Global Journal of Computer Science and Technology: G
Interdisciplinary
Volume 20 Issue 1 Version 1.0 Year 2020
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Modeling and Simulation of Genome Evolution Using Linear Boolean Functions Associated with One Dimensional Cellular Automata

By Prashanthi Govindarajan, Sathya Govindarajan
& Ethirajan Govindarajan

Abstract- Structural and functional behavior of genomes could be studied using one dimensional binary-valued three neighborhood cellular automata updating rules. These updating rules are linear Boolean functions, and they are applied to the adjoint sequences of adenine, (A), Thymine (T), Guanine (G) and Cytosine (C) corresponding to the characteristic sequence of a genome. This paper proposes the use of linear Boolean functions, and demonstrates the textural or fractal behavior of genome evolution in terms of nucleotide adjoints.

Keywords: linear boolean functions, cellular automata, genome evolution.

GJCST-G Classification: J.3

Modeling and Simulation of Genome Evolution Using Linear Boolean Functions Associated with One Dimensional Cellular Automata

Prashanthi Govindarajan [α], Sathya Govindarajan [σ] & Ethirajan Govindarajan [ρ]

Abstract- **Structural and functional behavior of genomes could be studied using one dimensional binary-valued three neighborhood cellular automata updating rules. These updating rules are linear Boolean functions, and they are applied to the adjoint sequences of adenine, (A), Thymine (T), Guanine (G) and Cytosine (C) corresponding to the characteristic sequence of a genome. This paper proposes the use of linear Boolean functions, and demonstrates the textural or fractal behavior of genome evolution in terms of nucleotide adjoints.**

Keywords: linear boolean functions, cellular automata, genome evolution.

I. INTRODUCTION

The four nucleotides A, T, G, and C get connected by phosphodiester bonds to form strands. Strand formation depends on innumerable factors related to inter and intra cellular parameters and functions. One cannot precisely say that a particular strand gets formed using such and such rules. The infinite possibilities of strand formation cannot be determined experimentally or in the framework of classical genetics. One can alternatively formulate a notion of "Language of Genomes" wherein one can finitely specify infinite strands, Fig. 1 shows a finitely generated quaternary tree structure of strand formation of nucleic acids.

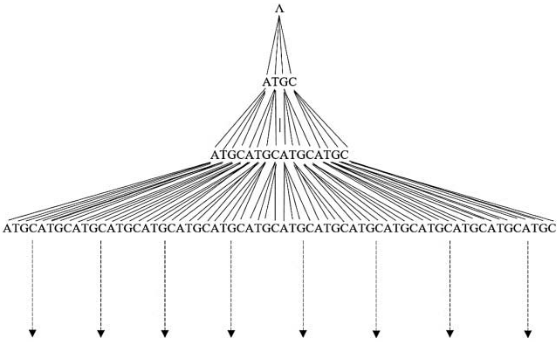


Fig. 1: Quaternary tree structure for strand formation

To be precise, Fig. 1 shows three levels of nucleotides. One can generate 64 strands of length 3. As the length increases, the number of strands increases as per the formula 4n, where n is the length of the strand. Strands of length three are called triplet codons or 3-tuple codons. Similarly, one can think of n-tuple codons where n is any number.

Parallel Prediction of Nucleotides using One Dimensional Cellular Automata

A genome sequence is a chain of four nucleotides A, T, G and C. The numerical representation of a genome sequence is a sequence of four numbers 1, 2, 3 and 4. Linear prediction of a strand could be carried out using linear prediction algorithms from a sub sequence of length 8. Alternatively, one can evolve generations of genome sequences from a given full-length genome sequence using one-dimensional cellular automata rules. Section 2 describes the notions of adjoints of nucleotides corresponding to a genome sequence. Section 3 describes the notions of cellular automata and linear Boolean functions. Section 4 provides the results of applying linear Boolean functions on adjoint strings of nucleotides. Section 5 demonstrates the results of combining evolution patterns of adjoint sequences dyadically. Section 6 presents various observations made from the study and proposes future perspectives of cellular automata-based genome analytics.

II. ADJOINTS OF NUCLEOTIDES

Adjoint of a particular nucleotide in a genome sequence is the binary sequence obtained by substituting the particular nucleotides in the genome sequence by 1's and the others by 0's. For example, let us consider a sample sequence G, A, A, T, G, A, T, T, A, C, C, A, A, G, G, C of length 16. Now the adjoint of adenine (A) is the binary string A(n) = 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0. The adjoint of thymine (T) is the binary string T(n) = 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0. The adjoint of guanine (G) is the binary string G(n) = 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0. The adjoint of cytocine (C) is binary string C(n) = 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1. The first segment of 40 nucleotides of a genome sequence of Brucella Suis 1330 is considered here for a case study. The actual length of the genome sequence of Brucella Suis 1330 is 5806. The sample sequence is given below.

GAATGATTACCAAGGCCAAGCTCAAGCTCTCCTTCCTTGG

Author α: Chief Executive Officer, PRC Global Technologies Inc. Ontario, Canada. e-mail: sathyag413@gmail.com

Author σ ρ: Director, PRC Global Technologies Inc. Ontario, Canada. e-mail: prashanthi89@gmail.com

Now the adjoints of this sample sequence of length 40 are given below.

A(n) = 0110010010011000011000011000000000000000
T(n) = 0001001100000000000000100000101001100110
G(n) = 1000100000000110000100000100000000000011
C(n) = 0000000001100001100010100010101100110000

A cellular automaton is an idealized parallel processing system consisting of an array of numbers (1-D, 2-D and more) realized using updating rules based on certain neighborhood. For example, a one-dimensional cellular automaton would consist of a finite length array as shown below.

III. CELLULAR AUTOMATA AND LINEAR BOOLEAN FUNCTIONS

A cellular automaton is an idealized parallel processing system consisting of an array of numbers (1-D, 2-D and more) realized using updating rules based on certain neighborhood. For example, a one dimensional cellular automaton would consist of a finite length array as shown below.

