

DNA Sequences with Forbidden Words and the Generalized Cantor Set

Zhuowei Yang¹, Ping Wang²

¹College of Mathematics and Statistics, Yunnan University, Kunming, China

²Department of Mathematics, Penn State University, Schuylkill Haven, PA, USA

Email: yangzhuowei@outlook.com, pxw10@psu.edu

How to cite this paper: Yang, Z.W. and Wang, P. (2019) DNA Sequences with Forbidden Words and the Generalized Cantor Set. *Journal of Applied Mathematics and Physics*, 7, 1687-1696.

<https://doi.org/10.4236/jamp.2019.78115>

Received: June 28, 2019

Accepted: August 9, 2019

Published: August 12, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this work, we establish relations between DNA sequences with missing subsequences (the forbidden words) and the generalized Cantor sets. Various examples associated with some generalized Cantor sets, including Hao's frame representation and the generalized Sierpinski Set, along with their fractal graphs, are also presented in this work.

Keywords

DNA Sequences, Generalized Cantor Set

1. Introduction

Researchers have been interested in the relationships between fractals and DNA structures for years. Just recently, Anitas and Slyamov [1] studied multiscale fractal representing DNA sequences using small-angle scattering analysis. Cattani and Pierro [2] conducted a multifractal analysis of binary images of DNA in order to define a methodological approach to the classification of DNA sequences. Badea and her collaborators [3] characterized the geometry of some medical images of tissues in terms of complexity parameters such as the fractal dimension (FD). Carlo Cattani presented analysis of DNA based on the indicator matrix together with some elementary approach to a fractal estimate of DNA sequences in the book [4] edited by Elloumi and Zomaya. Albrecht-Buehler [5] identified explicitly the GA-sequences as a class of fractal genomic sequences. Ainsworth [6] investigated how the cell's nucleus holds molecules that manage human's DNA in the right location. In a book edited by Crilly, Earnshaw and Jones, Voss applied standard spectral density measurement techniques to demonstrate the ubiquity of low frequency noise and long range fractal correlations.

The study of the genome or DNA sequences through fractal analysis is very interesting. DNA sequences can be seen as sequences over the alphabet $\Sigma = \{a, c, g, t\}$. Subsequences that do not appear in DNA are considered as forbidden words. A visualization method of the forbidden words in [7] [8] [9] [10] [11] has been designed by B.-L. Hao since 2000. This method is now called Hao's frame representation. Recently, C.-X. Huang and S.-L. Peng discussed this method in detail, and many beautiful graphics were provided in [12] [13]. From these geometric intuitions, it can be observed that these forbidden words demonstrate certain fractal properties. In fact in this work we generated some amazing fractal graphs associated with DNA sequences with forbidden words as shown in **Figure 1**.

It is important to explore the fractal generating mechanism that is associated with the forbidden words in the sequence. H. J. Jeffrey [14] [15] and P. Tiño [16] [17] tried to associate the forbidden words with the IFS (Iterated Functions Systems) using chaos game algorithm. Denote Σ^* as the set of all finite sequences over Σ . Then how to find a generating formula or the mapping $\sigma: \Sigma \rightarrow w \in \Sigma^*$, where w is a sequence that does not contain forbidden subsequences, or corresponding iteration method? As was pointed out by P. Tiño, the IFS is a multi-fractal and therefore the generating formula would be relatively complicated.

In order to detect the structures of some symbolic sequences, one has to find the properties of their topology and metric and be able to visualize these sequences. To do this, we have to provide a type of graphical representation together with their topology and metric properties so that we can directly reveal their corresponding fractal graphs. This kind of representation method is important and necessary.

For an alphabet with cardinal 3, the well known CGR method (that is, Chaos Game Representation method) was first introduced by M.F. Barnsley by considering the points in an equilateral triangle. The substrings of a string were shown graphically (see [18]). For an alphabet $\Sigma = \{a, c, g, t\}$ with cardinality 4, the CGR method was later generalized by H.J. Jeffrey so that the DNA sequences can be visualized (see [14] [15]). The authors have transformed the DNA sequences into pseudo random walk in a 2-dimensional plane or in a 3-dimensional space [19] [20] [21]. We notice here that an iterated function system can be applied to construct a graphical representation of some DNA sequences [16] [17]. The points in the unit square $[0,1] \times [0,1]$ can be used to denote the substrings of the DNA sequences. Consequently, the four vertices of the unit square are labelled as a, c, g, t .

