Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

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

Ethirajan Govindarajan

Received: 13 December 2019 Accepted: 2 January 2020 Published: 15 January 2020

Abstract

Structural and functional behavior of genomes could be studied using one dimensional binaryvalued 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.

Index terms— linear boolean functions, cellular automata, genome evolution.

1 Modeling and Simulation of Genome Evolution Using Linear Boolean Functions Associated with

One Dimensional Cellular Automata Prashanthi Govindarajan ? , Sathya Govindarajan ? & Ethirajan Govindarajan ? I. Introduction 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 he 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. T 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 ntuple codons where n is any number.

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 fulllength 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 automatabased genome analytics.

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, 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, 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, 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. A(n) = 01100100100110000110000110000000000000000 T(n) = 000100110000000000000100000101001100 1100 G(n) = 1000100000000110000100000100000000000011 C(n) = 00000000011000011000101000101011001100000

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 onedimensional cellular automaton would consist of a finite length array as shown below.

2 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 ?i-1, ?i, and ?i+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 ?ti-1, ?ti, and ?ti+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 threeneighborhood 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.

3 Linear Boolean Function

Decimal Equivalent 0 0(?? ? ???1 ?? ? ?? ?? ? ??+1) 1 (?? ? ???1 ?? ? ?? ?? ??+1) 2 (?? ? ???1 ?? ? ??) 3 (?? ? ???1 ?? ?? ?? ? ??+1) 4 (?? ? ???1 ?? ? ??+1) 5 (?? ? ???1 ?? ?? ?? ? ??+1)+(?? ? ???1 ?? ? ?? ?? ??+1) 6 (?? ? ???1 ?? ? ??+1) + (?? ? ???1 ?? ? ??) 7 (?? ? ???1 ?? ?? ?? ??+1) 8 (?? ? ???1 ?? ? ?? ?? ? ??+1) + (?? ? ???1 ?? ?? ?? ??+1) 9 (?? ? ???1 ?? ??+1)**10**(?? ? ???1 ?? ? ??) + (?? ? ???1 ?? ??+1) (?? ? ???1 ?? ??) (?? ? ???1 ?? ? ??+1) + (?? ? ???1 ?? ??) (?? ? ???1 ?? ??) + (?? ? ???1 ?? ??+1) (?? ? ???1) (?? ???1 ?? ? ?? ?? ? ??+1) (?? ? ?? ?? ? ??+1) (?? ???1 ?? ? ?? ?? ? ??+1) + (?? ? ???1 ?? ? ?? ?? ??+1) (?? ? ?? ?? ? ??+1) + (?? ? ???1 ?? ? ??) + (?? ???1 ?? ? ?? ?? ? ??+1) + (?? ? ???1 ?? ?? ?? ??+1) (?? ? ???1 ?? ?? ?? ??+1) + (?? ? ?? ?? ? ??+1) (?? ???1 ?? ? ?? ?? ? ??+1) + (?? ? ???1 ?? ??+1) (?? ? ?? ?? ? ??+1) + (?? ? ???1 ?? ??+1) (?? ???1 ?? ? ?? ?? ? ??+1) + (?? ? ???1 ?? ??) (?? ? ?? ?? ? ??+1) + (?? ? ???1 ?? ??+1) (?? ???1 ?? ? ?? ?? ? ??+1) + (?? ? ???1 ?? ??) + (?? ? ???1 ?? ??+1)

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.

The adjoints of the genome sequence segment are given below.

Adjoint A(n)Adjoint T(n) Adjoint G(n) Adjoint C(n)

