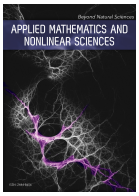
sciendo Applied Mathematics and Nonlinear Sciences 4(1) (2019) 267–278



Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

The self-similarity properties and multifractal analysis of DNA sequences

G. Durán-Meza [1], J. López-García [2] , J.L. del Río-Correa [1]

1 Departamento de Fśica, Universidad Autónoma Metropolitana Iztapalapa,
Avenida San Rafael Atlixco 186, Colonia Vicentina, 09340 Mexico City, Mexico

2 División de Matemáticas e Ingenier ía, UNAM-FES Acatlán, Estado de Mexico

Submission Info

Communicated by Jose Maria Amigo Garcia
Received December 15th 2018
Accepted April 7th 2019
Available online June 29th 2019

Abstract

In this work is presented a pedagogical point of view of multifractal analysis deoxyribonucleic acid (DNA) sequences is presented. The DNA sequences are formed by 4 nucleotides (adenine, cytosine, guanine, and tymine). Following Jeffrey's paper we associated a simple contractive function to each nucleotide, and constructed the Hutchinson's operator W, which was used to build covers of different sizes of the unitary square Q, thus $W^k(Q)$ is a cover of Q, conformed by 4^k squares Q_k of size 2^{-k}, as each Q_k corresponds to a unique subsequence of nucleotides with length $k : b_1 b_2 ... b_k$. Besides, it is obtained the optimal cover C_k to the fractal F generated for each DNA sequence was obtained. We made a multifractal decomposition of C_k in terms of the sets J_α conformed by the Q_k's with the same value of the Holder exponent α, and determined $f(\alpha)$, the Hausdorff dimension of J_α, using the curdling theorem.

Keywords: Multifractals DNA sequences self-affine Hutchinson operator Holder exponents.
MSC codes: 37A60, 37D40.

1 Introduction

The discovery of deoxyribonucleic acid (DNA) structure lies as a fundamental topic of 20th century biology and continues to be the favourite question of some contemporary scientists. Deoxyribonucleic Acid or DNA is a long linear polymer, formed by a large number of nucleotides. Each nucleotide contains a phosphate group, a sugar group and a nitrogen base. The four types of nitrogen bases are adenine (A), thymine (T), guanine (G) and cytosine (C). The uracil base (U) is presented in the RNA molecule. The DNA adopts a double helix form, which is a helical structure constituted by two complementary strands of nucleic acids. For this reason, DNA is also called α-helix.

Shortly speaking, DNA is a macromolecule formed by a large number of nucleotides. Thus, the genetic information is stored in the sequence of the bases. Therefore, the order of the bases determines the information available for the RNA production and proteins.

sciendo ISSN 2444-8656 doi:10.2478/AMNS.2019.1.00023 sciendo

Powerful methods for DNA sequencing have been developed. These methods serve to sequence complete genomes. For example, small genomes from viruses or fungi, and large genomes, such as the human genome, are made up of 3000 million base pairs [1].

In the National Center for Biotechnology Information (NCBI) directories are the databases that contain complete genomes, complete sequences of chromosomes, sequences of mRNAs, and proteins. The importance to analyse the large DNA databases in the Nonlinear Dynamics context is based on the work conducted earlier by Jeffrey [2], who proposed a graphic representation of these databases via an extended chaos game. Other contributions similarly based on a statistical description of DNA sequences take on a more structured form, such as Zu-Guo Yu et al [3], where the generalized dimensions D_q and its derivative, the *'analogous'* specific heat C_q , are calculated for the coding and noncoding length sequences of bacteria.

The Multifractal formalism was originally proposed to study various chaotic models derived from phenomena associated with turbulence [4]. This theory is used as an archetype of fractal measurements composed of interwoven fractal sets characterized by the Holder exponent, and it has become a crucial tool in the analysis of statistical data.

Section 2 is be dedicated to give the theoretical support of the DNA chaos game paraphrasing the ideas of Barnsley about definition of measures on fractals. In section 3 is showed a didactic approach using a pretty small genome applying the ideas of the last section. Sections 4 and 5 involves the use of these ideas to give a multifractal interpretation to the DNA sequences. In section 5, we discuss the coarse grained multifractal theory and using the curdling theorem. The alternative definition of the singularity spectra used in the Chabbra-Jensen algorithm are obtained. In section 6, we show the multifractal spectra for the six-mers of two bacterias, two archaea, the Homosapiens chromosome 21 and a fungus, and discuss the information contained in the singularity spectra for six DNA sequences.

2 DNA Chaos Game

We construct an unitary square Q with its corners labeled with a different basis of the genomic sequence $V = \{A, C, G, T\}$; where V indicates the vertices of Q on the cartesian plane which are $A = (0,0)$, $C = (0,1)$, $G = (1,1)$ and $T = (1,0)$. It is applied the Jeffrey's *"chaos game algorithm"*, and the Chaos Game Representation (CGR) [2] is obtained. It provides an objective meaning for studying DNA sequences in the context of fractal geometry and non-linear science.