In application, the frame representation method proposed by Hao *et al.* is more intuitive and visual [9] [10]. The unit square $[0,1] \times [0,1]$ is divided equally with vertical and horizontal lines so that there are 4^k congruent small squares with side length 2^{-k} and area 4^{-k} . For the alphabet $\Sigma = \{a, c, g, t\}$ with cardinality 4, each small square of side length 2^{-k} is used to denote the string in Σ^k ($k = 1, 2, 3, \dots$) regularly (See 1-, 2- and 3-frame graphs in **Figures 2(a)-(c)**).

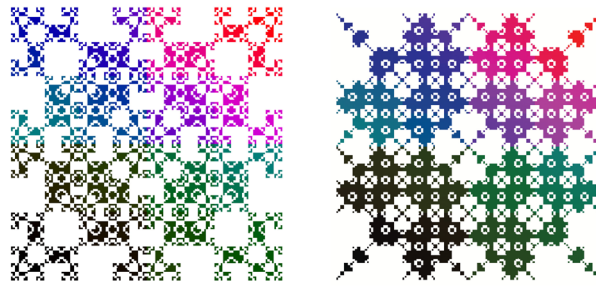


Figure 1. Graphs of some forbidden words.

g	t
a	c

(a)

gg	gt	tg	tt
ga	gc	ta	tc
ag	at	cg	ct
aa	ac	ca	cc

(b)

ggg	ggt	gtg	gtt	tgg	tgt	ttg	ttt
gga	ggc	gta	gtc	tga	tgc	tta	ttc
gag	gat	gcg	gct	tag	tat	tcg	tct
gaa	gac	gca	gcc	taa	tac	tca	tcc
agg	agt	atg	att	cgg	cgt	ctg	ctt
aga	agc	ata	atc	cga	cgc	cta	ctc
aag	aat	acg	act	cag	cat	cog	cct
aaa	aac	aca	acc	caa	cac	cca	ccc

(c)

Figure 2. The frame representation method of B.L. Hao *et al.* (a) 1-frame graph; (b) 2-frame graph; (c) 3-frame graph.

With the frame representation method of B.L. Hao, the repetition topology structure of the subsequences (*i.e.* the strings in Σ^k) of a DNA sequence can be easily visualized and efficiently drawn. The avoided or the under-represented short strings in the genome sequence form the forbidden words. These forbidden words are the reasons or the basis of the constructed fractals.

P. Tino [16] [17] proved the equivalence of the CGR method and the frame representation method of B.L. Hao *et al.* He noted that the cardinality of an alphabet can be generalized to a square integer ($|\Sigma| = b^2$ simultaneously for some integer b). We will in this paper extend the above methods and relax the restriction to the cardinality of an alphabet.

The order of this paper is as follows. In Section 2, we will first convert the problem into the discussion on certain type of generalized Cantor set, which can naturally correspond to multifractals, and then in Section 3, we will induce Hao's frame representation according to the principle that the correspondence between line segment and unit square is one-to-one [22]. Several examples, along with their fractal graphs, of some generalized Cantor sets are given at the end of this paper.

2. Forbidden Words and the Generalized Cantor Set

Rewrite the alphabet as $\Sigma = \{0, 1, 2, 3\}$. We first give the following definition.

Definition 2.1 Let $\Sigma = \{0, 1, 2, 3\}$. Denote B as the set consist of l finite sequences with length k (≥ 1):

$$B = \{t_{11} \cdots t_{1k}, t_{21} \cdots t_{2k}, \cdots, t_{l1} \cdots t_{lk}\}, \quad t_{ij} \in \Sigma, i = 1, \cdots, l, j = 1, \cdots, k. \quad (1)$$

Then call the infinite sequences over Σ

$$s = a_1 a_2 \cdots a_n \cdots, a_n \in \Sigma, \text{ and } a_1 a_{i+1} \cdots a_{i+k-1} \notin B, i = 1, 2, \cdots \quad (2)$$

the DNA sequence with no forbidden words B , a.k.a. allowed sequence.

It is known that when $x \in [0, 1]$ is expanded in ternary representation, the subset in $[0, 1]$

$$C = \{x = 0_3 x_1 x_2 \cdots x_n \cdots, x \in \{0, 2\}^{\mathbb{N}}\}$$

is called the Cantor set. Similarly, with quaternary expansion, we give the following definition.