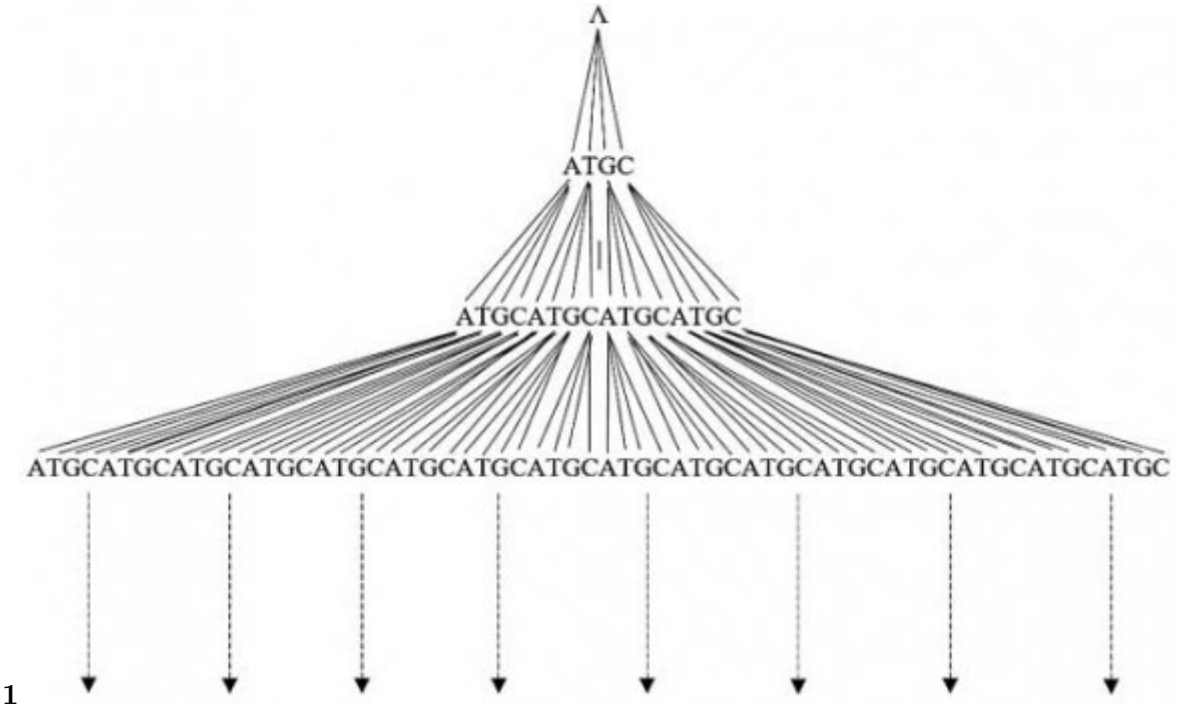
Cellular automata evolutions of adjoints of a genome are carried out using 256 rules of 123CA. As an example, rule number 137 of 123CA, that is, (? ? ???1 ? ? ?? ? ? ??+1) + (? ?? ? ??+1) is applied to adjoints of Brucella Suis 1330 genome and results shown below in Fig. **??**.Evolution of A(n) Evolution of T(n) Evolution of G(n) Evolution of C(n)

Fig. **??**: Evolution of adjoints using rule 137 of 123CA

The size of the images shown in Fig. **??** 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. **??**, 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. **??**, 4, 5 and 6 respectively.

4 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



1

Figure 1: Fig. 1 :

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

Figure 2:

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

34567

Figure 3: Fig. 3 :Fig. 4 :Fig. 5 :Fig. 6 :Fig. 7 :

Row Number	Column 1	Column 2	Column 3	Column 4	Column 5	Column 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

Figure 4: Fig. 8 :Fig. 9 :Fig. 10 :

Row Number	Column 1	Column 2	Column 3	Column 4	Column 5	Column 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

Figure 5:

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. [1]

.1 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.

[Ponting] , C P Ponting .

[Elgar and Vavouri] , G Elgar , T Vavouri .

[Gregory and Hebert] , T R Gregory , P D Hebert .

[Li et al.] , M Li , C Marin-Muller , U Bharadwaj , K H Chow , Q Yao , C; Chen , ; Marin-Muller , ; Bharadwaj , ; Chow , Yao .

[Nielsen and Johansen] , H Nielsen , S D Johansen .

[Genome Res] , 17568002. PMC1891343. *Genome Res* 17 (6) p. .

[Petrov and Hartl] , D A Petrov , D L Hartl .

[Ponicsan et al.] , S L Ponicsan , J F Kugel , Ja; Goodrich , Kugel .

[Häsler et al.] , J Häsler , T Samuelsson , K; Strub , Samuelsson .

[Walters et al.] , R D Walters , J F Kugel , Ja; Goodrich , Kugel .

[Genome Res] , 10.1101/gr.5290206.PMC1581435. 16963705. *Genome Res* 16 (10) p. .

[Khajavinia and Makalowski] , A Khajavinia , W Makalowski .

[Cobb et al.] , J Cobb , C Büsst , S Petrou , S Harrap , J; Ellis , Büsst .

[Petrou and Harrap] , ; Petrou , Harrap .

[Lu] , Yi-Fan Lu .

[Grünewald] , Thomas G Grünewald .

[Bernard et al.] , Virginie ; Bernard , Pascale ; Gilardi-Hebenstreit , Virginie ; Raynal , Surdez , ; Didier , Marie-Ming Aynaud . 10.1038/ng.3363.

[Mirabeau] , Olivier Mirabeau . 10.1038/ng.3363.

[Ahnert] , S E Ahnert .

