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.*

On the Notion of Percentage Nucleotide Concentration of Genome Sequences in Terms of Cellular Automata Evolutions of Adjoints Sequences

Ethirajan Govindarajan

Received: 8 December 2019 Accepted: 5 January 2020 Published: 15 January 2020

Abstract

This paper proposes a novel concept called ?Percentage Nucleotide Concentration of genomes? in terms of cellular automata evolutions of adjoints of Adenine, Thymine, Guanine, and Cytosine. The adjoints of the given a genome sequenceare the characteristic binary string sequences. For example, the adjoint of Adenine of a given genome sequence is a binary string consisting of 0?s and 1?s where 1?s corresponds to the presence of Adenine in the genome sequence. So, one can have four adjoint sequences of Adenine, Thymine, Guanine, and Cytosine corresponding to a given genome sequence. One-dimensional three neighborhood binary value cellular automata rules could be applied to an adjoint sequence and the desired number of evolutions obtained.These rules aredefined by linear Boolean functions and one can have 256 such linear Boolean functions.Nucleotide concentration is computed for an adjoint sequence and its variation evaluated for its successive evolutions based on a cellular automaton rule.

Index terms— cellular automata, evolutions of adjoints, linear boolean functions, nucleotide concentration in a genome.

1 On the Notion of Percentage Nucleotide Concentration of Genome Sequences in Terms of Cellular Automata Evolutions of Adjoints Sequences

Prashanthi Govindarajan ? , Sathya Govindarajan ? & Ethirajan Govindarajan ?

Abstract-This paper proposes a novel concept called ”Percentage Nucleotide Concentration of genomes” in terms of cellular automata evolutions of adjoints of Adenine, Thymine, Guanine, and Cytosine. The adjoints of the given a genome sequenceare the characteristic binary string sequences. For example, the adjoint of Adenine of a given genome sequence is a binary string consisting of 0’s and 1’s where 1’s corresponds to the presence of Adenine in the genome sequence. So, one can have four adjoint sequences of Adenine, Thymine, Guanine, and Cytosine corresponding to a given genome sequence. One-dimensional three neighborhood binary value cellular automata rules could be applied to an adjoint sequence and the desired number of evolutions obtained.These rules aredefined by linear Boolean functions and one can have 256 such linear Boolean functions.Nucleotide concentration is computed for an adjoint sequence and its variation evaluated for its successive evolutions based on a cellular automaton rule.

2 I. Introduction

he purpose of the research carried out and reported in this paper is whether it is possible to categorize a set of genomes like the human genome repository. The concept of ”%nucleotide concentration” introduced in this paper seems to show a way to accomplish this task. The genesis of the formulation of this concept originates

from chemistry, wherein the quantificational notion of percentage ionic concentration of hydrogen (pH value) is used to categorize solutions into three (i) water, whose pH value is 7, (ii) acidic solutions whose pH values are less than 7 and (iii) alkaline solutions whose pH values are greater than 7. On the same lines, an effort was made to categorize genome sets based on four values (i) % nucleotide concentration of Adenine (pA), (ii) % nucleotide concentration of Thymine (pT), (iii) % nucleotide concentration of Guanine (pG) and (iv) % nucleotide concentration of Cytosine (pC). It is reasonable to surmise that these values, possibly their compositions would categorize a given set of genomes. The formulation of the concept is briefly explained below. Section 2 of this paper describes the concept formulation.

Section 3 of this paper describes the fundamental notions of adjoints of a genome and their evolution using one dimensional cellular automata rules defined by linear Boolean functions. Section 4 provides experimental results of a case study pertaining to evaluation of Concentration of Nucleotides in terms of Adjoints of BrucellaSuis 1330 Genome Sequence.

3 II. Concept Formulation

Analogous to the notion of pH value of a solution, the values of pA, pT, pG and pC of a genome sequence and possibly composition of these values like the proportion pA:pT:pG:pC seems to pave a way to classify and characterize genome sets. The definition of "Percentage Nucleotide Concentration" of a genome sequence is given below.

4 Definition

Given a genome sequence, the number of a particular nucleotide, say A, present in that genome sequence is counted and the sum is divided by the total number of nucleotides in that genome sequence. The fraction when multiplied by 100 yields the "Percentage Concentration of Adenine pA". Similarly, one can evaluate pT, pG and pC.

5 One-Dimensional Three Neighborhood Cellular Automata Evolutions of Adjointsof a Genome Sequence

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 of BrucellaSuis 1330 for a case study. The actual length of the genome sequence of BrucellaSuis 1330 is 5806. A cellular T 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 twenty linear Boolean functions of cellular automata 123CA are listed below with their decimal equivalents.