The analytical process for the construction of CGR belongs to a DNA organism sequence that begins with the choice of an arbitrary point inside the square Q. Let $P(x_0, y_0)$ be the starting point. For simplicity, we take the middle point of Q, and let $V(x_V, y_V)$ be any corner of Q. $P(x_1, y_1)$ are the coordinates of a middle point of the segment $\overline{PV}$. These coordinates are given by,

$$\begin{aligned} x_1 &= \frac{1}{2}(x_0 + x_V) \\ y_1 &= \frac{1}{2}(y_0 + y_V) \end{aligned} \tag{1}$$

The matrix function of this process can be defined by

$$P_1 = \omega_V(P_0) = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \frac{x_V}{2} \\ \frac{y_V}{2} \end{pmatrix}. \tag{2}$$

The next points are produced by successive applications of the same process.

$$P_n = \omega_{V_n}(P_{n-1}) \;\; with \;\; P_0 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}. \tag{3}$$

It means that the DNA sequence can be expressed as a sequence of points produced by iterative applications of (3). Notice that the graph of the set of points P_n forms the CGR.

We plot the CGR of three different organisms in Fig. 1. This shows visually that the DNA sequence has some fractal patterns and a self-affine characteristics. Thus, it shows that the DNA sequence is not a random sequence.

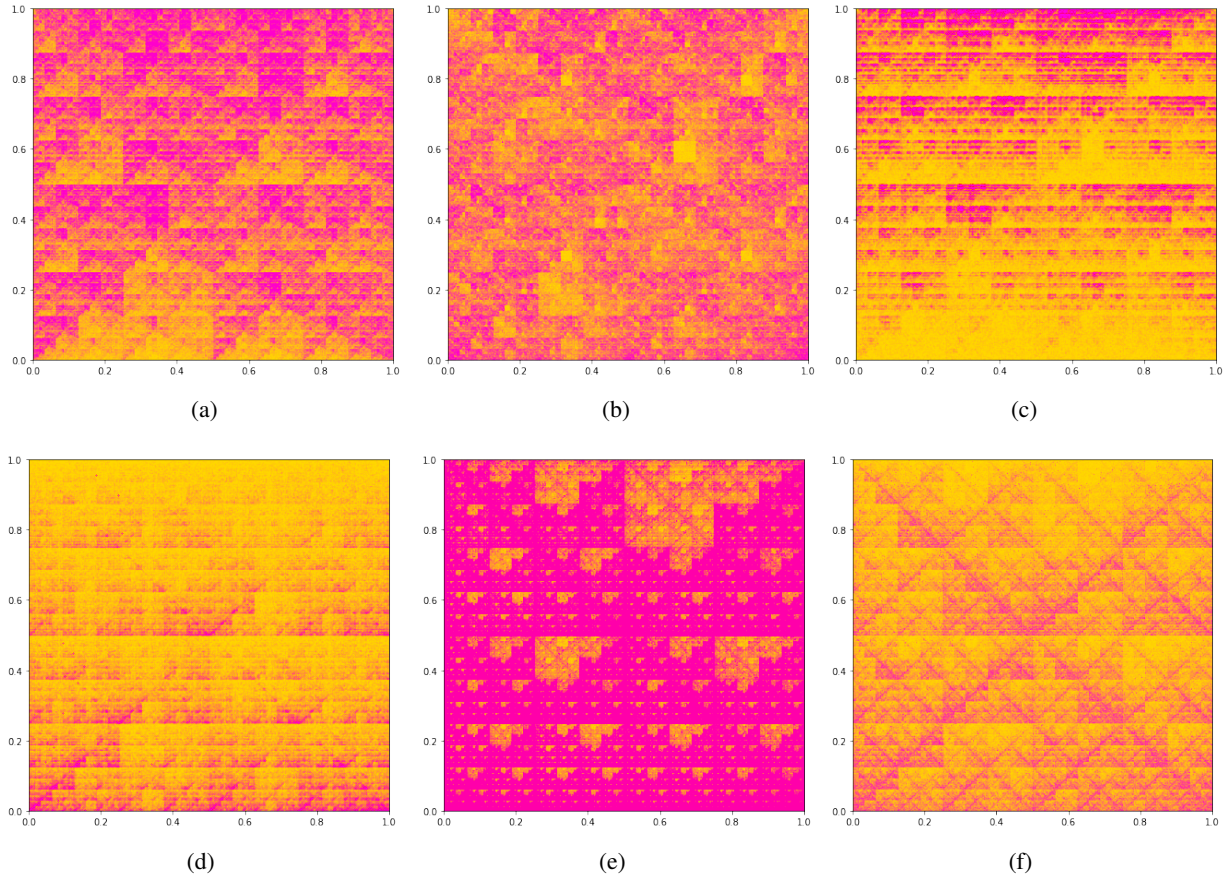


Fig. 1 The CGR of two bacterias (*Pirellula staleyi* and *Escherichia coli*), two archaea (*Nanoarchaeota archaeon* and *Halobacterium salinarum*), the *Homo sapiens chromosome 21* and a fungus *Encephalitozoon intestinalis*.

