greater than *Macaca mulatta*'s SOY of monkey and chimpanzee.

The correlation among succolarities of A, T, C and G are illustrated in the Table 2.

The correlation coefficient between the Suc-A and Suc-T is high and in contrast the correlation coefficient between the Suc-G and Suc-C is low.

Statistical autocorrelations

The statistical autocorrelations (σ) of the 4-adic representations of all sex genes of human along with their corresponding homologues are being determined. It has been found that the σ values lie in the interval (1.06, 1.945).

The sigma values of all sex genes of human and their homologues are normally distributed as shown in the Figure 13.

Fractal dimensions of threshold decomposition matrices

Here we considered the four different threshold decomposition matrices namely the template of A, T, C and G as we did in the 1.1 (D) for each of the sex-genes (Figure 6). Then we have determined the fractal dimension of the threshold decomposed matrices.

In the Figure 14, it is seen that the fractal dimension of the template of A, T, G and C lies in the interval (1.71, 1.81), (1.72, 1.89), (1.70, 1.90) and (1.74, 1.87) respectively.

The fractal dimensions of the decomposed template of A and T follow normal distribution whereas fractal dimensions of the other templates do not follow the same.

From the Figure 7, it is seen that the fractal dimensions of the threshold decomposition matrices for A, T, G and C of genes ZFX (from Human X- chromosome) and ZFY (from Human Y-chromosome) are almost same although the genomic template are entirely different in terms of ordering of nucleotides. The aforesaid fact holds good for all the one to one corresponding genes from X and Y chromosomes.

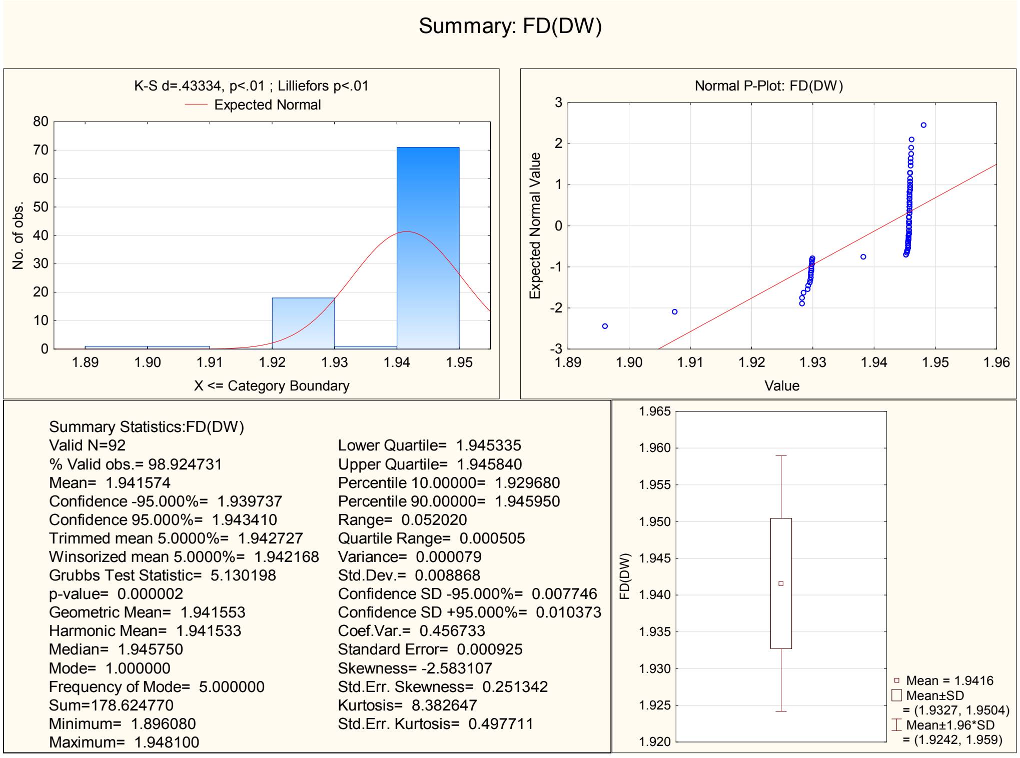
Fractal dimensions of skeleton of threshold decompositions

In the earlier subsection, the fractal dimensions of threshold decomposition matrices are found. Let us now find out the fractal dimension of the morphological skeleton of all decomposed threshold matrices.