[Peng et al.] , C.-K Peng , S V Buldyrev , A L Goldberger , S Havlin , F Sciortino , M Simons , H E Stanley .

[Li and Kaneko] , W Li , , K Kaneko .

[Rands et al. ()] '8.2% of the Human Genome Is Constrained: Variation in Rates of Turnover across Functional Element Classes in the Human Lineage'. Chris M Rands , Stephen Meader , Chris P Ponting , Gerton Lunter . 25057982. PMC4109858. *PLoS Genet* 2014. 10 (7) p. e1004525.

[An integrated encyclopedia of DNA elements in the human genome Bibcode:2012Natur.489...57T ()] 'An integrated encyclopedia of DNA elements in the human genome'. 10.1038/nature11247.PMC3439153. 22955616. *Bibcode:2012Natur.489...57T*, 2012. 489 p. .

[Another source is genome duplication followed by a loss of function due to redundancy] *Another source is genome duplication followed by a loss of function due to redundancy*,

[Cidre-Aranaz and Tirode] 'Chimeric EWSR1-FLI1 regulates the Ewing sarcoma susceptibility gene EGR2 via a GGAA microsatellite'. Florencia ; Cidre-Aranaz , Franck Tirode . 10.1038/ng.3363. 26214589. *Nature Genetics* 47 (9) p. .

[Kellis ()] 'Defining functional DNA elements in the human genome'. M Kellis . 24753594. PMC4035993. *Bibcode: 2014PNAS..111.6131K*, 2014. 111 p. .

[Hawkins et al. (2006)] 'Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium'. J S Hawkins , H Kim , J D Nason , R A Wing , J F Wendel . 10.1101/gr.5282906.PMC1581434. 16954538. *Genome Res* Oct 2006. 16 (10) p. .

[Marshall et al. (1994)] 'Dollo's law and the death and resurrection of genes'. C R Marshall , E C Raff , R A Raff , ; Raff , Raff . 7991619. PMC45421. *Bibcode:1994PNAS...9112283M*, December 1994. 91 p. .

[Piegu et al. (2006)] *Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in Oryza australiensis*, B Piegu , R Guyot , N Picault , A Roulin , A Sanyal , A Saniyal , H Kim , K Collura . Oct 2006. (a wild relative of rice)

[Pennisi (2012)] 'ENCODE Project Writes Eulogy for Junk DNA'. E Pennisi . 10.1126/science.337.6099.1159. 22955811. *Science* 6 September 2012. 337 (6099) p. .

[Ludwig (2002)] 'Functional evolution of noncoding DNA'. M Z Ludwig . 10.1016/S0959-437X(02)00355-6. 12433575. *Current Opinion in Genetics & Development* December 2002. 12 (6) p. .

[Biémont and Vieira ()] 'Genetics: Junk DNA as an evolutionary force'. Christian ; Biémont , C Vieira . 17024082. *Bibcode: 2006Natur.443..521B*, 2006. 443 p. .

[Goodrich (2010)] 'Genomic gems: SINE RNAs regulate mRNA production'. Goodrich . 10.1016/j.gde.2010.01.004.PMC2859989. 20176473. *Current Opinion in Genetics & Development* February 2010. 20 (2) p. .

[Visel et al. (2009)] 'Genomic Views of Distant-Acting Enhancers'. A Visel , E M Rubin , L A Pennacchio . 19741700. PMC2923221. *Bibcode: 2009Natur.461..199V*, September 2009. 461 p. .

[Johansen ()] 'Group I introns: Moving in new directions'. Johansen . 10.4161/rna.6.4.9334. 19667762. *RNA Biol* 2009. 6 (4) p. .

[Fink ()] 'How much noncoding DNA do eukaryotes require?" (PDF)'. T M A Fink . 10.1016/j.jtbi.2008.02.005. 18384817. *J. Theor. Biol* 2008. 252 (4) p. .

[Nelson et al. (2001)] 'Human endogenous retroviruses: transposable elements with potential?'. P N Nelson , P Hooley , D Roden , H Davari Ejtehadi , P Rylance , P Warren , J Martin , P G Murray . 10.1111/j.1365-2249.2004.02592.x.PMC1809191. 15373898. PMID 11237011. *International Human Genome Sequencing Consortium*, Oct 2004. February 2001. 138 p. . (Nature. Bibcode: 2001Natur.409..860L.)

[Wahls ()] 'Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells'. W P Wahls . 10.1016/0092-8674(90)90719-U. 2295091. *Cell* 1990. 60 (1) p. .