An IFS is associated with a DNA sequence. We construct four affine transformations [5] to explore regions in Q.

$$\begin{aligned}
\omega_A &= \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\omega_C &= \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix} \\
\omega_G &= \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \\
\omega_T &= \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}
\end{aligned} \tag{4}$$

It is important to note that there is a relationship between the chaos game algorithm and the Bernoulli

mapping. This is possible, thanks to the chaos game properties. The sequence of points $\mathbf{P} = \{P_1 P_2 \dots P_n\}$ and the sequence of DNA have a one-to-one relation. The $\mathbf{P}$ is unique for each genome. Knowing the coordinates of P_n, the coordinates of P_{n-1} can be determined by using a binary representation for the x and y coordinates. It is possible to show that the inverse relationship of (1) is:

$$\begin{aligned} x_{n-1} &= 2x_n \mod 1 \\ y_{n-1} &= 2y_n \mod 1 \end{aligned} \tag{5}$$

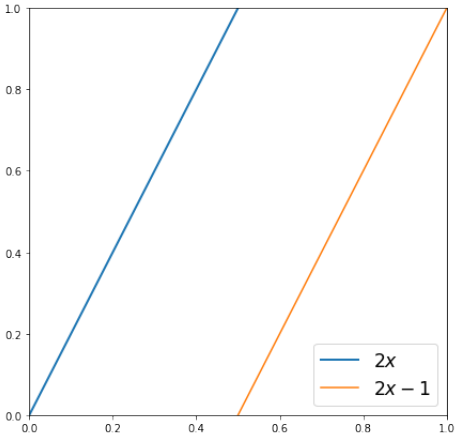


Fig. 2 Bernoulli map.

Any point of $\mathbf{P}$ contains the information of which are all the bases of DNA sequence up to that point. Since, knowing P_n, all the points prior to it can be calculated, and using each point corresponds to a base of the genome which are the bases before the $n-th$.

The Hutchinson's operator of (4) is $W(B) = \omega_A(B) \bigcup \omega_C(B) \bigcup \omega_G(B) \bigcup \omega_T(B)$ where $B \subset Q$. The application of each function to Q provides an square $Q_V = \omega_V(Q)$ with side $1/2$. The action of Hutchinson's operator on Q forms the unitary square, because the union of four squares of side $1/2$ is again Q. In other words, Q is invariant under W [6].

$$Q = W(Q) \tag{6}$$

When we apply W twice to Q, we have

$$W^2(Q) = \bigcup_{V_2} \bigcup_{V_1} \omega_{V_2} \circ \omega_{V_1}(Q) = \bigcup_{V_2} \bigcup_{V_1} Q_{V_2 V_1} \tag{7}$$

in this way, a grid with 4^2 squares is obtained,, each one with side $(1/2)^2$. For m applications of W, we obtain 4^m sub-squares of side $(1/2)^m$, which conform Q.

We have considered subregions of Q instead of chaos game钬榿points, due to the fact that P_0 is the midpoint of Q, $P_1 = \omega_V(P_0)$ is the midpoint of Q_{V_1}, $P_2 = \omega_{V_2} \circ \omega_{V_1}(P_0)$ is the midpoint of $Q_{V_2 V_1}$, and so on. Thus, $P_N = \omega_{V_N} \circ \cdots \circ \omega_{V_1}(P_0)$ is the midpoint of $Q_{V_N \cdots V_1}$.

Then, each DNA sequence generates a large collection of squares of sides $\left[(1/2), (1/2)^2, ..., (1/2)^N\right]$. The squares satisfy the following property:

$$Q_{V_N V_{N-1} \cdots V_2 V_1} \subset Q_{V_N V_{N-1} \cdots V_2} \subset \circ \circ \circ \subset Q_{V_N V_{N-1}} \subset Q_{V_N} \subset Q. \tag{8}$$

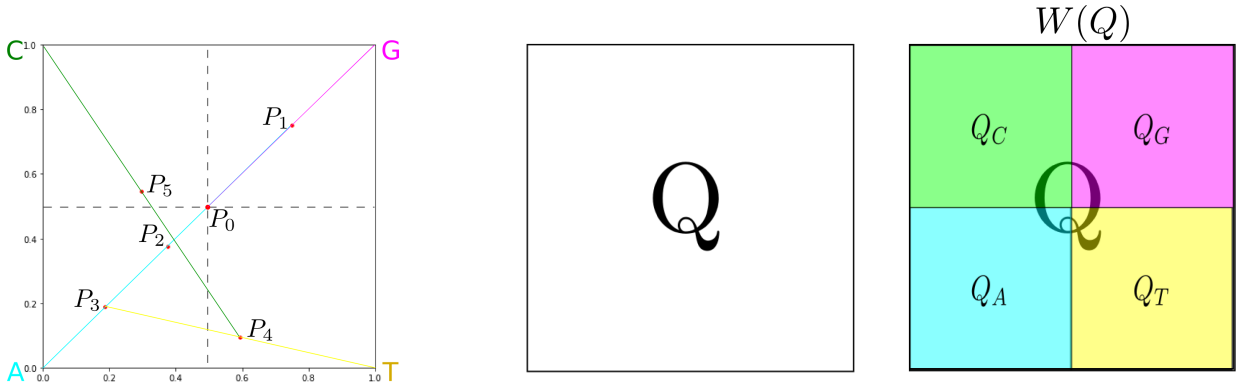
Consequently, we can observe that the number of monomers $V = (A, C, G, T)$ in a DNA sequence matches the number of points inside Q_V and the number of dimers $V_1 V_2$ matches with the number of points inside $Q_{V_2 V_1}$. On the whole, the number of R-mers is given by the number of points inside of $Q_{V_R \cdots V_2 V_1}$.