In the Figure 15, it is determined that the fractal dimensions of

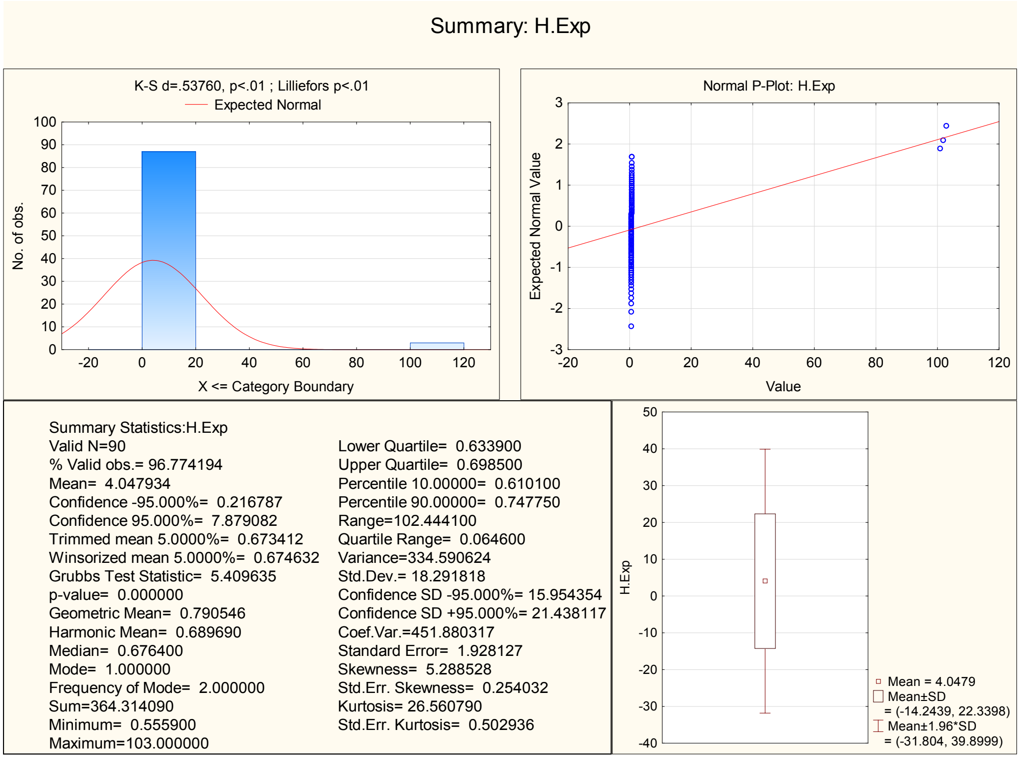
Table 2. Correlation coefficients for Succolarities of A, T, C and G.

Variables	SUC-A	SUC-T	SUC-C	SUC-G
SUC-A	1	0.994016	0.974306	0.244261
SUC-T	0.994016	1	0.957539	0.239863
SUC-C	0.974306	0.957539	1	0.251441
SUC-G	0.244261	0.239863	0.251441	1



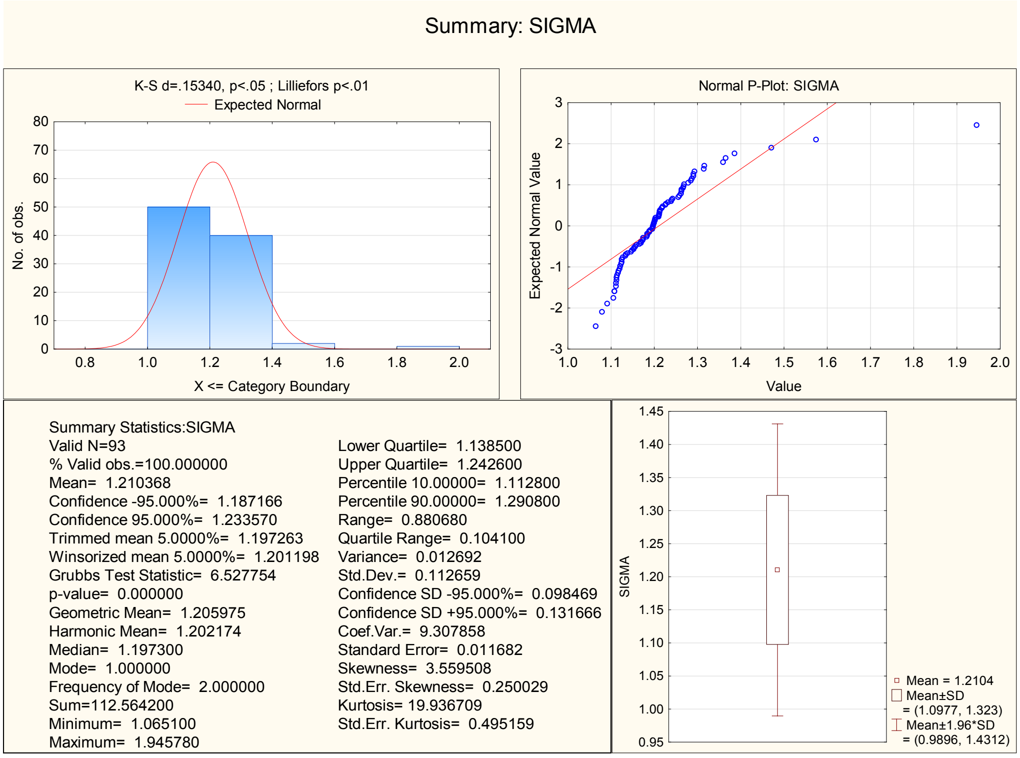
```
Summary Statistics:FD(DW)
Valid N=92                             Lower Quartile= 1.945335
% Valid obs.= 98.924731                Upper Quartile= 1.945840
Mean= 1.941574                         Percentile 10.00000= 1.929680
Confidence -95.000%= 1.939737          Percentile 90.00000= 1.945950
Confidence 95.000%= 1.943410           Range= 0.052020
Trimmed mean 5.0000%= 1.942727         Quartile Range= 0.000505
Winsorized mean 5.0000%= 1.942168      Variance= 0.000079
Grubbs Test Statistic= 5.130198        Std.Dev.= 0.008868
p-value= 0.000002                      Confidence SD -95.000%= 0.007746
Geometric Mean= 1.941553               Confidence SD +95.000%= 0.010373
Harmonic Mean= 1.941533                Coef.Var.= 0.456733
Median= 1.945750                       Standard Error= 0.000925
Mode= 1.000000                         Skewness= -2.583107
Frequency of Mode= 5.000000            Std.Err. Skewness= 0.251342
Sum=178.624770                         Kurtosis= 8.382647
Minimum= 1.896080                      Std.Err. Kurtosis= 0.497711
Maximum= 1.948100
```