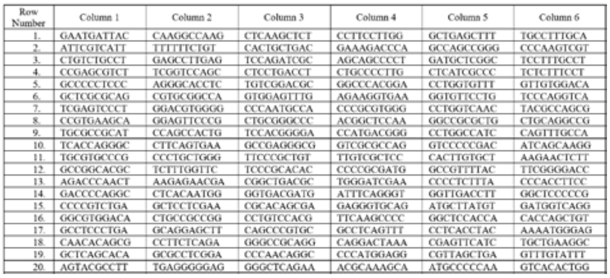
---	---	---	i-1	i	i+1	---	---	---

Consider an ith cell in the array. This cell has a neighbor i-1 on its left and another i+1 on its right. All three put together is called a three-neighborhood. One can assign a site (cell) variable ξi-1, ξi, and ξi+1 to the three-neighborhood cells. At a particular instant of time, these variables take on numerical values, say either a 0 or a 1. In such a case, the variables are denoted as ξti-1, ξti, and ξti+1. The value of the ith cell at the next instant of time is evaluated using an updating rule that involves the present values of the ith, (i-1)th and (i+1)th cells. This updating rule is essentially a linear Boolean function of three variables. One can construct 256 linear Boolean functions as updating rules of one-dimensional three-neighborhood binary-valued cellular automata. Each rule defines an automaton by itself. So, one-dimensional binary-valued three-neighborhood cellular automata (123CA) rules could be used to model adjoints of a genome sequence. The first thirty linear Boolean functions of cellular automata 123CA are listed below with their decimal equivalents.

Linear Boolean Function	Decimal Equivalent
0	0
$(\bar{\xi}_{i-1}\bar{\xi}_i\bar{\xi}_{i+1})$	1
$(\bar{\xi}_{i-1}\bar{\xi}_i\xi_{i+1})$	2
$(\bar{\xi}_{i-1}\bar{\xi}_i)$	3
$(\bar{\xi}_{i-1}\xi_i\bar{\xi}_{i+1})$	4
$(\bar{\xi}_{i-1}\bar{\xi}_{i+1})$	5
$(\bar{\xi}_{i-1}\xi_i\bar{\xi}_{i+1})+(\bar{\xi}_{i-1}\bar{\xi}_i\xi_{i+1})$	6
$(\bar{\xi}_{i-1}\bar{\xi}_{i+1})+(\bar{\xi}_{i-1}\bar{\xi}_i)$	7
$(\bar{\xi}_{i-1}\xi_i\xi_{i+1})$	8
$(\bar{\xi}_{i-1}\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\xi_i\xi_{i+1})$	9
$(\bar{\xi}_{i-1}\xi_{i+1})$	10
$(\bar{\xi}_{i-1}\bar{\xi}_i) + (\bar{\xi}_{i-1}\xi_{i+1})$	11
$(\bar{\xi}_{i-1}\xi_i)$	12
$(\bar{\xi}_{i-1}\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\xi_i)$	13
$(\bar{\xi}_{i-1}\xi_i) + (\bar{\xi}_{i-1}\xi_{i+1})$	14
$(\bar{\xi}_{i-1})$	15
$(\xi_{i-1}\bar{\xi}_i\bar{\xi}_{i+1})$	16
$(\bar{\xi}_i\bar{\xi}_{i+1})$	17
$(\xi_{i-1}\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\bar{\xi}_i\xi_{i+1})$	18
$(\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\bar{\xi}_i)$	19
$(\xi_{i-1}\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\xi_i\bar{\xi}_{i+1})$	20
$(\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\bar{\xi}_{i+1})$	21
$(\xi_{i-1}\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\xi_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\bar{\xi}_i\xi_{i+1})$	22
$(\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\bar{\xi}_i)$	23
$(\xi_{i-1}\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\xi_i\xi_{i+1})$	24
$(\bar{\xi}_{i-1}\xi_i\xi_{i+1}) + (\bar{\xi}_i\bar{\xi}_{i+1})$	25
$(\xi_{i-1}\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\xi_{i+1})$	26
$(\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\xi_{i+1})$	27
$(\xi_{i-1}\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\xi_i)$	28
$(\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\xi_i)$	29
$(\xi_{i-1}\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\xi_i) + (\bar{\xi}_{i-1}\xi_{i+1})$	30

IV. CELLULAR AUTOMATA EVOLUTIONS OF GENOME ADJOINTS

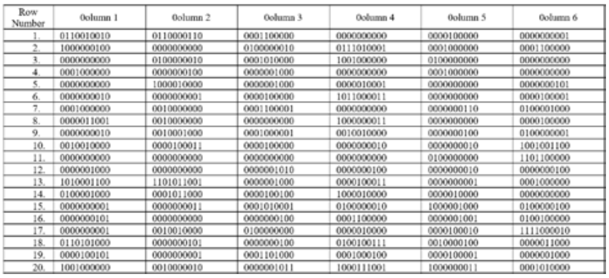
The genome sequence of Brucella Suis 1330 is considered here for a case study. Due to space limitations, a part of the genome sequence and its adjoints are shown below. As defined already, adjoint of genome sequence concerning a particular nucleotide is the binary string obtained by marking a '1' in the place of that particular nucleotide and by marking a '0' in the places of other nucleotides. A segment consisting of 60 nucleotides of Brucella Suis 1330 is shown below.



Row Number	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6
1.	GAATGATTAC	CAAGGCCAAG	CTCAAGCTCT	CCTTCCTTGG	GCTGAGCTTT	TGCCTTTGCA
2.	ATTCGTCATT	TTTTTTCTGT	CACTGCTGAC	GAAAGACCCA	GCCAGCCGGG	CCCAAGTCGT
3.	CTGTCTGCCT	GAGCCTTGAG	TCCAGATCGC	AGCAGCCCCT	GATGCTCGGC	TCCTTTGCCT
4.	CCGAGCGTCT	TCGGTCCAGC	CTCCTGACCT	CTGCCCCTTG	CTCATCGCCC	TCTCTTTCCT
5.	GCCCCCTCCC	AGGGCACCTC	TGTCGGACGC	GGCCCACGGA	CCTGGTGTTT	GTTGTGGACA
6.	GCTCGCGCAG	CGTGCGGCCA	GTGGAGTTTG	AGAAGGTGAA	GGTGTTCCTG	TCCCAGGTCA
7.	TCGAGTCCCT	GGACGTGGGG	CCCAATGCCA	CCCGCGTGGG	CCTGGTCAAC	TACGCCAGCG
8.	CCGTGAAGCA	GGAGTTCCCG	CTGCGGGCCC	ACGGCTCCAA	GGCCGCGCTG	CTGCAGGCCG
9.	TGCGCCGCAT	CCAGCCACTG	TCCACGGGGA	CCATGACGGG	CCTGGCCATC	CAGTTTGCCA
10.	TCACCAGGGC	CTTCAGTGAA	GCCGAGGGCG	GTCGCGCCAG	GTCCCCCGAC	ATCAGCAAGG
11.	TGCGTGCCCG	CCCTGCTGGG	TTCCCGCTGT	TTGTCGCTCC	CACTTGTGCT	AAGAACTCTT
12.	GCCGGCACGC	TCTTTGGTTC	TCCCGCACAC	CCCCGCGATG	GCCGTTTTAC	TTCGGGGACC
13.	AGACCCAACT	AAGAGAACGA	CGGCTGACGC	TGGGATCGAA	CCCCTCTTTA	CCCACCTTCC
14.	GACCCCAGGC	CTCACAATGG	GGTGACGATG	ATTTCAGGGT	GGTTGACCTT	GGCTCCCCCG
15.	CCCCGTCTGA	GCTCCTCGAA	CGCACAGCGA	GAGGGTGCAG	ATGCTTATGT	GATGGTCAGG
16.	GGCGTGGACA	CTGCCGCCGG	CCTGTCCACG	TTCAAGCCCC	GGCTCCACCA	CACCAGCTGT
17.	GCCTCCCTGA	GCAGGAGCTT	CAGCCCGTGC	GCCTCAGTTT	CCTCACCTAC	AAAATGGGAG
18.	CAACACAGCG	CCTTCTCAGA	GGGCCGCAGG	CAGGACTAAA	CGAGTTCATC	TGCTGAAGGC
19.	GCTCAGCACA	GCGCCTCGGA	CCCAACAGGC	CCCATGGAGG	CGTTAGCTGA	GTTTGTATTT
20.	AGTACGCCTT	TGAGGGGGAG	GGGCTCAGAA	ACGCAAAGCA	ATGCCCCCAA	GTCACACTGG