The assertion that the DNA sequence is not a random sequence involves the appearance of statistical techniques for the analysis of the different distributions of the DNA sequences.

3 A simple example

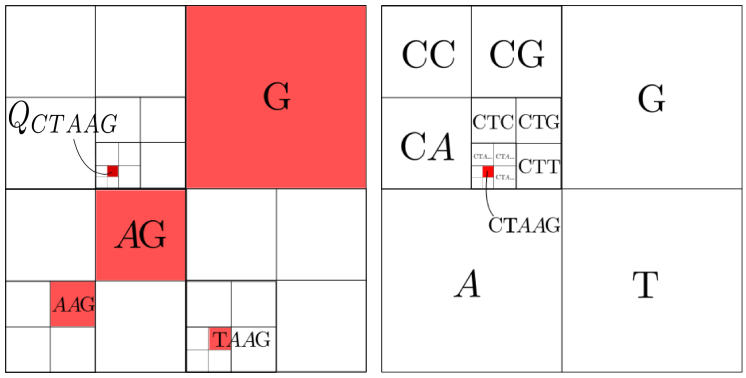
This section exemplifies how chaos game works. Through a playful model, we propose to illustrate the theoretical support of DNA chaos game. In this context, a small DNA sequence *GAATC* it is a toy genome.

Lets us take our toy *GAATC*. As we have pointed out in (2), the sequence points of *GAATC* are: $\{P_1 = \omega_G(P_0), P_2 = \omega_A(P_1), P_3 = \omega_A(P_2), P_4 = \omega_T(P_3), P_5 = \omega_C(P_4)\}$, we can see its graph or CGR in figure 3(a).



(a) (b)

Fig. 3 (a): The CGR of *GAATC*. (b): Application of the Hutchinson's operator to the square.



(a) (b)

Fig. 4 Subregion's description of *GAATC*, in (a) application of functions, in (b) localization.

The application of Hutchinson's operator to Q and the SFI (4) may be illustrated by the figure 3(b). Therefore, we can show the subregion's descriptions for this small sequence, where we have obtained sub-squares of side $\left[(1/2),(1/2)^2,(1/2)^3,(1/2)^4,(1/2)^5\right]$ in figure 4. A tiny sequence provides a didactic point of view to motivate the reader to make his or her own games on the computer.

We must emphasize that, there exist a subtle distinction between two processes. One of them is the functions' application, it means that the application of (4) is given according to the characters apparition order in the sequence and it is described graphically in the figure 4(a). The other corresponds to the location of sub-squares, where given a sub-square, we can find it, using the property (8). It means that in the localization processes, the sequence must be read in a reverse order as shown in the figure 4(b).

4 DNA sequence's singular measure

Figure 1 shows that a fractal F could be generated by a certain DNA sequence and F is contained in Q. From the graphical point of view, 21 Human chromosome's CGR is the practical proof of self-similarity. If we look closely 1(d), we will notice that the upper right quadrant Q_G, is repeated at every subregion on Q but in reduced scales and different tonalities. We see that the tonalities vary in some nontrivial manner; it can be related to a complex measure. Therefore, this is our motivation to use the multifractal theory.

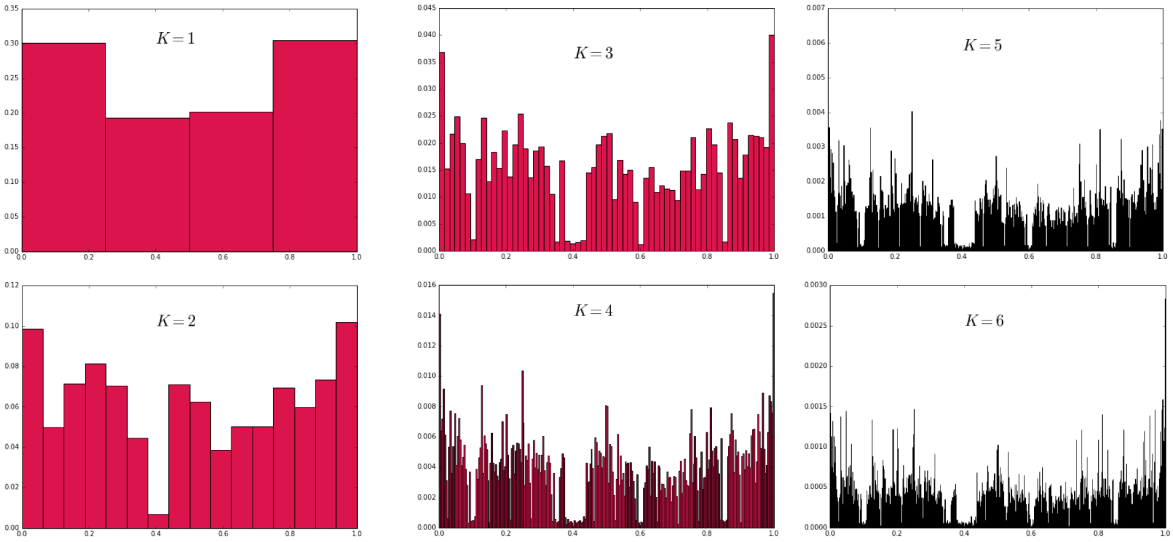


Fig. 5 21 Human chromosome histograms for $R = 1$ to $R = 6$.