Figure 11. Descriptive Statistics of FD of DNA Walk.



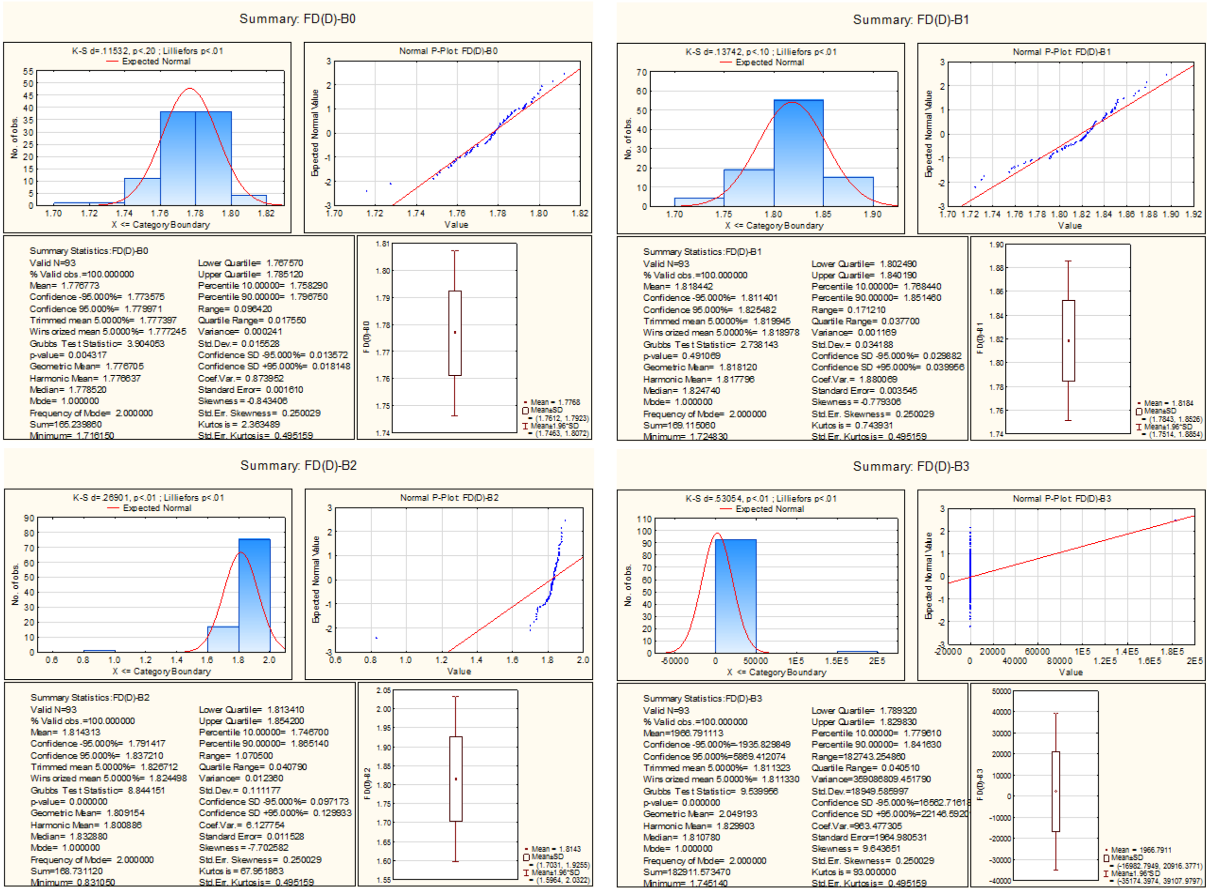
```
Summary Statistics:H.Exp
Valid N=90                             Lower Quartile= 0.633900
% Valid obs.= 96.774194                Upper Quartile= 0.698500
Mean= 4.047934                         Percentile 10.00000= 0.610100
Confidence -95.000%= 0.216787          Percentile 90.00000= 0.747750
Confidence 95.000%= 7.879082           Range=102.444100
Trimmed mean 5.0000%= 0.673412         Quartile Range= 0.064600
Winsorized mean 5.0000%= 0.674632      Variance=334.590624
Grubbs Test Statistic= 5.409635        Std.Dev.= 18.291818
p-value= 0.000000                      Confidence SD -95.000%= 15.954354
Geometric Mean= 0.790546               Confidence SD +95.000%= 21.438117
Harmonic Mean= 0.689690                Coef.Var.=451.880317
Median= 0.676400                       Standard Error= 1.928127
Mode= 1.000000                         Skewness= 5.288528
Frequency of Mode= 2.000000            Std.Err. Skewness= 0.254032
Sum=364.314090                         Kurtosis= 26.560790
Minimum= 0.555900                      Std.Err. Kurtosis= 0.502936
Maximum=103.000000
```