The adjoints of the genome sequence segment are given below.

Adjoint A(n)



Row Number	0olumn 1	0olumn 2	0olumn 3	0olumn 4	0olumn 5	0olumn 6
1.	0110010010	0110000110	0001100000	0000000000	0000100000	0000000001
2.	1000000100	0000000000	0100000010	0111010001	0001000000	0001100000
3.	0000000000	0100000010	0001010000	1001000000	0100000000	0000000000
4.	0001000000	0000000100	0000001000	0000000000	0001000000	0000000000
5.	0000000000	1000010000	0000001000	0000010001	0000000000	0000000101
6.	0000000010	0000000001	0000100000	1011000011	0000000000	0000100001
7.	0001000000	0010000000	0001100001	0000000000	0000000110	0100001000
8.	0000011001	0010000000	0000000000	1000000011	0000000000	0000100000
9.	0000000010	0010001000	0001000001	0010010000	0000000100	0100000001
10.	0010010000	0000100011	0000100000	0000000010	0000000010	1001001100
11.	0000000000	0000000000	0000000000	0000000000	0100000000	1101100000
12.	0000001000	0000000000	0000001010	0000000100	0000000010	0000000100
13.	1010001100	1101011001	0000001000	0000100011	0000000001	0001000000
14.	0100001000	0001011000	0000100100	1000010000	0000010000	0000000000
15.	0000000001	0000000011	0001010001	0100000010	1000001000	0100000100
16.	0000000101	0000000000	0000000100	0001100000	0000001001	0100100000
17.	0000000001	0010010000	0100000000	0000010000	0000100010	1111000010
18.	0110101000	0000000101	0000000100	0100100111	0010000100	0000011000
19.	0000100101	0000000001	0001101000	0001000100	0000100001	0000001000
20.	1001000000	0010000010	0000001011	1000111001	1000000011	0001010000

Adjoint T(n)

Row Number	0olumn 1	0olumn 2	0olumn 3	0olumn 4	0olumn 5	0olumn 6
1.	0001001100	0000000000	0100000101	0011001100	0010000111	1000111000
2.	0110010011	1111110101	0001001000	0000000000	0000000000	0000001001
3.	0101010001	0000011000	1000001000	0000000001	0010010000	1001110001
4.	0000000101	1000100000	0100100001	0100000110	0100100000	1010111001
5.	0000001000	0000000010	1010000000	0000000000	0010010111	0110100000
6.	0010000000	0010000000	0100001110	0000001000	0010110010	1000000100
7.	1000010001	0000010000	0000010000	0000001000	0010010000	1000000000
8.	0001000000	0000110000	0100000000	0000010000	0000000010	0100000000
9.	1000000001	0000000010	1000000000	0001000000	0010000010	0001110000
10.	1000000000	0110001000	0000000000	0100000000	0100000000	0100000000
11.	1000100000	0001001000	1100000101	1101000100	0001101001	0000001011
12.	0000000000	1011100110	1000000000	0000000010	0000111100	1100000000
13.	0000000001	0000000000	0000100000	1000010000	0000101110	0000001100
14.	0000000000	0100000100	0010000010	0111000001	0011000011	0001000000
15.	0000010100	0010010000	0000000000	0000010000	0100110101	0010010000
16.	0000100000	0100000000	0010100000	1100000000	0001000000	0000000101
17.	0001000100	0000000011	0000000100	0001000111	0010000100	0000100000
18.	0000000000	0011010000	0000000000	0000001000	0000110010	1001000000
19.	0010000000	0000010000	0000000000	0000100000	0011000100	0111010111
20.	0010000011	1000000000	0000100000	0000000000	0100000000	0100000100

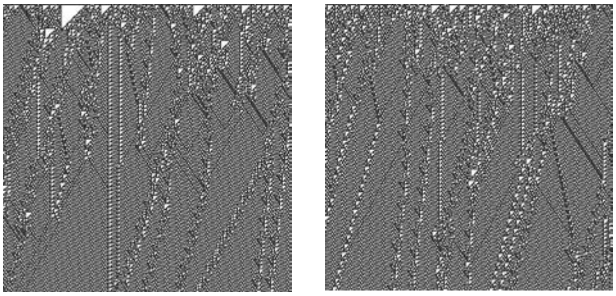
Adjoint G(n)