Definition 2.2 When $x \in [0, 1]$ is represented in quaternary expansion

$$x = \sum_{n=1}^{\infty} \frac{a_n}{4^n} = 0_4 a_1 a_2 \cdots a_n \cdots, a_n \in \Sigma, \quad (3)$$

we call

$$C_G = \{x = 0_4 a_1 a_2 \cdots a_n \cdots, a_n \in \Sigma, a_i a_{i+1} \cdots a_{i+k-1} \notin B, i = 1, 2, \cdots\}, k \geq 1 \quad (4)$$

the generalized Cantor set.

Apparently, the discussions on DNA sequences (1) (2) that contain no forbidden words B can be converted into the discussion on the generalized Cantor set C_G .

Let $b_k^i = 4^{k-1} t_{i1} + \cdots + 4 t_{ik-1} + t_{ik}$, $b_k^i \in \{0, 1, \cdots, 4^k - 1\}$, $i = 1, 2, \cdots, l$, and

$$B' = \{b_k^1, b_k^2, \cdots, b_k^l\}, \quad (5)$$

Then, the condition $a_i a_{i+1} \cdots a_{i+k-1} \notin B, i = 1, 2, \cdots$ in Definition 2.2 can be rewritten as

$$4^{k-1} a_i + 4^{k-2} a_{i+1} + \cdots + a_{i+k} \notin B', i = 1, 2, \cdots$$

Theorem 2.1 The generalized Cantor set C_G can be inducted by using an iteration method.

Proof. In fact, for the $(k-1)$ th step of the quaternary expansion of $x \in [0, 1]$, there is

$$x = \frac{a_1}{4} + \cdots + \frac{a_{k-1}}{4^{k-1}} + \frac{x_{k-1}}{4^{k-1}}, 0 \leq x_{k-1} \leq 1, a_1, \cdots, a_k \in \Sigma. \quad (6)$$

Let

$$x_{k-1} = \frac{a_k}{4} + \frac{x_k}{4}, 0 \leq x_k \leq 1, a_k \in \Sigma \text{ and } 4^{k-1} a_1 + \cdots + a_k \notin B'. \quad (7)$$

Substitute (7) into (6),

$$x = \frac{a_1}{4} + \cdots + \frac{a_k}{4^k} + \frac{x_k}{4^k}, 0 \leq x_k \leq 1, 4^{k-1} a_1 + \cdots + a_k \notin B'. \quad (8)$$

In general, we let

$$x_{i-1+k} = \frac{a_{i+k}}{4} + \frac{x_{i+k}}{4}, 0 \leq x_{i+k} \leq 1, a_{i+k} \in \Sigma \text{ and } 4^{k-1} a_i + \cdots + a_{i+k} \notin B'. \quad (9)$$

and as $i \rightarrow \infty$, we obtain the generalized Cantor set C_G (2.2).

$$\left[\sum_{j=1}^k \frac{a_j}{4^j}, \sum_{j=1}^k \frac{a_j}{4^j} + \frac{1}{4^k} \right], a_j \in \Sigma \text{ are } 4^k \text{ intervals in } [0, 1] \text{ with length } \frac{1}{4^k}.$$

From the iteration Equation (7) in the theorem, the iteration acts differently on the l subintervals than on the $4^k - l$ intervals. Hence we have [11].

Corollary 2.3 *The generalized Cantor set C_G is multifractal.*

Proof. In the construction of the generalized Cantor sets C_G , measures on removed portions are redistributed to the neighboring sections repeatedly. Thus C_G is multifractal.

Obviously, the generalized Cantor sets are applicable for all p -carry representation (p is an integer).

3. The Hao's Frame Representation of the Generalized Cantor Set C_G

The theoretic foundation of the construction of DNA sequences can be seen in [12]. The subintervals in the quaternary expansion of $x \in [0, 1]$ can be one-to-one corresponding to the subsquares that are obtained by repeatedly equally dividing the unit square (and its subsquares) into 4 smaller subsquares. Cantor sets are created in one dimension in $[0, 1]$ while Sierpinski sets are constructed in two dimension within $[0, 1] \times [0, 1]$. Using the corresponding relationship between the unit interval and the unit square, we can convert the discussion on the generalized Cantor sets into the discussion on the generalized Sierpinski sets on the unit square.