[Mauger et al. (2015)] 'IFNL3 mRNA structure is remodeled by a functional non-coding polymorphism associated with hepatitis C virus clearance'. David M Mauger , David B Goldstein , Thomas J Urban , Kevin M Weeks , Shelton S Bradrick . 10.1038/srep16037. 26531896. *Scientific Reports* 4 November 2015. 5 p. 16037.

[Khurana (2013)] 'Integrative annotation of variants from 1092 humans: application to cancer genomics'. ; Khurana . 10.1126/science.1235587.PMC3947637. 24092746. *Science* April 2013. 342 (6154) p. .

[Goodrich (2009)] 'InvAluable junk: the cellular impact and function of Alu and B2 RNAs'. Goodrich . 10.1002/iub.227.PMC4049031. 19621349. *IUBMB Life* Aug 2009. 61 (8) p. .

[Doolittle and Ford ()] 'Is junk DNA bunk? A critique of ENCODE'. W Doolittle , Ford . 10.1073/pnas.1221376110.PMC3619371. 23479647. *Bibcode: 2013PNAS..110.5294D*, (USA) 2013. 110 p. .

[Callaway (2010)] 'Junk DNA gets credit for making us who we are'. Ewen Callaway . *New Scientist* Edward E. Max, M.D., Ph.D. 57 (ed.) March 2010. 56. (Ayala FJ)

[Carey ()] *Junk DNA: A Journey Through the Dark Matter of the Genome*, Nessa Carey . 2015. Columbia University Press. (ISBN 9780231170840)

[Kaneko ()] 'Long-Range Correlation and Partial 1/falpha Spectrum in a Non-Coding DNA Sequence" (PDF)'. K Kaneko
Bibcode:1992EL.....17..655L . 10.1209/02955075/17/7/014. *Europhys. Lett* 1992. 17 (7) p. .

[Buldyrev et al. ()] 'Long-range correlations in nucleotide sequences'. S V Buldyrev , Goldberger , S; Al; Havlin , F; Sciortino , M; Simons , Stanley , He . 10.1038/356168a0. 1301010. *Bibcode:1992Natur.356..168P*, 1992. 356 p. .

[Buldyrev et al. ()] 'Long-range correlations properties of coding and noncoding DNA sequences: GenBank analysis'. S V Buldyrev , A L Goldberger , S Havlin , R N Mantegna , M Matsa , C.-K Peng , M Simons , H E Stanley; Goldberger , A Havlin , S Mantegna , R Matsa , M Peng , C.-K Simons , M Stanley , H . 10.1103/PhysRevE.51.5084. *Phys. Rev. E* 1995. 1995PhRvE. 51 (5) p. .

[Chen (2009)] 'MicroRNAs: Control and Loss of Control in Human Physiology and Disease'. Chen . 10.1007/s00268-008-9836-x.PMC2933043. 19030926. *World J Surg* April 2009. 33 (4) p. .

[Costa ()] *Non-coding RNAs and Epigenetic Regulation of Gene Expression: Drivers of Natural Selection*, Fabrico Costa . 2012. 1904455948. Morris, Kevin V: Caister Academic Press. (7 Non-coding RNAs, Epigenomics, and Complexity in Human Cells)

[Morris ()] *Non-Coding RNAs and Epigenetic Regulation of Gene Expression: Drivers of Natural Selection*, Kevin Morris . 2012. 1904455948. Norfolk, UK: Caister Academic Press.

[Palazzo and Lee ()] 'Noncoding RNA: what is functional and what is junk?'. Alexander F Palazzo , Eliza S Lee . 10.3389/fgene.2015.00002. 25674102. *Frontiers in* 1664-8021. 2015. 6 (2) .

[Graur et al. ()] 'On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE" (PDF)'. Dan Graur , Yichen Zheng , Nicholas Price , Ricardo B R Azevedo1 , Rebecca A Zufall , Eran Elhaik . 10.1093/gbe/evt028.PMC3622293. 23431001. *Genome Biology and Evolution* 2013. 5 (3) p. .

[Waterhouse and Hellens (2015)] 'Plant biology: Coding in non-coding RNAs'. Peter M Waterhouse , Roger P Hellens . 10.1038/nature14378. *Nature* 25 March 2015. 520 (7545) p. .

.1 Acknowledgement

[Hartl ()] 'Pseudogene evolution and natural selection for a compact genome'. Hartl . 10.1093/jhered/91.3.221. 10833048. *J. Hered* 2000. 91 (3) p. .

[Tutar ()] 'Pseudogenes'. Y Tutar . 10.1155/2012/424526.PMC3352212. 22611337. *Comp Funct Genomics* 2012. 2012. p. 424526.