Row Number	0olumn 1	0olumn 2	0olumn 3	0olumn 4	0olumn 5	0olumn 6
1.	1000100000	0001100001	0000010000	0000000011	1001010000	0100000100
2.	0000100000	0000000010	0000100100	1000100000	1000100111	0000010010
3.	0010001000	1010000101	0000100010	0100100000	1001000110	0000001000
4.	0010101000	0011000010	0000010000	0010000001	0000001000	0000000000
5.	1000000000	0111000000	0100110010	1100000110	0001101000	1001011000
6.	1000101001	0101011000	1011010001	0100110100	1101000001	0000011000
7.	0010100000	1100101111	0000001000	0001010111	0001100000	0001000101
8.	0010100100	1101000001	0010111000	0011000000	1100101001	0010011001
9.	0101001000	0001000001	0000011110	0000100111	0001100000	0010001000
10.	0000001110	0000010100	1001011101	1001010001	1000000100	0000100011
11.	0101010001	0000100111	0000010010	0010010000	0000010100	0010000000
12.	1001100010	0000011000	0000100000	0000101001	1001000000	0001111000
13.	0100000000	0010100010	0110010010	0111000100	0000000000	0000000000
14.	1000000110	0000000011	1101001001	0000001110	1100100000	1100000001
15.	0000100010	1000000100	0100001010	1011101001	0010000010	1001100011
16.	1101011000	0010010011	0001000001	0000010000	1100000000	0000010010
17.	1000000010	1001101000	0010001010	1000001000	0000000000	0000011101
18.	0000000101	0000000010	1110010011	0011000000	0101000000	0100100110
19.	1000010000	1010000110	0000000110	0000011011	0100010010	1000100000
20.	0100010000	0101111101	1110000100	0010000100	0010000000	1000000011

Adjoint C(n)

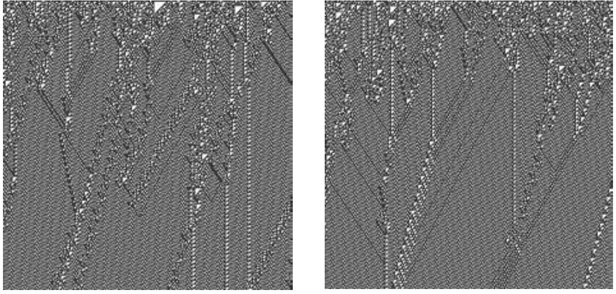
Row Number	1olumn 1	1olumn 2	1olumn 3	1olumn 4	1olumn 5	1olumn 6
1.	0000000001	1000011000	1010001010	1100110000	0100001000	0011000010
2.	0001001000	0000001000	1010010001	0000001110	0110011000	1110000100
3.	1000100110	0001100000	0110000101	0010011110	0000101001	0110000110
4.	1100010010	0100011001	1011000110	1001111000	1010010111	0101000110
5.	0111110111	0000101101	0001000101	0011101000	1100000000	0000000010
6.	0101010100	1000100110	0000000000	0000000000	0000001100	0111000010
7.	0100001110	0001000000	1110000110	1110100000	1100001001	0010110010
8.	1100000010	0000001110	1001000111	0100101100	0011010100	1001000110
9.	0010110100	1100110100	0110100000	1100001000	1100011001	1000000110
10.	0101100001	1001000000	0110000010	0010101100	0011111001	0010010000
11.	0010001110	1110010000	0011101000	0000101011	1010000010	0000010100
12.	0110010101	0100000001	0111010101	1111010000	0110000001	0010000011
13.	0001110010	0000000100	1001000101	0000001000	1111010000	1110110011
14.	0011110001	1010100000	0000010000	0000100000	0000001100	0010111110
15.	1111001000	0101101000	1010100100	0000000100	0001000000	0000001000
16.	0010000010	1001101100	1100011010	0010001111	0010110110	1011001000
17.	0110111000	0100000100	1001110001	0110100000	1101011001	0000000000
18.	1001010010	1100101000	0001101000	1000010000	1000001001	0010000001
19.	0101001010	0101101000	1110010001	1110000000	1000001000	0000000000
20.	0000101100	0000000000	0001010000	0101000010	0001111100	0010101000

Cellular automata evolutions of adjoints of a genome are carried out using 256 rules of 123CA. As an example, rule number 137 of 123CA, that is, $(\bar{\xi}_{i-1}\bar{\xi}_i\bar{\xi}_{i+1}) + (\xi_i\xi_{i+1})$ is applied to adjoints of Brucella Suis 1330 genome and results shown below in Fig. 2.



Evolution of A(n)

Evolution of T(n)



Evolution of G(n)

Evolution of C(n)

Fig. 2: Evolution of adjoints using rule 137 of 123CA

The size of the images shown in Fig. 2 is 500x500, though the actual size is 5806x500. The first 500 columns of the actual images are clipped and presented here for visual clarity. From Fig. 2, it is clear that the evolution pattern of each adjoint is different. One can observe that there are certain fractal patterns in the evolutions and such fractals are distributed in the images very differently. For instance, the zoomed in versions of the evolution patterns of A(n), T(n), G(n) and C(n) using rule 137 are shown in Figs. 3, 4, 5 and 6 respectively.

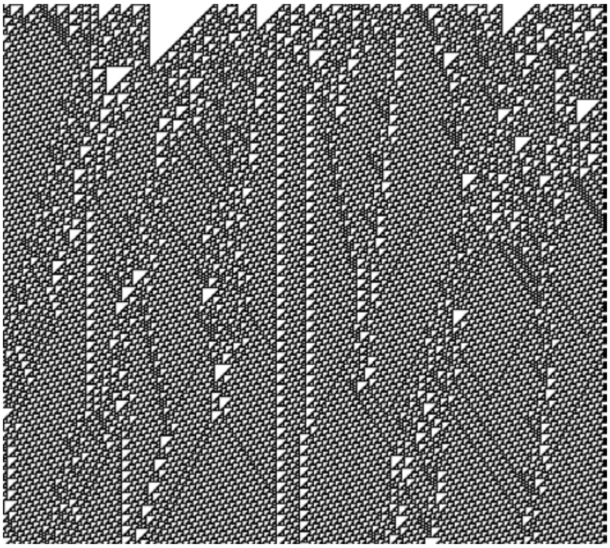


Fig. 3: Zoomed in version of evolution pattern of A(n)

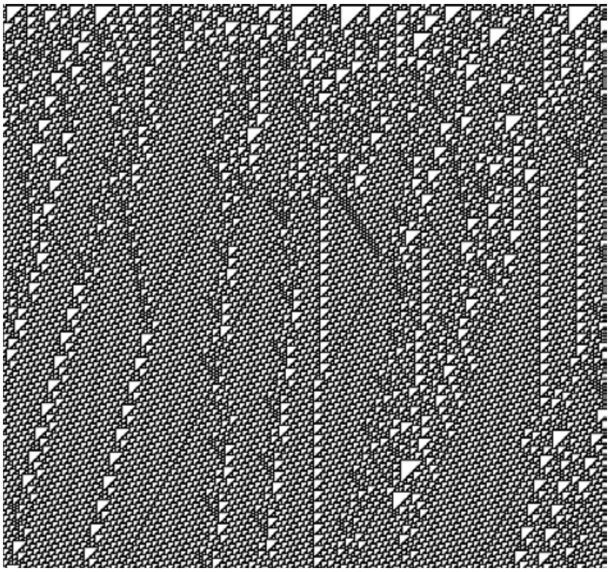


Fig. 4: Zoomed in version of evolution pattern of T(n)