Let $0 \leq \xi, \eta \leq 1$. The binary expansion of (ξ, η) is

$$(\xi, \eta) = \left(\sum_{n=1}^{\infty} \frac{c_n}{2^n}, \sum_{n=1}^{\infty} \frac{d_n}{2^n} \right), c_n, d_n \in \{0, 1\}. \quad (10)$$

The expansion can be related to the quaternary expansion of $x \in [0, 1]$ as follows:

$$a_i = c_i + 2d_i = \begin{cases} 0 & : (c_i, d_i) = (0, 0) \\ 1 & : (c_i, d_i) = (1, 0) \\ 2 & : (c_i, d_i) = (0, 1) \\ 3 & : (c_i, d_i) = (1, 1) \end{cases} \quad (11)$$

Thus the forbidden words b_k^i in B' can be represented as

$$b_k^i = 4^{k-1} (c_{i_1} + 2d_{i_1}) + \cdots + 4 (c_{i_{k-1}} + 2d_{i_{k-1}}) + (c_{i_k} + 2d_{i_k}), \quad (12)$$

$$c_{i_1}, d_{i_1}, \dots, c_{i_k}, d_{i_k} \in \{0, 1\}, i = 1, 2, \dots, l$$

Definition 3.1 Let $0 \leq \xi, \eta \leq 1$ and the binary expansion of (ξ, η) is (10). Then call

$$S_G = \left\{ (\xi, \eta) = \left(\sum_{n=1}^{\infty} \frac{c_n}{2^n}, \sum_{n=1}^{\infty} \frac{d_n}{2^n} \right), c_n, d_n \in \{0, 1\}, \right. \\ \left. 4^{k-1} (c_{i_1} + 2d_{i_1}) + \cdots + 4 (c_{i_{k-1}} + 2d_{i_{k-1}}) + (c_{i_k} + 2d_{i_k}) \notin B' \right\} \quad (13)$$

the generalized Sierpinski set that corresponds to the the generalized Cantor set $C_G(4)$.

Theorem 3.1 The generalized Sierpinski set S_G can be inducted by iterating method.

Proof. The $(k-1)$ th binary expansion of (ξ, η) is

$$(\xi, \eta) = \left(\sum_{n=1}^{k-1} \frac{c_n}{2^n} + \frac{\xi_{k-1}}{2^{k-1}}, \sum_{n=1}^{k-1} \frac{d_n}{2^n} + \frac{\eta_{k-1}}{2^{k-1}} \right), 0 \leq \xi_{k-1}, \eta_{k-1} \leq 1. \quad (14)$$

Let

$$(\xi_{k-1}, \eta_{k-1}) = \left(\frac{c_k}{2} + \frac{\xi_k}{2}, \frac{d_k}{2} + \frac{\eta_k}{2} \right), c_k, d_k \in \{0, 1\}, 0 \leq \xi_k, \eta_k \leq 1, \quad (15)$$

$$4^{k-1}(c_1 + 2d_1) + \cdots + 4(c_{k-1} + 2d_{k-1}) + (c_k + 2d_k) \notin B', c_k, d_k \in \{0, 1\}$$

Substitute (15) into (14), we have

$$(\xi, \eta) = \left(\sum_{n=1}^k \frac{c_n}{2^n} + \frac{\xi_k}{2^k}, \sum_{n=1}^k \frac{d_n}{2^n} + \frac{\eta_k}{2^k} \right),$$

$$4^{k-1}(c_1 + 2d_1) + \cdots + 4(c_{k-1} + 2d_{k-1}) + (c_k + 2d_k) \notin B'$$

Generally, let

$$(\xi_{k+i-1}, \eta_{k+i-1}) = \left(\frac{c_{k+i}}{2} + \frac{\xi_{k+i}}{2}, \frac{d_{k+i}}{2} + \frac{\eta_{k+i}}{2} \right), 0 \leq \xi_{k+i}, \eta_{k+i} \leq 1, c_{k+i}, d_{k+i} \in \{0, 1\},$$

$$4^{k-1}(c_1 + 2d_1) + \cdots + 4(c_{k+i-1} + 2d_{k+i-1}) + (c_{k+i} + 2d_{k+i}) \notin B'$$

Noticing the corresponding relationship between numbers and the subsquares, naturally we have Hao's frame representation. The second-order Hao's frame representation can be inducted from the corresponding relationship illustrated in **Figure 3**.

The next few examples illustrate analytic structure of some DNA sequences along with the fractal graphs of the relevant generalized Cantor sets.