Figure 12. Descriptive Statistics of Hurst exponent (2-adic).



```
Summary Statistics:SIGMA
Valid N=93                          Lower Quartile=  1.138500
% Valid obs.=100.000000             Upper Quartile=  1.242600
Mean=  1.210368                     Percentile 10.00000=  1.112800
Confidence -95.000%=  1.187166      Percentile 90.00000=  1.290800
Confidence 95.000%=  1.233570       Range=  0.880680
Trimmed mean 5.0000%=  1.197263     Quartile Range=  0.104100
Winsorized mean 5.0000%=  1.201198  Variance=  0.012692
Grubbs Test Statistic=  6.527754    Std.Dev.=  0.112659
p-value=  0.000000                  Confidence SD -95.000%=  0.098469
Geometric Mean=  1.205975           Confidence SD +95.000%=  0.131666
Harmonic Mean=  1.202174            Coef.Var.=  9.307858
Median=  1.197300                   Standard Error=  0.011682
Mode=  1.000000                     Skewness=  3.559508
Frequency of Mode=  2.000000        Std.Err. Skewness=  0.250029
Sum=112.564200                      Kurtosis= 19.936709
Minimum=  1.065100                  Std.Err. Kurtosis=  0.495159
Maximum=  1.945780
```

Figure 13. Descriptive statistics of statistical autocorrelations.



```
Summary Statistics:FD(D)-B0
Valid N=93                          Lower Quartile=  1.767570
% Valid obs.=100.000000             Upper Quartile=  1.785120
Mean=  1.776773                     Percentile 10.00000=  1.758290
Confidence -95.000%=  1.773575      Percentile 90.00000=  1.796750
Confidence 95.000%=  1.779971       Range=  0.096420
Trimmed mean 5.0000%=  1.777397     Quartile Range=  0.017550
Winsorized mean 5.0000%=  1.777245  Variance=  0.000241
Grubbs Test Statistic=  3.904053    Std.Dev.=  0.015528
p-value=  0.004317                  Confidence SD -95.000%=  0.013572
Geometric Mean=  1.776705           Confidence SD +95.000%=  0.018148
Harmonic Mean=  1.776637            Coef.Var.=  0.873952
Median=  1.778520                   Standard Error=  0.001610
Mode=  1.000000                     Skewness= -0.843406
Frequency of Mode=  2.000000        Std.Err. Skewness=  0.250029
Sum=165.239860                      Kurtosis=  2.363489
Minimum=  1.716150                  Std.Err. Kurtosis=  0.495159
```

```
Summary Statistics:FD(D)-B1
Valid N=93                          Lower Quartile=  1.802490
% Valid obs.=100.000000             Upper Quartile=  1.840190
Mean=  1.818442                     Percentile 10.00000=  1.768440
Confidence -95.000%=  1.811401      Percentile 90.00000=  1.851460
Confidence 95.000%=  1.825482       Range=  0.171210
Trimmed mean 5.0000%=  1.819945     Quartile Range=  0.037700
Winsorized mean 5.0000%=  1.818978  Variance=  0.001169
Grubbs Test Statistic=  2.738143    Std.Dev.=  0.034188
p-value=  0.491069                  Confidence SD -95.000%=  0.029882
Geometric Mean=  1.818120           Confidence SD +95.000%=  0.039956
Harmonic Mean=  1.817796            Coef.Var.=  1.880069
Median=  1.824740                   Standard Error=  0.003545
Mode=  1.000000                     Skewness= -0.779306
Frequency of Mode=  2.000000        Std.Err. Skewness=  0.250029
Sum=169.115060                      Kurtosis=  0.743931
Minimum=  1.724830                  Std.Err. Kurtosis=  0.495159
```

```
Summary Statistics:FD(D)-B2
Valid N=93                          Lower Quartile=  1.813410
% Valid obs.=100.000000             Upper Quartile=  1.854200
Mean=  1.814313                     Percentile 10.00000=  1.746700
Confidence -95.000%=  1.791417      Percentile 90.00000=  1.865140
Confidence 95.000%=  1.837210       Range=  1.070500
Trimmed mean 5.0000%=  1.826712     Quartile Range=  0.040790
Winsorized mean 5.0000%=  1.824498  Variance=  0.012360
Grubbs Test Statistic=  8.844151    Std.Dev.=  0.111177
p-value=  0.000000                  Confidence SD -95.000%=  0.097173
Geometric Mean=  1.809154           Confidence SD +95.000%=  0.129933
Harmonic Mean=  1.800886            Coef.Var.=  6.127754
Median=  1.832880                   Standard Error=  0.011528
Mode=  1.000000                     Skewness= -7.702582
Frequency of Mode=  2.000000        Std.Err. Skewness=  0.250029
Sum=168.731120                      Kurtosis= 67.951863
Minimum=  0.831050                  Std.Err. Kurtosis=  0.495159
```

```
Summary Statistics:FD(D)-B3
Valid N=93                          Lower Quartile=  1.789320
% Valid obs.=100.000000             Upper Quartile=  1.829830
Mean=1966.791113                    Percentile 10.00000=  1.779610
Confidence -95.000%=-1935.829849    Percentile 90.00000=  1.841630
Confidence 95.000%=5869.412074      Range=182743.254860
Trimmed mean 5.0000%=  1.811323     Quartile Range=  0.040510
Winsorized mean 5.0000%=  1.811330  Variance=359086809.451790
Grubbs Test Statistic=  9.539956    Std.Dev.=18949.585997
p-value=  0.000000                  Confidence SD -95.000%=16562.7161
Geometric Mean=  2.049193           Confidence SD +95.000%=22146.5920
Harmonic Mean=  1.829903            Coef.Var.=963.477305
Median=  1.810780                   Standard Error=1964.980531
Mode=  1.000000                     Skewness=  9.643651
Frequency of Mode=  2.000000        Std.Err. Skewness=  0.250029
Sum=182911.573470                   Kurtosis= 93.000000
Minimum=  1.745140                  Std.Err. Kurtosis=  0.495159
```

Figure 14. Descriptive statistics of fractal dimension of threshold decompositions.

the skeletons of template of A, T, G and C lies in the interval (1.74, 1.80), (1.66, 1.78), (1.65, 1.77) and (1.66, 1.87) respectively. The fractal dimensions of the skeleton of decomposed templates of A and T are normally distributed whereas fractal dimensions of the others template do not follow the same as it happened for threshold decomposition matrices.

In cases of ZFX and ZFY genes, the fractal dimensions of the morphological skeletons of the decomposed threshold matrices are almost same. Interestingly, the same is true for all human X and Y chromosomal genes namely as it is evident from the quantitative details (*Supp. Met. 1*) (Figure 8).