The description of subregions provides a systematic way for the construction of a cover C_R, this is called the optimal cover of F. The C_R is composed of squares Q_R of side λ_R, each Q_R has a statistical measure $\mu(Q_R) \neq 0$ given by the frequency with which the DNA sequence visits Q_R. The aim of this description is that each Q_R of the cover C_R is associated with a R-mer of the DNA sequence.

Summarizing, we can say that a R-mer is a DNA subsequence of size R. It means, that this subsequence has R basis, if it is within the DNA sequence, then the probabilistic measure of a R-mer depends on the total number with which this appears in the DNA sequence. The cover C_R gives the number of all the different possible R-mers: 4^R.

In order to get the frequency of each different R-mer of a DNA sequence with N basis, we denote $N_{V_1V_2...V_R}$ as the number of times R-mer appears in the total sequence. It can be easily noticed that the total number of R-mers in the entire sequence is $N-R+1$. Therefore the frequency of a given subsequence is

$$F_{V_1V_2...V_R} = \frac{N_{V_1V_2...V_R}}{N-R+1} \tag{9}$$

Now, we may say by the fact that the R-mer is inside of the square $Q_{V_1V_2...V_R}$, that is its probabilistic measure μ and its side λ are:

$$\mu_{V_1V_2...V_R} = F_{V_1V_2...V_R} \quad \& \quad \lambda(V_1V_2...V_R) = \left(\frac{1}{2}\right)^R \tag{10}$$

To understand R-mers statistical behaviour more thoroughly, it is necessary to make an intermediate step where we establish a one-to-one relationship between sub-squares of Q and sub-intervals of the unit interval [6].

Thus, we use a quaternary basis associating each nitrogenous base with a number: $A = 0$, $C = 1$, $G = 2$ and $T = 3$, then the DNA sequence becomes a sequence of digits. This makes it easy the calculation of the

frequencies in the environment of a programming language (python 3.6.0). We have plotted the histograms of 21 Human chromosome in figure 5. We see the same behavioral patterns as in a multiplicative process' measure.

5 Multifractal analysis

In this section we use the coarse grained multifractal theory to characterize the different R-mers that are contained in the DNA sequence, as each R-mer has associated a sub-square Q_R of size $(1/2)^R$, these sub-squares conform the cover $C_R = \{Q_R\}$ of the fractal F generated by the DNA sequence. Each Q_R has a coarse grained Holder's exponent, which is given by:

$$\alpha(Q_{V_1V_2\ldots V_R}) = \frac{\ln \mu(Q_{V_1V_2\ldots V_R})}{\ln \lambda(Q_{V_1V_2\ldots V_R})} = \frac{\ln \mu(Q_{V_1V_2\ldots V_R})}{\ln \lambda_R} \tag{11}$$

The fractal F is covered by interwoven subsets J_α, which are defined as:

$$J_\alpha = \{Q_{V_1V_2\ldots V_R} | \alpha(Q_{V_1V_2\ldots V_R}) \in (\alpha, \alpha + \Delta\alpha)\} \tag{12}$$

The pre-fractal dimension of each J_α is given by

$$f_R(\alpha) = -\frac{\ln N_R(\alpha)}{\ln \lambda_R} \tag{13}$$

where $N_R(\alpha)$ is the cardinality of J_α. The multifractal spectra of the cover means the knowledge of the pre-fractal dimension of all subsets J_α. We determine these spectra using the Chabbra-Jensen algorithm [7].

More details of the algorithm than we present in this paper can be found in [8]. In this algorithm, is fundamental the definition of the escort probability measure of $\mu(Q_{V_1V_2\ldots V_R})$ of order q is fundamental, which is given by [9]:

$$\mu(q, Q_{V_1V_2\ldots V_R}) = \frac{[\mu(Q_{V_1V_2\ldots V_R})]^q}{Z_R(q)} \tag{14}$$

where

$$Z_R(q) = \sum_j [\mu(Q_{V_1V_2\ldots V_R})]^q \tag{15}$$

The sum over j is taken over all R-mers with a measure different from zero. Taking into account that:

$$if \quad Q_{V_1V_2\ldots V_R} \subset J_\alpha \quad then \quad \mu_\alpha(Q_{V_1V_2\ldots V_R}) = \lambda_R^\alpha \tag{16}$$

and that $Z_R(q)$ satisfies the power law:

$$Z_R(q) = \lambda_R^{\tau(q)} \quad or \quad \tau(q) = \frac{\ln Z_R(q)}{\ln \lambda_R} \tag{17}$$