Example 3.2 Let $\Sigma = \{0, 1, 2, 3\}$, $B = \{00, 11, 22\}$. Then $B' = \{0, 5, 10\}$. Hence the arithmetic expression of the generalized Cantor set is

$$x = 0_4 a_1 a_2 \cdots a_n \cdots, a_n \in \{0, 1, 2, 3\} \text{ and } 4a_i + a_{i+1} \neq 0, 5, 10, i = 1, 2, \cdots$$

And the symbolic sequence is

$$a_1 a_2 \cdots a_n \cdots, a_n \in \Sigma, a_i a_{i+1} \notin B, i = 1, 2, \cdots$$

which is shown graphically in **Figure 4**.

Example 3.3 Let $\Sigma = \{0, 1, 2, 3\}$, $B = \{10, 20, 30\}$. Then $B' = \{4, 8, 12\}$. Hence the arithmetic expression of the generalized Cantor set is

$$x = 0_4 a_1 a_2 \cdots a_n \cdots, a_n \in \{0, 1, 2, 3\} \text{ and } 4a_i + a_{i+1} \neq 4, 8, 12, i = 1, 2, \cdots$$

And the symbolic sequence is

$$a_1 a_2 \cdots a_n \cdots, a_n \in \Sigma, a_i a_{i+1} \notin B, i = 1, 2, \cdots$$

with graphs **Figure 5**:

Example 3.4 Let $\Sigma = \{0, 1, 2, 3\}$, $B = \{011, 022, 100, 133, 200, 233, 311, 322\}$. Then $B' = \{5, 10, 16, 31, 32, 47, 53, 58\}$. Hence the arithmetic expression of the generalized Cantor set is

$$x = 0_4 a_1 a_2 \cdots a_n \cdots, a_n \in \{0, 1, 2, 3\} \text{ and } 4^2 a_i + 4a_{i+1} + a_i \notin B', i = 1, 2, \cdots$$

And the symbolic sequence is

$$a_1 a_2 \cdots a_n \cdots, \quad a_n \in \Sigma, a_i a_{i+1} \notin B, i = 1, 2, \dots$$

which are shown below

Similarly, we could produce the following amazing fractal graphs shown in **Figure 6**, **Figure 7**, of different DNA sequences with various forbidden words.

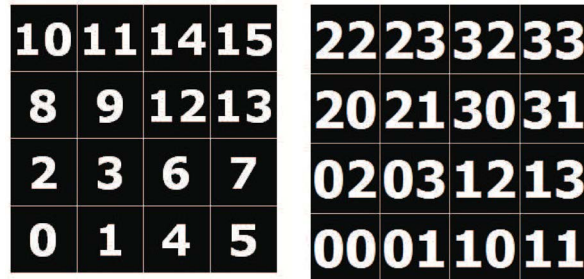


Figure 3. Hao's frame representation of $k = 2$.

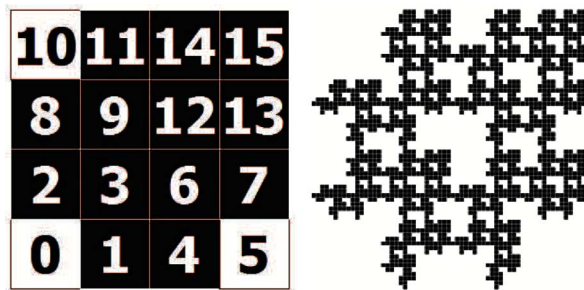


Figure 4. $B' = \{0, 5, 10\}$.

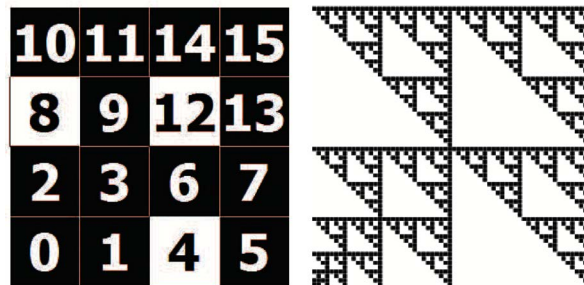


Figure 5. $B' = \{4, 8, 12\}$.