[Zheng et al. (2007)] *Pseudogenes in the ENCODE regions: Consensus annotation, analysis of transcription, and evolution*, D Zheng , A Frankish , R Baertsch . June 2007.

[Ayala ()] 'Pseudogenes: are they "junk" or functional DNA?'. Ayala . 14616058. *Annu. Rev. Genet* 2003. 37 p. .

[Carroll (2008)] 'Regulating Evolution'. Sean B Carroll . 10.1038/scientificamerican0508-60. 1844. *Scientific American* May 2008. 298 (5) p. .

[Mckie (2013)] *Scientists attacked over claim that 'junk DNA' is vital to life*, Robin Mckie . 24 February 2013. (The Observer)

[Ellis (2008)] 'Searching for functional genetic variants in non-coding DNA'. Ellis . 10.1111/j.1440-1681.2008.04880.x. 18307723. *Clin. Exp. Pharmacol. Physiol* April 2008. 35 (4) p. .

[Orgel et al. (1980)] 'Selfish DNA: the ultimate parasite'. L E Orgel , Fh; Crick , Crick . 10.1038/284604a0. 7366731. *Nature* April 1980. 284 p. . (Bibcode:1980Natur.284..604O.)

[Doolittle et al. ()] 'Selfish genes, the phenotype paradigm and genome evolution'. W F Doolittle , C; Sapienza , Sapienza . 10.1038/284601a0. 6245369. *Bibcode: 1980Natur.284..601D*, 1980. 284 p. .

[Ohno ()] 'So Much "junk'. Susumu Ohno . *DNA in Our Genome. Gordon and Breach* H. H. Smith (ed.) 1972. 2013-05-15. p. .

[Eddy ()] 'The C-value paradox, junk DNA, and ENCODE'. Sean Eddy . *Curr Biol* 2012. 22 (21) p. .

[Palazzo et al. ()] 'The Case for Junk DNA'. Alexander F Palazzo , T Gregory , Ryan . 10.1371/journal.pgen.1004351.ISSN1553-7404. *PLoS Genetics* 2014. 10 (5) p. e1004351.

[Mattick and Dinger ()] 'The extent of functionality in the human genome'. J S Mattick , M E Dinger . 10.1186/1877-6566-7-2. *The HUGO Journal* 2013. 7 (1) p. 2.

[Hebert (1999)] 'The modulation of DNA content: proximate causes and ultimate consequences'. Hebert . 10.1101/gr.9.4.317(inactive2015-02-01. 10207154. *Genome Res* April 1999. 9 (4) p. .

[Subirana et al. (2010)] 'The most frequent short sequences in noncoding DNA'. J A Subirana , X; Messeguer , Messeguer . 10.1093/nar/gkp1094.PMC2831315. PMID. *Nucleic Acids Res* March 2010. 19966278. 38 (4) p. .

[Dileep ()] 'The place and function of noncoding DNA in the evolution of variability'. V Dileep . doi:10.5779/hypothesis. v7i1.146. *Hypothesis* 2009. 7 (1) p. e7.

[Makalowski (2007)] 'The term "junk DNA" repelled mainstream researchers from studying noncoding genetic material for many years'. Makalowski . 17503549. *Scientific American* May 2007. 296 (5) p. . (What is "junk" DNA, and what is it worth?)

[Stojic (2016)] 'Transcriptional silencing of long noncoding RNA GNG12-AS1 uncouples its transcriptional and product-related functions'. L Stojic . *nature.com. Nature. Retrieved* 21 Feb 2016.

[Vavouri (2008)] 'Tuning in to the signals: noncoding sequence conservation in vertebrate genomes'. Vavouri . 10.1016/j.tig.2008.04.005. 18514361. *Trends Genet* July 2008. 24 (7) p. .

[Strub (2007)] 'Useful 'junk': Alu RNAs in the human transcriptome'. Strub . 10.1007/s00018-007-7084-0. 17514354. *Cell. Mol. Life Sci* July 2007. 64 (14) p. .

[Hardison ()] 'What fraction of the human genome is functional?'. R C Hardison . 10.1101/gr.116814.110.PMC3205562. 21875934. *Genome Research* 2011. 21 p. .

[Smith (2013)] 'Widespread purifying selection on RNA structure in mammals'. M A Smith . 10.1093/nar/gkt596.PMC3783177. 23847102. *Nucleic Acids Research* June 2013. 41 (17) p. .

[Worlds Record Breaking Plant: Deletes its Noncoding "Junk" DNA". Design Trend (2013)] *Worlds Record Breaking Plant: Deletes its Noncoding "Junk" DNA". Design & Trend*, May 12. 2013. (References Références Referencias 1. Retrieved 2013-06-04)