Fractal dimension of protein plots of genes

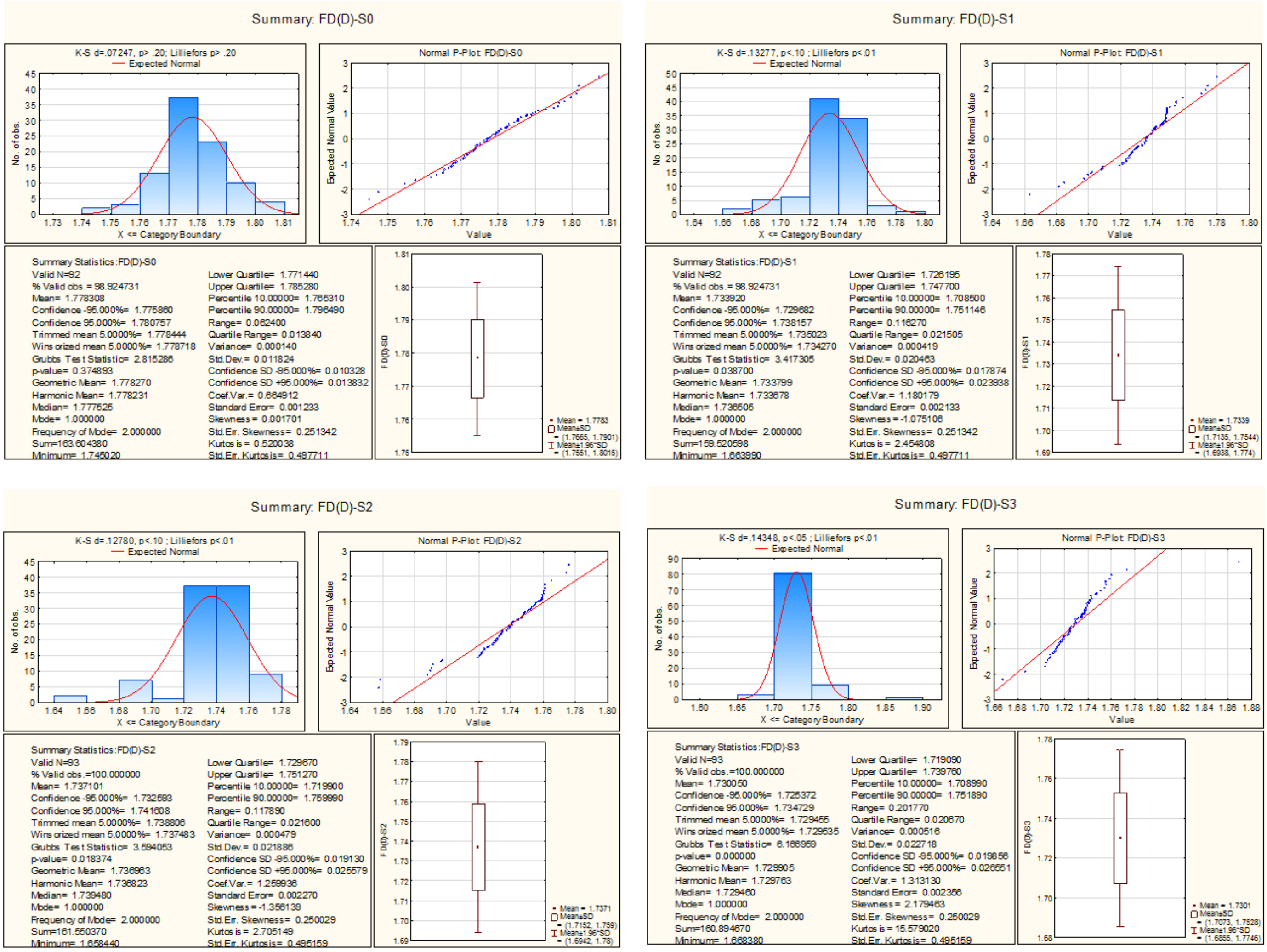
The fractal dimension of the protein plots of Accessible Residues (AR), Buried Residues (BR), Alpha Helix (Chou & Fasman) (ALCF), Amino Acid Composition (ACC), Beta Sheet (Chou & Fasman) (BSCF), Beta Turn (Chou & Fasman) (BTCF), Coil (Deleage & Roux) (CDR), Hydrophobicity (Aboderin) (HA), Molecular Weight (MW), and Polarity (P) of the human X and Y chromosomal genes are lie in the interval (1.81, 1.94), (1.77, 1.94), (1.36, 1.94), (1.79, 1.94), (1.79, 1.94), (1.78, 1.94), (1.84, 1.94), (1.79, 1.94), (1.81, 1.94), (1.81, 1.94), (1.88, 1.94) and (1.81, 1.94) respectively as shown in the descriptive statistics Figure 9.

The fractal dimensions of the protein plots of Accessible Residues (AR), Buried Residues (BR), Amino Acid Composition (ACC), Beta Sheet (Chou & Fasman) (BSCF), Beta Turn (Chou & Fasman) (BTCF), Molecular Weight (MW) and Polarity (P) are following normal distribution (Figure 9).

In cases of the gene pairs (ZFX, ZFY) and (UTX, UTY), the fractal dimensions are agreed almost for all of the protein plots except one or two (Figure 10). In our conviction, these non-agreements make them different as X and Y chromosomal gene. Interestingly, this fact is true for all the X and Y chromosomal homologues of human as evident from the quantitative data (*Supp. Met-2*).