6 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) 11 (?? ? ???1 ?? ??) 12 (?? ? ???1 ?? ? ??+1) + (?? ? ???1 ?? ??) 13 (?? ? ???1 ?? ??) + (?? ? ???1 ?? ??+1) 14 (?? ? ???1) 15 (?? ???1 ?? ? ?? ?? ? ??+1) 16 (?? ? ?? ?? ? ??+1) 17 (?? ???1 ?? ? ?? ?? ? ??+1) + (?? ? ???1 ?? ? ?? ?? ??+1) 18 (?? ? ?? ?? ? ??+1) + (?? ? ???1 ?? ? ??) 19 (?? ???1 ?? ? ?? ?? ? ??+1) + (?? ? ???1 ?? ?? ?? ? ??+1)**20**

For the case study rule number 90 is applied to the adjoints of BrucellaSuis 1330 genome sequence and 500 evolutions generated. Rule 90 is shown below.(? ???1 ? ? ??+1) + (? ? ???1 ? ??+1)**90**

Since the image of the 500 evolutions of BrucellaSuis 1330 is large, a small portion of the images are presented in this paper.

7 Concentration of Nucleotidesin Adjoints of Brucellasuis 1330 Genome Sequence

The values of pA, pT, pG and pCof the BrucellaSuis 1330 genome sequence are computed for the adjoints A(n), T(n),G(n) and C(n) and their 500 evolutions using 123CA rules based one linear Boolean functions. Fig. **??** shows the evolutions of the adjoints of A(n), T(n). G(n) and C(n) using the linear Boolean function rule 90 of 123CA. The values are tabulated and the corresponding graphs shown subsequently. Table 1 shows the pA values of A(n) of BrucellaSuis 1330 genome sequence and the 500 generations of A(n) using rule 90 of 123CA. Figs. **??** and 3 shows the graphs of the variations of pA values of all generations. Table 2 shows the pT values of T(n) of BrucellaSuis 1330 genome sequence and the 500 generations of T(n) using rule 90 of 123CA. Figs. **??**and 5 shows the graph of the variations of pT values of all generations. Table 3 shows the pG values of G(n) of BrucellaSuis 1330 genome sequence and the 500 generations of G(n) using rule 90 of 123CA. Fig. **??** shows the graph of variations of pA values of all generations. Table **??** shows the pC values of C(n) of BrucellaSuis 1330 genome sequence and 500 generations of C(n) using rule 90 of 123CA. Fig. **??** shows the graph of the variations of pC values of all generations. This paper proposes a novel concept called "Percentage Nucleotide Concentration of genomes" in terms of cellular automata evolutions of adjoints of Adenine, Thymine, Guanine, and Cytosine. The research carried out and reported in this paper exhibits the possibility to categorize a set of genomes like the human genome repository. In short, the concept of "Percentage Nucleotide Concentration (PNC)" introduced in this paper seems to show a way to accomplish this task. [1]

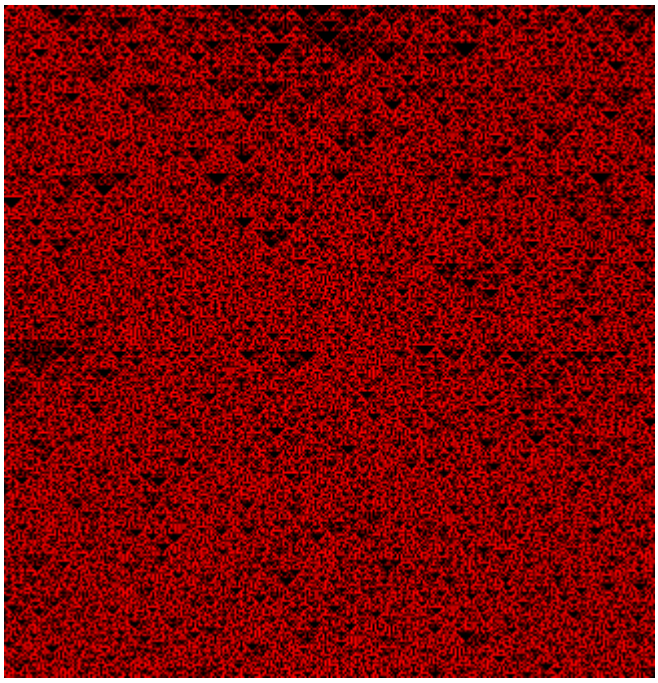
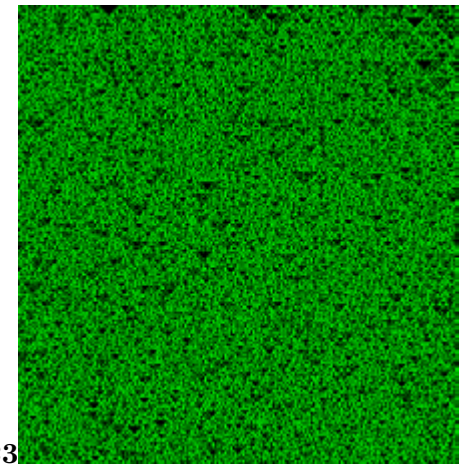
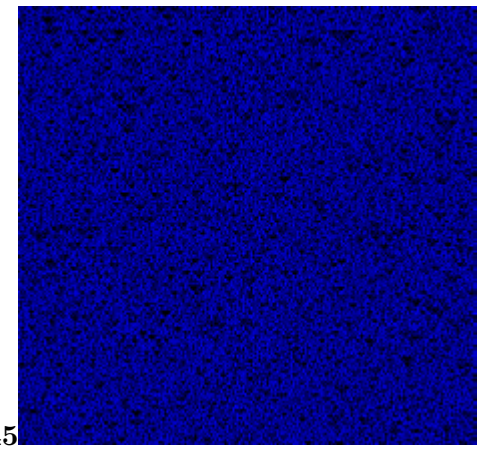