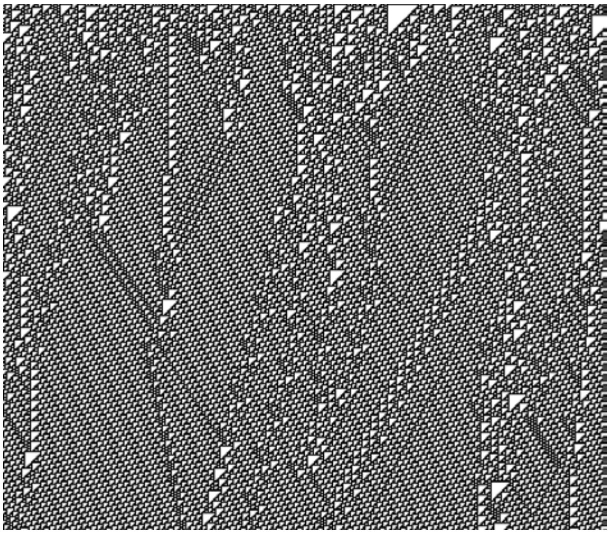


Fig. 5: Zoomed in version of evolution pattern of G(n)

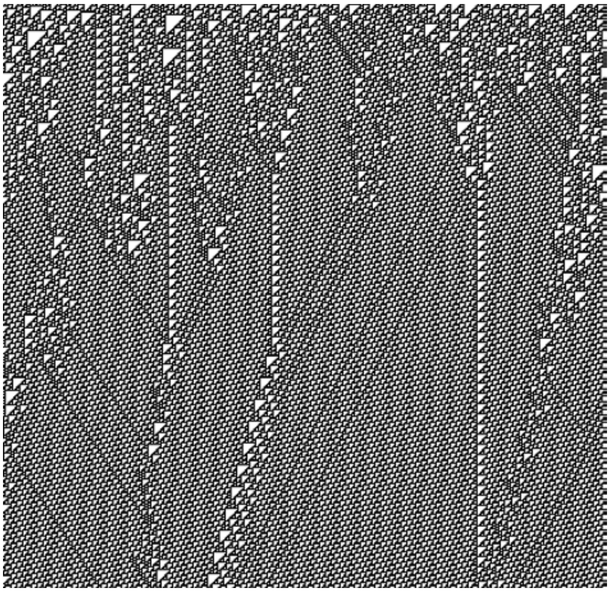
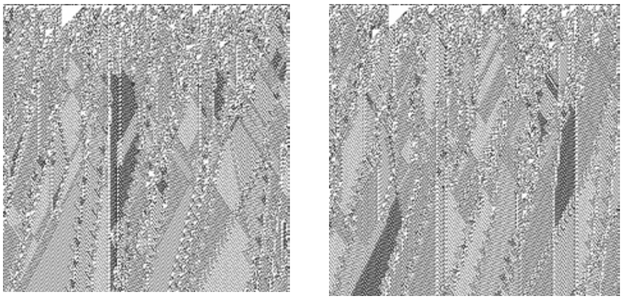


Fig. 6: Zoomed in version of evolution pattern of C(n)

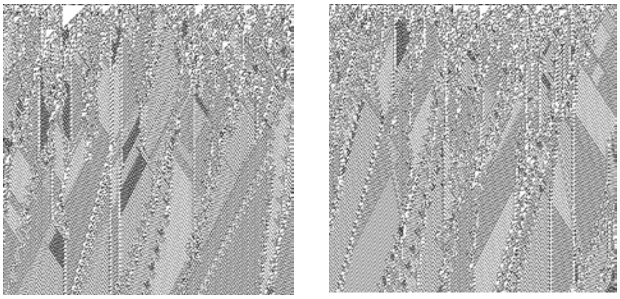
V. Dyadic Operations between Cellular Automata Evolutions of Genome Adjoints

It is a matter of an empirical study to examine the conjoint behavior of various evolution patterns of adjoints and it could be carried out by combining evolution patterns of adjoints dyadically. The various dyadic operations are (i) Boolean addition, (ii) Boolean subtraction, (iii) Boolean multiplication, (iv) Boolean division, (v) Dyadic relation of maximum and (vi) Dyadic relation of minimum. Out of these six different dyadic operations and relations, the Boolean operation of binary addition is considered here for the intended study.



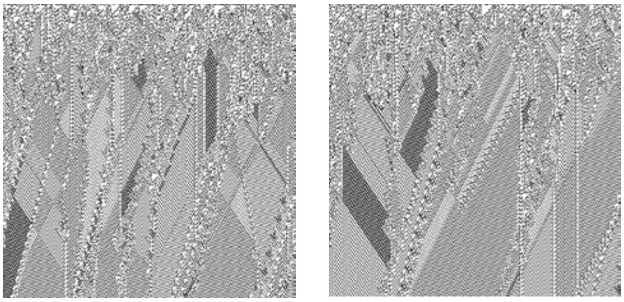
Addition of patterns of A and T

Addition of patterns of A and G



Addition of patterns of A and C

Addition of patterns of T and G



Addition of patterns of T and C

Addition of patterns of G and C

Fig. 7: Boolean addition of evolution patterns of adjoints

The zoomed in versions of the Boolean additions of evolution patterns of A(n), T(n), G(n) and C(n) using rule 137 are shown in Figs. 7, 8, 9 and 10 respectively.