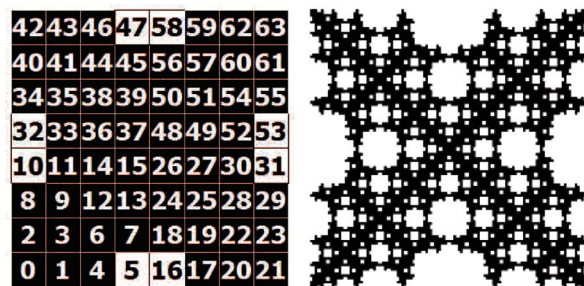


Figure 6. $B' = \{5, 10, 16, 31, 32, 47, 53, 58\}$.

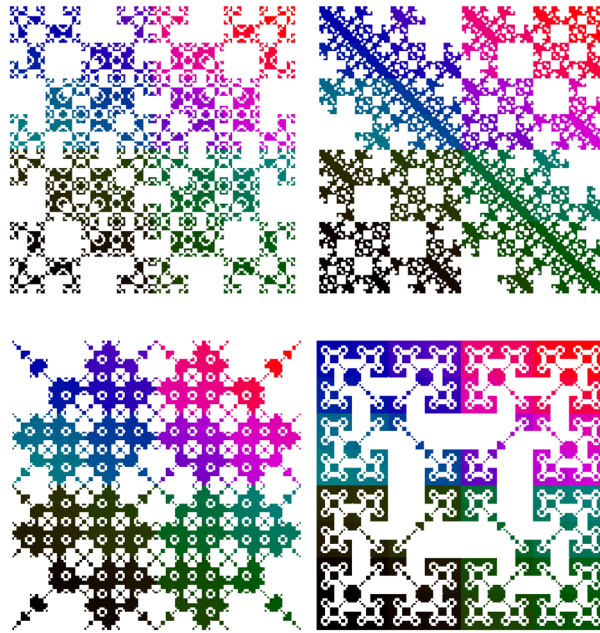


Figure 7. Other examples.

4. Conclusion

We established relations between the generalized Cantor sets and some DNA sequences with missing words. And we have associated Hao's frame representations and the generalized Sierpinski set with the generalized Cantor sets. The authors are interested in applying the analytical representation method to study the graphical results of space filling research works (cf. [23] [24] [25]).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Mircea Anitas, E. and Slyamov, A. (2017) Structural Characterization of Chaos Game Fractals Using Small-Angle Scattering Analysis. *PLoS ONE*, **12**, e0181385. <https://doi.org/10.1371/journal.pone.0181385>
- [2] Cattani, C. and Pierro, G. (2013) on the Fractal Geometry of DNA by the Binary Image Analysis. *Bulletin of Mathematical Biology*, **75**, 1544-1570. <https://doi.org/10.1007/s11538-013-9859-9>
- [3] Badea, A.F., *et al.* (2013) Fractal Analysis of Elastographic Images for Automatic Detection of Diffuse Diseases of Salivary Glands: Preliminary Results. *Computational and Mathematical Methods in Medicine*, **2013**, Article ID: 347238. <https://doi.org/10.1155/2013/347238>
- [4] Elloumi, M. and Zomaya, A.Y. (2014) Biological Knowledge Discovery Handbook. John Wiley & Sons, Hoboken, NJ. <https://doi.org/10.1002/9781118617151>
- [5] Albrecht-Buehler, G. (2012) Fractal Genome Sequences. *Gene*, **498**, 20-27. <https://doi.org/10.1016/j.gene.2012.01.090>