The q-measure of a set J_α is given by:

$$\mu(q, J_\alpha) = N_R(\alpha)\mu_\alpha(q, Q_{V_1V_2\ldots V_R}) = \lambda_R^{-f_R(\alpha)} \frac{[\mu(Q_{V_1V_2\ldots V_R})]^q}{Z_R(q)} = \lambda_R^{q\alpha - \tau(q) - f_R(\alpha)} \tag{18}$$

As the q-measure only takes values between zero and one, then

$$\tau(q) \leq q\alpha - f(\alpha) \tag{19}$$

For all values of λ_R, the q-measure obtains its maximal value for a special value $\alpha^*(q)$, which satisfies:

$$\tau(q) = q\alpha^*(q) - f(\alpha^*(q)) \tag{20}$$

We note that (19) and (20) imply that

$$\tau(q) = \inf_{\alpha}(q\alpha - f(\alpha)) \tag{21}$$

where the infimum is taken over all possible values of α. This expression establishes that $\tau(q)$ and $f(q)$ are a couple of Legendre transforms.

The behaviour of the q-measure around $\alpha^*(q)$ is found using a logarithmic expansion of (18):

$$\ln\mu(q,J_\alpha) = \ln\mu(q,J_{\alpha^*}) + (\alpha-\alpha^*)\left\{q - \frac{df(\alpha)}{d\alpha}\right]_{\alpha^*}\right\}\ln\lambda_R + \frac{(\alpha-\alpha^*)^2}{2}\left\{-\frac{d^2f(\alpha)}{d\alpha^2}\right]_{\alpha^*}\right\}\ln\lambda_R \tag{22}$$

Then, the results of this expansion are:

$$q = \frac{df(\alpha)}{d\alpha}\Big]_{\alpha^*(q)} \quad and \quad -\ln\lambda_R\frac{d^2f(\alpha)}{d\alpha^2}\Big]_{\alpha^*(q)} < 0 \tag{23}$$

Therefore, the q-measure for values of α near α^* takes the form:

$$\mu(q,J_\alpha) = \frac{1}{\sqrt{2\pi\sigma_\alpha}}\exp-\frac{(\alpha-\alpha^*(q))^2}{2\sigma_\alpha^2} \quad with \quad \sigma_\alpha^2 = \|f''(\alpha^*(q))\ln\lambda_R\| \tag{24}$$

It shows that q-measure concentrates around $J_\alpha^*(q)$, and when $\lambda_R \to 0$, J_α^* is the support of the q-measure, i.e.

$$\lim_{\lambda_R\to 0}\mu(q,J_\alpha) = \delta(\alpha-\alpha^*(q)) \tag{25}$$

then, J_{α^*} is the curdling set of the q-measure, and the result (21) is called the curdling theorem [8].

The value of $\alpha^*(q)$ is found taking the derivative of (20) and using (22).

$$\alpha^*(q) = \frac{d\tau(q)}{dq} = \frac{1}{\ln\lambda_R}\frac{d\ln Z_R(q)}{dq} \tag{26}$$

where in the second term of the r.h.s. of (26), (17) was used. Finally, using (14) and (15), we obtain that:

$$\alpha^*(q) = \frac{1}{\ln\lambda_R}\sum_{V_1V_2\ldots V_R}\mu(q,Q_{V_1V_2\ldots V_R})\ln\mu(Q_{V_1V_2\ldots V_R}) \tag{27}$$

The fractal dimension of J_{α^*} is determined using (20). i.e.

$$f(\alpha^*(q)) = q\alpha^*(q) - \tau(q) = q\alpha^*(q) - \frac{\ln Z_R(q)}{\ln\lambda_R} \tag{28}$$

introducing (25) in the last expression, we obtained:

$$f(\alpha^*(q)) = \frac{1}{\ln\lambda_R}\sum_{V_1V_2\ldots V_R}\mu(q,Q_{V_1V_2\ldots V_R})\ln\mu(q,Q_{V_1V_2\ldots V_R}) \tag{29}$$

The relations (27) and (29) are the alternative definition of the singularity spectrum used in the Chhabbra and Jensen algorithm. The curve $f_R(\alpha)$ versus α is obtained when the q-parameter is eliminated, but these expressions contain more information about the singularity spectra, as we will discuss in the next section.

6 Concluding Remarks

We begin discussing the information contained in the Chhabbra-Jensen algorithm, considering different values of the q-parameter. The simplest case is $q = 0$, when the q-measure is given by:

$$\mu\left(q=0, Q_{V_1V_2\ldots V_R}\right) = \frac{1}{Z_R(q=0)} = \frac{1}{4^R} \tag{30}$$

It corresponds to consider that all the R-mers have the same probability. The value of the fractal dimension of the curdling set, is evaluated using (29), then

$$f(\alpha^*(q=0)) = \frac{1}{\ln\lambda_R}\left(4^R\right)\mu(q=0,Q_R)\ln\mu(q=0,Q_R) = \frac{-R\ln 4}{-R\ln 2} = 2 \tag{31}$$