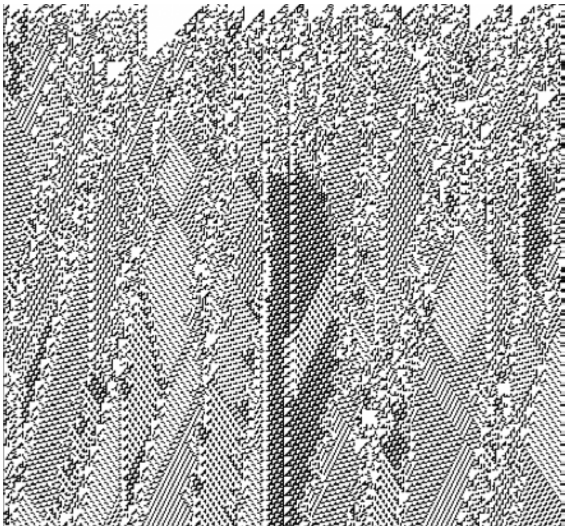


Fig. 8: Zoomed in version of addition of patterns of A and T

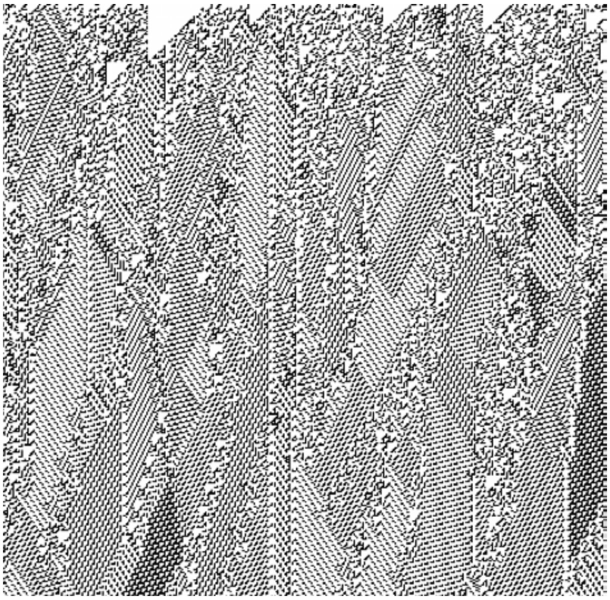


Fig. 9: Zoomed in version of addition of patterns of A and G

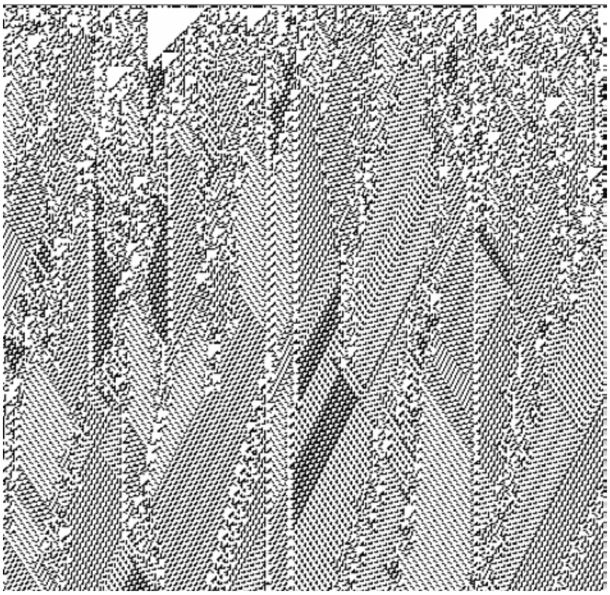


Fig. 10: Zoomed in version of addition of patterns of A and C

VI. OBSERVATIONS AND CONCLUSIONS

From the above empirical study, it is observed that cellular automata modeling and simulation of evolutions of adjoints of a given genome sequence and the inter-pattern operations and relations exhibit distinct patterns of fractals and fractal distributions. The novel technique and results presented in this paper are outcome of prolonged research carried out in the mathematical modeling of genomes and their evolutions. It is evident that one can as well look into the possibilities of genome editing using such cellular automata tools.

ACKNOWLEDGEMENT

The authors express their profound gratitude to the management of Pentagram Research Centre Private Limited, Hyderabad for their boundless support in carrying out research in their premises. The authors further put on record the invaluable guidance of Prof. Dr E G Rajan, President of Pentagram Research and data scientists Mr. Rahul Sharma, Mr. Srikanth Maddikunta and Mr. Shubham Karande.

REFERENCES RÉFÉRENCES REFERENCIAS

1. "Worlds Record Breaking Plant: Deletes its Noncoding "Junk" DNA". Design & Trend. May 12, 2013. Retrieved 2013-06-04.
2. Pennisi, E. (6 September 2012). "ENCODE Project Writes Eulogy for Junk DNA". Science 337 (6099): 1159–1161. doi:10.1126/science.337.6099.1159. PMID 22955811.
3. The ENCODE Project Consortium (2012). "An integrated encyclopedia of DNA elements in the human genome". Nature 489 (7414): 57–74. Bibcode:2012Natur.489...57T.doi:10.1038/nature11 247. PMC 3439153. PMID 22955616..
4. Costa, Fabrico (2012). "7 Non-coding RNAs, Epigenomics, and Complexity in Human Cells". In Morris, Kevin V. Non-coding RNAs and Epigenetic Regulation of Gene Expression: Drivers of Natural Selection. Caister Academic Press. ISBN 1904455948.
5. Carey, Nessa (2015). Junk DNA: A Journey Through the Dark Matter of the Genome. Columbia University Press. ISBN 9780231170840.
6. Robin McKie (24 February 2013). "Scientists attacked over claim that 'junk DNA' is vital to life". The Observer.
7. Sean Eddy (2012) The C-value paradox, junk DNA, and ENCODE, Curr Biol 22(21):R898–R899.
8. Doolittle, W. Ford (2013). "Is junk DNA bunk? A critique of ENCODE". Proc Natl Acad Sci USA 110 (14): 5294–5300. Bibcode: 2013PNAS..110.5294D. doi:10.1073/pnas.1221376110. PMC 3619371. PMID 23479647.
9. Palazzo, Alexander F.; Gregory, T. Ryan (2014). "The Case for Junk DNA". PLoS Genetics 10 (5): e1004351. doi:10.1371/journal.pgen.1004351. ISSN 1553-7404.
10. Dan Graur, Yichen Zheng, Nicholas Price, Ricardo B. R. Azevedo1, Rebecca A. Zufall and Eran Elhaik (2013). "On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE" (PDF). Genome Biology and Evolution 5 (3): 578–90. doi:10.1093/gbe/evt028. PMC 3622293. PMID 23431001.