- [6] Ainsworth, C. (2009) Cells Go Fractal. *Nature-International Weekly Journal of Science*.
- [7] Falconer, K.J. (1998) Techniques in Fractal Geometry. John Wiley & Sons, Hoboken, NJ.
- [8] Hao, B.-L., Xie, H.-M. and Chen, G.-Y. (2000) Factorizable Language: From Dynamics to Bacterial Complete Genomes. *Physica A: Statistical Mechanics and Its Applications*, **288**, 10-20. [https://doi.org/10.1016/S0378-4371\(00\)00411-8](https://doi.org/10.1016/S0378-4371(00)00411-8)
- [9] Hao, B.-L., Xie, H.-M., Chen, G.-Y. and Chen, G.-Y. (2000) Avoided Strings in Bacterial Complete Genomes and a Related Combinatorial Problem. *Annals of Combinatorics*, **4**, 247-255. <https://doi.org/10.1007/PL00001279>
- [10] Yu, Z.-G., Hao, B.-L., Xie, H.-M. and Chen, G.-Y. (2000) Dimensions of Fractals Related to Languages Defined by Tagged Strings in Complete Genomes. *Chaos, Solitons & Fractals*, **11**, 2215-2222. [https://doi.org/10.1016/S0960-0779\(99\)00141-1](https://doi.org/10.1016/S0960-0779(99)00141-1)
- [11] Crilly, A.J., Earnshaw, R. and Jones, H. (2013) Applications of Fractals and Chaos: The Shape of Things. Springer-Verlag, Berlin, Heidelberg.
- [12] Huang, C.X. and Peng, S.L. (2006) A Note on Fractals of One Forbidden Word and Their Box Dimensions. *Fractals*, **14**, 327-337. <https://doi.org/10.1142/S0218348X06003325>
- [13] Huang, C.X. and Peng, S.L. (2008) Fractals of Forbidden Words and Approximating Their Box Dimensions. *Physica A: Statistical Mechanics and Its Applications*, **387**, 703-716. <https://doi.org/10.1016/j.physa.2007.09.009>
- [14] Jeffrey, H.J. (1990) Chaos Game Representation of Genetic Sequences. *Nucleic Acids Research*, **18**, 2163-2170. <https://doi.org/10.1093/nar/18.8.2163>
- [15] Jeffrey, H.J. (1992) Chaos Game Visualization of Sequences. *Computers & Graphics*, **16**, 25-33. [https://doi.org/10.1016/0097-8493\(92\)90067-6](https://doi.org/10.1016/0097-8493(92)90067-6)
- [16] Tiño, P. (1999) Spatial Representation of Symbolic Sequences Iterative Function Systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, **29**, 386-393. <https://doi.org/10.1109/3468.769757>
- [17] Tiño, P. (2002) Multifractal Properties of Hao's Geometric Representations of DNA Sequences. *Physica A: Statistical Mechanics and Its Applications*, **304**, 480-494. [https://doi.org/10.1016/S0378-4371\(01\)00574-X](https://doi.org/10.1016/S0378-4371(01)00574-X)
- [18] Barnsley, M.F. (1988) Fractals Everywhere. Academic Press, Cambridge, MA.
- [19] Berthelsen, C.L., Glazier, J.A. and Skolnik, M.H. (1992) Global Fractal Dimension of Human DNA Sequences Treated as Pseudorandom Walks. *Physical Review A*, **45**, 8902-8913. <https://doi.org/10.1103/PhysRevA.45.8902>
- [20] Stanley, H.E., Buldyrev, S.V., Goldberger, A.L., Goldberger, Z.D., Havlin, S., Mantegna, R.N., Ossadnik, S.M., Peng, C.K. and Simons, M. (1994) Statistical Mechanics in Biology: How Ubiquitous Are Long-Range Correlations? *Physica A: Statistical Mechanics and Its Applications*, **205**, 214-253. [https://doi.org/10.1016/0378-4371\(94\)90502-9](https://doi.org/10.1016/0378-4371(94)90502-9)
- [21] Leong, P.M. and Morgenthaler, S. (1995) Random Walk and Gap Plots of DNA Sequences. *Bioinformatics*, **11**, 503-507. <https://doi.org/10.1093/bioinformatics/11.5.503>
- [22] Solov'yev, V.V., Korolev, S.V. and Lim, H.A. (1993) A New Approach for the Classification of Functional Regions of DNA Sequences Based of Fractal Representation. *International Journal of Genomic Research*, **1**, 109-128.
- [23] Bagga, S., Girdhar, A., Trivedi, M.C. and Yang, Y.Z. (2016) RMI Approach to Cluster Based Cache Oblivious Peano Curves. 2016 *Second International Conference on*

Computational Intelligence & Communication Technology, Ghaziabad, India, 12-13 February 2016, 89-95. <https://doi.org/10.1109/CICT.2016.26>

- [24] Makarov, B.M. and Podkorytov, A.N. (2017) On the Coordinate Functions of Peano Curves. *St. Petersburg Mathematical Journal*, **28**, 115-125.
<https://doi.org/10.1090/spmj/1441>
- [25] Platos, J., Kromer, P. and Snasel, V. (2015) Efficient Area Association Using Space Filling Curves. 2015 *International Conference on Intelligent Networking and Collaborative Systems*, Taipei, 2-4 September 2015, 322-326.
<https://doi.org/10.1109/INCoS.2015.39>