This is the fractal dimension of Q which is the support of F, and it is the maximum value of the fractal dimension. As follows from (27), the value of the $\alpha^*(q=0)$ is given by:

$$\alpha^*(q=0) = \frac{1}{\ln\lambda_R}\frac{1}{4^R}\sum_{V_1V_2\ldots V_R}\ln\mu(Q_{V_1V_2\ldots V_R}) = \frac{1}{4^R}\sum_{V_1V_2\ldots V_R}\alpha(Q_{V_1V_2\ldots V_R}) \tag{32}$$

This is the average of the Holder's exponent of all the members of the cover C_R. When $q = 1$, the q-measure reduces to the measure of the Q_R, i.e. $\mu(q=1,Q_{V_1V_2\ldots V_R}) = \mu(Q_{V_1V_2\ldots V_R})$. From (27) and (29), we have

$$\alpha^*(q=1,Q_{V_1V_2\ldots V_R}) = f(\alpha^*(q=1)) = -\frac{1}{R\ln 2}\sum_{V_1V_2\ldots V_R}\mu(Q_{V_1V_2\ldots V_R})\ln\mu(Q_{V_1V_2\ldots V_R}) \tag{33}$$

Thus, for the measure $\mu(Q_R)$, the fractal dimension of the curdling set J_{α^*} is identical with the value of the Holder's exponent which characterize this set, but this value is given by the average entropy of the cover C_R of it. We analyse the behavior of $\mu(q,Q_R)$ when q goes to infinity. For large values of q, the measure that gives the maximum contribution to $Z_R(q)$ is

$$\mu_{MAX} = \max_{V_1V_2\ldots V_R}\mu(Q_{V_1V_2\ldots V_R}) = \max_{V_1V_2\ldots V_R}\left(\frac{1}{2^R}\right)^{\alpha(V_1V_2\ldots V_R)} = \frac{1}{2^{R\alpha_{min}}} \tag{34}$$

Then, for large values of q, we have

$$Z_R(q) \approx N_R(\alpha_{min})\left[\mu_{MAX}(Q_R)\right]^q \tag{35}$$

Thus, using this result into (14), we obtain

$$\mu(q,Q_{V_1V_2\ldots V_3}) = \frac{1}{N_R(\alpha_{min})}\left[\frac{\mu(Q_{V_1V_2\ldots V_3})}{\mu_{MAX}(Q_R)}\right]^q \tag{36}$$

Therefore,

$$\lim_{q\to\infty}\mu(q,Q_{V_1V_2\ldots V_3}) = \frac{1}{N_R(\alpha_{min})}\delta(\mu(Q_{V_1V_2\ldots V_3}) - \mu_{MAX}(Q_R)) \tag{37}$$

Thus, when it is taking the limit of (27) when $q \to \infty$, we obtain

$$\lim_{q\to\infty}\alpha(q) = \frac{\ln\mu_{MAX}(Q_R)}{\ln\lambda_R} = \alpha_{\min}(Q_R) \tag{38}$$

Similarly, using the fact that $\mu_{\min}(Q_R) = \lambda_R^{\alpha_{\max}}$, the behavior of the Holder's exponent can be obtained when $q \to -\infty$, i.e.

$$\lim_{q\to\infty}\alpha(q) = \alpha_{\min}(Q_R) \tag{39}$$

The important characteristics of the fractal spectra $f_R(\alpha)$ are the following:

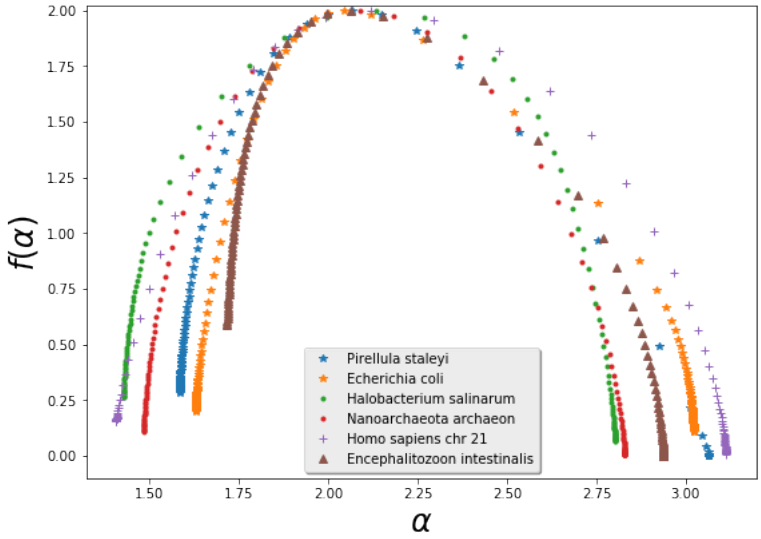


Fig. 6 Multifractal spectra of two bacterias: *Pirellula staleyi* and Escherichia coli, two archaea: *Nanoarchaeota archaeon* and *Halobacterium salinarum*, the *Homo sapiens chromosome 21* and a fungi *Encephalitozoon intestinalis*

1. The curve is always convex upward.
2. The q parameter is the slope of the straight line which is tangent to the curve $f_R(\alpha^*)$ in any value of α^*. Both properties follow from (23).
3. The equation of this straight line is $y = q[\alpha - \alpha^*(q)] + f(\alpha^*(q))$. Therefore, it intersects the y-axis in $-\tau(q)$, as follows from (22).
4. The maximum value of $f_R(\alpha)$ occurs at $\alpha^*(q=0) = \alpha_0$. At this point f is equal to the fractal dimension of the support.
5. The $f_R(\alpha)$ curve is tangent to the line $y = \alpha$, the point of tangency occurs at α_1, this value corresponds with the information of fractal dimension of the curdling set.
6. The range of values of α are given by $\alpha_{\min} = \alpha(q \to \infty) \leq \alpha \leq \alpha_{\max}(q \to \infty)$.