11. Ponting, CP; Hardison, RC (2011). "What fraction of the human genome is functional?". Genome Research 21: 1769–1776. doi:10.1101/gr.116814.110. PMC 3205562. PMID 21875934.
12. Kellis, M.; et al. (2014). "Defining functional DNA elements in the human genome". PNAS 111 (17): 6131–6138. Bibcode: 2014PNAS..111.6131K. doi: 10.1073 /pnas.1318948111. PMC 4035993. PMID 24753594.
13. Chris M. Rands, Stephen Meader, Chris P. Ponting and Gerton Lunter (2014). "8.2% of the Human Genome Is Constrained: Variation in Rates of Turnover across Functional Element Classes in the Human Lineage". PLoS Genet 10 (7): e1004525. doi:10.1371 /journal.pgen.1004525. PMC 4109858. PMID 25057982.
14. Mattick JS, Dinger ME (2013). "The extent of functionality in the human genome". The HUGO Journal 7 (1): 2. doi:10.1186/1877-6566-7-2.
15. Morris, Kevin, ed. (2012). Non-Coding RNAs and Epigenetic Regulation of Gene Expression: Drivers of Natural Selection. Norfolk, UK: Caister Academic Press. ISBN 1904455948.
16. Elgar G, Vavouri T; Vavouri (July 2008). "Tuning in to the signals: noncoding sequence conservation in vertebrate genomes". Trends Genet. 24 (7): 344–52. doi:10.1016/j.tig.2008.04.005. PMID 18514361.
17. Gregory TR, Hebert PD; Hebert (April 1999). "The modulation of DNA content: proximate causes and ultimate consequences". Genome Res. 9 (4): 317–24. doi:10.1101/gr.9.4.317 (inactive 2015-02-01). PMID 10207154.
18. Wahls, W.P.; et al. (1990). "Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells". Cell 60 (1): 95–103. doi:10.1016/0092-8674(90)90719-U. PMID 2295091.
19. Waterhouse, Peter M.; Hellens, Roger P. (25 March 2015). "Plant biology: Coding in non-coding RNAs". Nature 520 (7545): 41–42. doi:10.1038/nature14378.
20. Li M, Marin-Muller C, Bharadwaj U, Chow KH, Yao Q, Chen C; Marin-Muller; Bharadwaj; Chow; Yao; Chen (April 2009). "MicroRNAs: Control and Loss of Control in Human Physiology and Disease". World J Surg 33 (4): 667–84. doi:10.1007/s00268-008-9836-x. PMC 2933043. PMID 19030926.
21. Visel A, Rubin EM, Pennacchio LA (September 2009). "Genomic Views of Distant-Acting Enhancers". Nature 461 (7261): 199–205. Bibcode: 2009Natur.461..199V. doi:10.1038/ nature 08451. PMC 2923221.PMID 19741700.
22. Nielsen H, Johansen SD; Johansen (2009). "Group I introns: Moving in new directions". RNA Biol 6 (4): 375–83. doi:10.4161/rna.6.4.9334. PMID 19667762.
23. Zheng D, Frankish A, Baertsch R; et al. (June 2007). "Pseudogenes in the ENCODE regions: Consensus annotation, analysis of transcription, and evolution". Genome Res. 17 (6): 839–51.doi:10.1101/gr.5586307. PMC 1891343. PMID 17568002.
24. Marshall CR, Raff EC, Raff RA; Raff; Raff (December 1994). "Dollo's law and the death and resurrection of genes". Proc. Natl. Acad. Sci. U.S.A. 91 (25): 12283–7. Bibcode:1994PNAS...9112283M.doi:10.1073/pnas.91.25 12283. PMC 45421. PMID 7991619.
25. Tutar, Y. (2012). "Pseudogenes". Comp Funct Genomics 2012: 424526. doi:10.1155/2012/424526. PMC 3352212. PMID 22611337.
26. Petrov DA, Hartl DL; Hartl (2000). "Pseudogene evolution and natural selection for a compact genome". J. Hered. 91 (3): 221–7. doi:10.1093/jhered/91.3.221. PMID 10833048.
27. Ponicsan SL, Kugel JF, Goodrich JA; Kugel; Goodrich (February 2010). "Genomic gems: SINE RNAs regulate mRNA production". Current Opinion in Genetics & Development 20 (2): 149–55. doi:10.1016/j.gde. 2010.01.004. PMC 2859989. PMID 20176473.
28. Häsler J, Samuelsson T, Strub K; Samuelsson; Strub (July 2007). "Useful 'junk': Alu RNAs in the human transcriptome". Cell. Mol. Life Sci. 64 (14): 1793–800. doi:10.1007/s00018-007-7084-0.PMID 17514354.
29. Walters RD, Kugel JF, Goodrich JA; Kugel; Goodrich (Aug 2009). "InvAluable junk: the cellular impact and function of Alu and B2 RNAs". IUBMB Life 61 (8): 831–7. doi:10.1002/iub.227. PMC 4049031.PMID 19621349.
30. Nelson, PN.; Hooley, P.; Roden, D.; Davari Ejtehadi, H.; Rylance, P.; Warren, P.; Martin, J.; Murray, PG. (Oct 2004). "Human endogenous retroviruses: transposable elements with potential?". Clin Exp Immunol 138 (1): 1–9. doi:10.1111/j.1365-2249.2004.02592.x. PMC 1809191. PMID 15373898.
31. International Human Genome Sequencing Consortium (February 2001). "Initial sequencing and analysis of the human genome". Nature 409 (6822): 879–888. Bibcode: 2001Natur.409..860L. doi:10.1038/35057062. PMID 11237011.
32. Piegu, B.; Guyot, R.; Picault, N.; Roulin, A.; Sanyal, A.; Saniyal, A.; Kim, H.; Collura, K.; et al. (Oct 2006). "Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice". Genome Res 16 (10): 1262–9. doi:10.1101/gr.5290206. PMC 1581435. PMID 16963705.
33. Hawkins, JS.; Kim, H.; Nason, JD.; Wing, RA.; Wendel, JF. (Oct 2006). "Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium". Genome Res 16 (10): 1252–61. doi:10.1101/gr.5282906. PMC 1581434. PMID 16954538.