Quantitative classification of human sex genes

All human sex genes including their homologues are classified based on the order of polystring mean ($P_M^{\ N}$) and polystring standard deviation ($P_{SD}^{\ N}$). The genes are tabulated in the *Suppl. Met. 3A and Supp. Met. 3B.* The polystring mean ($P_M^{\ N}$) and polystring standard deviation ($P_{SD}^{\ N}$) of the human sex gene pair (UTX, UTY) is same as



```
Summary Statistics:FD(D)-S0
Valid N=92                              Lower Quartile= 1.771440
% Valid obs.= 98.924731                 Upper Quartile= 1.785280
Mean= 1.778308                          Percentile 10.00000= 1.765310
Confidence -95.000%= 1.775860           Percentile 90.00000= 1.796490
Confidence 95.000%= 1.780757            Range= 0.062400
Trimmed mean 5.0000%= 1.778444          Quartile Range= 0.013840
Winsorized mean 5.0000%= 1.778718       Variance= 0.000140
Grubbs Test Statistic= 2.815286         Std.Dev.= 0.011824
p-value= 0.374893                       Confidence SD -95.000%= 0.010328
Geometric Mean= 1.778270                Confidence SD +95.000%= 0.013832
Harmonic Mean= 1.778231                 Coef.Var.= 0.664912
Median= 1.777525                        Standard Error= 0.001233
Mode= 1.000000                          Skewness= 0.001701
Frequency of Mode= 2.000000             Std.Err. Skewness= 0.251342
Sum=163.604380                          Kurtosis= 0.520038
Minimum= 1.745020                       Std.Err. Kurtosis= 0.497711
```

```
Summary Statistics:FD(D)-S1
Valid N=92                              Lower Quartile= 1.726195
% Valid obs.= 98.924731                 Upper Quartile= 1.747700
Mean= 1.733920                          Percentile 10.00000= 1.708500
Confidence -95.000%= 1.729682           Percentile 90.00000= 1.751146
Confidence 95.000%= 1.738157            Range= 0.116270
Trimmed mean 5.0000%= 1.735023          Quartile Range= 0.021505
Winsorized mean 5.0000%= 1.734270       Variance= 0.000419
Grubbs Test Statistic= 3.417305         Std.Dev.= 0.020463
p-value= 0.038700                       Confidence SD -95.000%= 0.017874
Geometric Mean= 1.733799                Confidence SD +95.000%= 0.023938
Harmonic Mean= 1.733678                 Coef.Var.= 1.180179
Median= 1.736505                        Standard Error= 0.002133
Mode= 1.000000                          Skewness= -1.075106
Frequency of Mode= 2.000000             Std.Err. Skewness= 0.251342
Sum=159.520598                          Kurtosis= 2.454808
Minimum= 1.663990                       Std.Err. Kurtosis= 0.497711
```

```
Summary Statistics:FD(D)-S2
Valid N=93                              Lower Quartile= 1.729670
% Valid obs.=100.000000                 Upper Quartile= 1.751270
Mean= 1.737101                          Percentile 10.00000= 1.719900
Confidence -95.000%= 1.732593           Percentile 90.00000= 1.759990
Confidence 95.000%= 1.741608            Range= 0.117890
Trimmed mean 5.0000%= 1.738806          Quartile Range= 0.021600
Winsorized mean 5.0000%= 1.737483       Variance= 0.000479
Grubbs Test Statistic= 3.594053         Std.Dev.= 0.021886
p-value= 0.018374                       Confidence SD -95.000%= 0.019130
Geometric Mean= 1.736963                Confidence SD +95.000%= 0.025579
Harmonic Mean= 1.736823                 Coef.Var.= 1.259936
Median= 1.739480                        Standard Error= 0.002270
Mode= 1.000000                          Skewness= -1.356139
Frequency of Mode= 2.000000             Std.Err. Skewness= 0.250029
Sum=161.550370                          Kurtosis= 2.705149
Minimum= 1.658440                       Std.Err. Kurtosis= 0.495159
```

```
Summary Statistics:FD(D)-S3
Valid N=93                              Lower Quartile= 1.719090
% Valid obs.=100.000000                 Upper Quartile= 1.739760
Mean= 1.730050                          Percentile 10.00000= 1.708990
Confidence -95.000%= 1.725372           Percentile 90.00000= 1.751890
Confidence 95.000%= 1.734729            Range= 0.201770
Trimmed mean 5.0000%= 1.729455          Quartile Range= 0.020670
Winsorized mean 5.0000%= 1.729535       Variance= 0.000516
Grubbs Test Statistic= 6.166959         Std.Dev.= 0.022718
p-value= 0.000000                       Confidence SD -95.000%= 0.019856
Geometric Mean= 1.729905                Confidence SD +95.000%= 0.026551
Harmonic Mean= 1.729763                 Coef.Var.= 1.313130
Median= 1.729460                        Standard Error= 0.002356
Mode= 1.000000                          Skewness= 2.179463
Frequency of Mode= 2.000000             Std.Err. Skewness= 0.250029
Sum=160.894670                          Kurtosis= 15.579020
Minimum= 1.668380                       Std.Err. Kurtosis= 0.495159
```

Figure 15. Descriptive statistics of fractal dimension of skeleton of threshold decompositions.

tabulated in the *Suppl. Met. 3A. and Supp. Met. 3B*. In contrary, the gene pair (ZFX, ZFY) has the same polystring mean order but polystring SD order of ZFX is TAGC whereas the same of ZFY is TACG. Even for the homologues of CD99 in three different species namely *Homo sapiens, Pan troglodytes and Bos taurus* have the same polystring mean order namely TGCA but the polystring SD of CD99 of *Pan troglodytes and Bos taurus* same but CD99 for human is different from other two. This fact reveals that the ordering of nucleotides in a gene is most important feature to distinguish them from each other and make them unique. The above phenomena are equally true for other human sex genes and their homologues.