We conclude that there are at least four important values of the Holder's exponent, $\alpha_0, \alpha_1, \alpha_{\min}$ and $\alpha_{\max}$, which characterize globally the fractal spectra. The value α_0 where $f_R(\alpha)$ reaches its maxima; the value of α_1 which determines the fractal dimension of the curdling set, the other two values are used for determining the width of the spectrum $W = \alpha_{\max} - \alpha_{\min}$, and the skew shape of the spectrum $r = (\alpha_{\max} - \alpha_0)/(\alpha_0 - \alpha_{\min})$. The skew parameter r determines which fractal exponents are dominant. Right skew shape of the spectrum with $r > 1$ is more complex than the left skew shape with $r < 1$ [10–12].

The geometrical shape of the multifractal spectra illustrates the level of multifractality of the DNA sequences [10]. In figure 6, we show the multifractal spectra of the sixth-mers of several DNA sequences, all of them were constructed with the range of q's between $q_{\min} = -30$ and $q_{\max} = 30$; with steps $\Delta q = 0.5$.

The values of the quantities that we select for evaluating the "randomness" are reported in Table 1. The maximum of all curves is 2; which is due to that all take the square Q as the support of F. The parameter r shows that the *Halobacterium salinarum*' spectrum has a greater symmetry than the others, and the rest have a right-skewed shape. The bacteria have a higher value of W and r, while the archaea have a lower value of them. This suggests that the multifractal spectrum of bacteria presents a greater complexity than the spectrum of archaea. This suggests that the use of the multifractal technique could be used for a quantitative classification of archaea and bacteria.

Table 1 Relevant parameters of multifractals of Fig. 6 : the value of α where the maximum value occurs α_0, the wide of spectrum W and the spectrum skew shape r.

Organism	α_0	W	r
Pirellula staleyi	2.0673621898	1.4825717654943124651	2.06044807855
Escherichia coli	2.04444319969	1.394197126734667779	2.34682776089
Halobacterium salinarum	2.13387945312	1.3762064622503003001	0.944824001754
Nanoarchaeota archaeon	2.08756863398	1.346437367927127315	1.22732173982
Homo sapiens chr 21	2.11728218486	1.705128432757234025	1.40041368164
Encephalitozoon intestinalis	2.06212028856	1.222195891979336313	2.53446680707

References

[1] J.M. Berg, L. Stryer, J.L. Tymoczko, and J.M. Macarulla. *Bioquímica*. Reverté, 2007.

[2] H.Joel Jeffrey. Chaos game visualization of sequences. *Computers & Graphics*, 16(1):25 – 33, 1992.

[3] Zu-Guo Yu, Vo Anh, and Ka-Sing Lau. Multifractal characterisation of length sequences of coding and noncoding segments in a complete genome. *Physica A: Statistical Mechanics and its Applications*, 301(1):351 – 361, 2001.

[4] U. Frisch and A.N. Kolmogorov. *Turbulence: The Legacy of A. N. Kolmogorov*. Cambridge University Press, 1995.

[5] Heinz-Otto Peitgen, Hartmut Jürgens, and Dietmar Saupe. *Chaos and fractals: new frontiers of science*. Springer Science & Business Media, 2006.

[6] M.F. Barnsley. *Fractals Everywhere: New Edition*. Dover Books on Mathematics. Dover Publications, 2013.

[7] Chhabra, A., and Jensen, R. V.. Direct determination of the $f(\alpha)$ singularity spectrum. *Phys. Rev. Lett.*, 62:1327, 1989.

[8] JL del Río-Correa and J López-García. Shannon entropy and hausdorff dimension in multifractals. *Revista Mexicana de Física*, 58(1):13–20, 2012.

[9] Ch Beck and F Schlögl. Thermodynamics of chaotic systems (cambridge nonlinear science series 4), 1993.

[10] Longfeng Zhao, Wei Li, Chunbin Yang, Jihui Han, Zhu Su, and Yijiang Zou. Multifractality and network analysis of phase transition. *PloS one*, 12(1):e0170467, 2017.

[11] Darko Stošić, Dusan Stošić, Tatijana Stošić, and H Eugene Stanley. Multifractal analysis of managed and independent float exchange rates. *Physica A: Statistical Mechanics and its Applications*, 428:13–18, 2015.

[12] YU Shimizu, Stefan Thurner, and Klaus Ehrenberger. Multifractal spectra as a measure of complexity in human posture. *Fractals*, 10(01):103–116, 2002.