34. Ehret CF, De Haller G; De Haller (1963). "Origin, development, and maturation of organelles and organelle systems of the cell surface in Paramecium". Journal of Ultrastructure Research. 9 Supplement 1: 1, 3–42. doi:10.1016/S0022-5320(63)80088-X. PMID 14073743.
35. Dan Graur, The Origin of Junk DNA: A Historical Whodunnit
36. Gregory, T. Ryan, ed. (2005). The Evolution of the Genome. Elsevier. pp. 29–31. ISBN 0123014638. Comings (1972), on the other hand, gave what must be considered the first explicit discussion of the nature of "junk DNA," and was the first to apply the term to all noncoding DNA."; "For this reason, it is unlikely that any one function for noncoding DNA can account for either its sheer mass or its unequal distribution among taxa. However, dismissing it as no more than "junk" in the pejorative sense of "useless" or "wasteful" does little to advance the understanding of genome evolution. For this reason, the far less loaded term "noncoding DNA" is used throughout this chapter and is recommended in preference to "junk DNA" for future treatments of the subject."
37. Ohno, Susumu (1972). H. H. Smith, ed. So Much "junk" DNA in Our Genome. Gordon and Breach, New York. pp. 366–370. Retrieved 2013-05-15.
38. Doolittle WF, Sapienza C; Sapienza (1980). "Selfish genes, the phenotype paradigm and genome evolution". Nature 284 (5757): 601–603. Bibcode: 1980Natur.284..601D. doi:10.1038/284601a0. PMID 6245369.
39. Another source is genome duplication followed by a loss of function due to redundancy.
40. Orgel LE, Crick FH; Crick (April 1980). "Selfish DNA: the ultimate parasite". Nature 284 (5757): 604–7. Bibcode:1980Natur.284..604O.doi:10.1038/284604a0. PMID 7366731.
41. Khajavinia A, Makalowski W; Makalowski (May 2007). "What is "junk" DNA, and what is it worth?". Scientific American 296 (5): 104. doi:10.1038/scientificamerican0307-104. PMID 17503549. The term "junk DNA" repelled mainstream researchers from studying noncoding genetic material for many years
42. Biémont, Christian; Vieira, C (2006). "Genetics: Junk DNA as an evolutionary force". Nature 443 (7111): 521–4. Bibcode: 2006Natur.443..521B. doi:10.1038/443521a. PMID 17024082.
43. Palazzo, Alexander F.; Lee, Eliza S. (2015). "Non-coding RNA: what is functional and what is junk?". Frontiers in Genetics6: 2. doi: 10.3389/fgene.2015.00002. ISSN 1664-8021. PMID 25674102.
44. Ludwig MZ (December 2002). "Functional evolution of noncoding DNA". Current Opinion in Genetics & Development 12 (6): 634–9. doi:10.1016/S0959-437X(02)00355-6. PMID 12433575.
45. Cobb J, Büsst C, Petrou S, Harrap S, Ellis J; Büsst; Petrou; Harrap; Ellis (April 2008). "Searching for functional genetic variants in non-coding DNA". Clin. Exp. Pharmacol. Physiol. 35 (4): 372–5. doi:10.1111/j.1440-1681.2008.04880.x. PMID 18307723.
46. E Khurana; et al. (April 2013). "Integrative annotation of variants from 1092 humans: application to cancer genomics". Science 342 (6154): 372–5. doi:10.1126/science.1235587. PMC 3947637. PMID 24092746.
47. Lu, Yi-Fan; Mauger, David M.; Goldstein, David B.; Urban, Thomas J.; Weeks, Kevin M.; Bradrick, Shelton S. (4 November 2015). "IFNL3 mRNA structure is remodeled by a functional non-coding polymorphism associated with hepatitis C virus clearance". Scientific Reports 5: 16037. doi:10.1038/srep16037. PMID 26531896.
48. Grünewald, Thomas G P; Bernard, Virginie; Gilardi-Hebenstreit, Pascale; Raynal, Virginie; Surdez, Didier; Aynaud, Marie-Ming; Mirabeau, Olivier; Cidre-Aranaz, Florencia; Tirode, Franck. "Chimeric EWSR1-FLI1 regulates the Ewing sarcoma susceptibility gene EGR2 via a GGAA microsatellite". Nature Genetics 47 (9): 1073–1078. doi:10.1038/ng.3363. PMC 4591073. PMID 26214589.
49. Subirana JA, Messeguer X; Messeguer (March 2010). "The most frequent short sequences in non-coding DNA". Nucleic Acids Res. 38 (4): 1172–81. doi:10.1093/nar/gkp1094. PMC 2831315. PMID 19966278.
50. S. E. Ahnert; T. M. A. Fink (2008). "How much non-coding DNA do eukaryotes require?" (PDF). J. Theor. Biol. 252 (4): 587–592. doi:10.1016/j.jtbi.2008.02.005. PMID 18384817.
51. Smith MA; et al. (June 2013). "Widespread purifying selection on RNA structure in mammals". Nucleic Acids Research 41 (17): 8220–8236. doi:10.1093/nar/gkt596. PMC 3783177. PMID 23847102.
52. Dileep, V (2009). "The place and function of non-coding DNA in the evolution of variability". Hypothesis 7 (1): e7. doi:10.5779/ hypothesis. v7i1.146.
53. Carroll, Sean B.; et al. (May 2008). "Regulating Evolution". Scientific American 298 (5): 60–67. doi:10.1038/scientificamerican0508-60. PMID 18444326.
54. Stojic, L. "Transcriptional silencing of long noncoding RNA GNG12-AS1 uncouples its transcriptional and product-related functions". nature.com. Nature. Retrieved 21 Feb 2016.
55. Callaway, Ewen (March 2010). "Junk DNA gets credit for making us who we are". New Scientist.
56. "Plagiarized Errors and Molecular Genetics", talkorigins, by Edward E. Max, M.D., Ph.D.

57. Balakirev ES, Ayala FJ; Ayala (2003). "Pseudogenes: are they "junk" or functional DNA?". Annu. Rev. Genet. 37: 123–51. doi:10.1146/annurev.genet.37.040103.103949. PMID 14616058.
58. C.-K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H. E. Stanley; Buldyrev, SV; Goldberger, AL; Havlin, S; Sciortino, F; Simons, M; Stanley, HE (1992). "Long-range correlations in nucleotide sequences". Nature 356 (6365): 168–70. Bibcode:1992Natur.356..168P. doi:10.1038/356168a0. PMID 1301010.
59. W. Li and, K. Kaneko; Kaneko, K (1992). "Long-Range Correlation and Partial 1/falpha Spectrum in a Non-Coding DNA Sequence" (PDF). Europhys. Lett 17 (7): 655–660. Bibcode:1992EL.....17..655L. doi:10.1209/02955075/17/7/014.
60. S. V. Buldyrev, A. L. Goldberger, S. Havlin, R. N. Mantegna, M. Matsa, C.-K. Peng, M. Simons, and H. E. Stanley; Goldberger, A.; Havlin, S.; Mantegna, R.; Matsa, M.; Peng, C.-K.; Simons, M.; Stanley, H. (1995). "Long-range correlations properties of coding and noncoding DNA sequences: GenBank analysis". Phys. Rev. E 51 (5): 5084–5091. Bibcode: 1995PhRvE..51.5084B.doi:10.1103/PhysRevE.51.5084.