Conclusion and future endeavours

In this paper, a quantitative and deterministic detail is adumbrated through which a given string of nucleotides can be inferred as a human X or Y chromosomal gene without seeking any biological experiment. This would help us in screening any given stretch of nucleotides of specific length as a Human sex-gene homologue. This quantitative detail of genomic imprints of sex genes would enable biologist to understand them in more precise way from the very genomic composition level and these understanding are the next challenge of current Genomics. It is noted that the proposed deterministic model is not only meant for human sex genes or its homologue but also can be treated as a standard prototype for other genes and genomes. In our future endeavours we would like to validate the model through the biological experiment.

Authors contributions

Sk. S. Hassan conceptualized the problem and experiments and performed entire research with the rest of the authors of the article.

Acknowledgements

The author authors are indebted to his colleague *Dr. Sudhakar Sahoo* and the former Director *Professor Swadheenananda Pattanayak* of Institute of Mathematics and Applications, Bhubaneswar for their kind help and suggestions.

Conflicts of interest

We all authors of the manuscript certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

References

1. Watson JD (1990) The human genome project: past, present, and future. *Science* 248: 44-49.[Crossref]
2. Sawicki MP, Samara G, Hurwitz M,Passaro E Jr (1993) Human Genome Project. *Am J Surg* 165: 258–264.
3. Collins FS, Morgan M, PatrinosA (2003) The Human Genome Project: Lessons from Large-Scale Biology. *Science* 300: 286-290.[Crossref]
4. Collins FS, McKusick VA (2001) Implications of the Human Genome Project for medical science. *JAMA* 285: 540-544.[Crossref]
5. Friedmann T, Roblin R (1972) Gene therapy for human genetic disease? *Science* 175: 949-955.[Crossref]
6. U. S. National Library of medicine (2012) “Genetics Home Reference”.
7. Andrea Ballabio, David Nelson and Steve Rozen (2006) Genetics of disease: The sex chromosomes and human disease. *Curr Opin Genet Dev* 16: 1–4. [Crossref]
8. Zhao L, Koopman P (2012) SRY protein function in sex determination: thinking outside the box. *Chromosome Res* 20: 153-162.[Crossref]
9. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.[Crossref]
10. Chandra HS (1985) Is human X chromosome inactivation a sex-determining device? *Proc Natl Acad Sci U S A* 82: 6947-6949.[Crossref]
11. Darling SM, Banting GS, Pym B, Wolfe J, Goodfellow PN (1986) Cloning an expressed gene shared by the human sex chromosomes. *Proc Natl Acad Sci U S A* 83: 135-139.[Crossref]
12. Lu X, Wu CI (2011) Sex, sex chromosomes and gene expression. *BMC Biol*9: 30.[Crossref]
13. Kent J, Wheatley SC, Andrews JE, Sinclair AH, Koopman P (1996) A male-specific role for SOX9 in vertebrate sex determination. *Development* 122: 2813-2822.[Crossref]
14. Mandelbrot BB (1982) “The fractal geometry of nature”. New York, ISBN 0-7167-1186-9.
15. Avnir D (1998) “Is the geometry of Nature fractal”. *Science* 279: 39.
16. Develi K, Babadagli T (1998) Quantification of natural fracture surfaces using fractal geometry. *Math Geology* 30: 971-998.
17. Hassan SS, Choudhury PP, Goswami A (2013) Underlying mathematics in diversification of human olfactory receptors in different loci. *Interdiscip Sci* 5: 270-273.[Crossref]
18. Hassan SS, Pal Choudhury P, DayaSagar BS, Chakraborty S, Guha R, et al. (2011) Quantitative Description of Genomic Evolution of Olfactory Receptors, (Communicated).
19. Carlo C (2010) Fractals and Hidden Symmetries in DNA. *Math Prblm in Engng* 507056.
20. Zu-Guo Y (2002) Fractals in DNA sequence analysis. *Chinese Physics* 11: 1313-1318.
21. deMelo RHC, Conci A (2008) Succolarity: Defining a Method to calculate this Fractal Measure.” ISBN: 978-80-227-2856-0 291-294.
22. Deschavanne PJ, Giron A, Vilain J, Fagot G (1999) Genomic signature: characterization and classification of species assessed by chaos game representation of sequences. *Mol Biol Evol* 16: 1391-1399. [Crossref]