Figure 1:



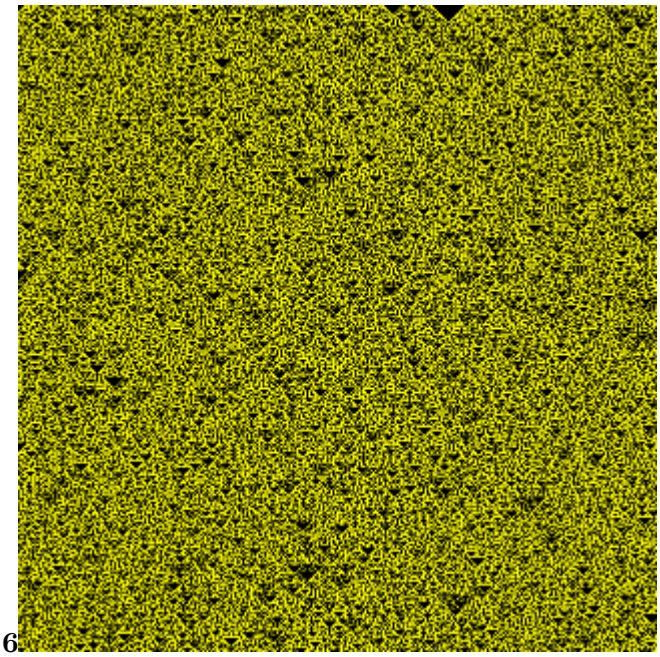
23

Figure 2: Fig. 2 :Fig. 3 :



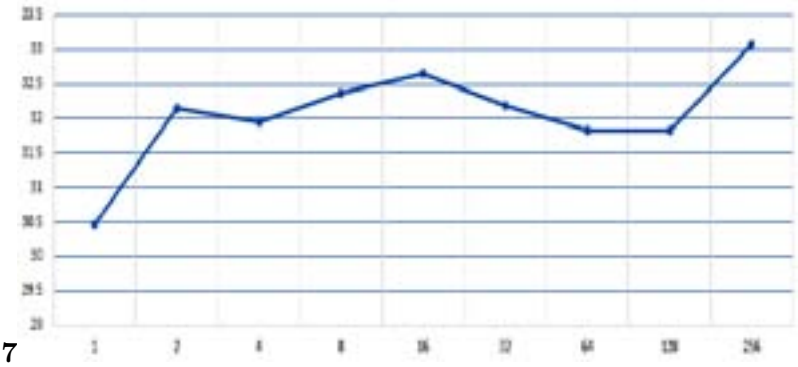
45

Figure 3: Fig. 4 :Fig. 5 :



6

Figure 4: Fig. 6 :



7

Figure 5: Fig. 7 :

1	19.18705	21	42.7661	41	42.31829	61	50.08612	81	43.1278
2	30.50293	22	49.13882	42	48.93214	62	50.24113	82	48.65656
3	30.29625	23	48.67379	43	48.36376	63	52.10127	83	49.3455
4	40.80262	24	50.55115	44	50.72339	64	48.7599	84	49.65553
5	30.38236	25	42.95556	45	46.60696	65	30.89907	85	49.31106
6	41.62935	26	47.06762	46	49.56941	66	41.54323	86	49.63831
7	41.52601	27	48.62212	47	50.22391	67	41.78136	87	50.15501
8	48.51877	28	50.51671	48	50.51671	68	47.24423	88	49.68998
9	31.01963	29	48.69101	49	42.88667	69	41.93937	89	48.31209
10	41.02652	30	50.49948	50	48.19153	70	48.39821	90	49.53496
11	41.14709	31	49.96555	51	49.39718	71	48.50155	91	50.93007
12	48.57044	32	49.39718	52	49.56941	72	49.19049	92	50.15501
13	41.49156	33	30.83018	53	48.58767	73	42.30107	93	49.17327
14	48.19153	34	41.54323	54	50.27558	74	48.32931	94	50.08612
15	47.15811	35	41.6638	55	49.65553	75	49.0527	95	50.20668
16	50.06889	36	48.536	56	50.44781	76	50.20668	96	50.39614
17	31.34688	37	41.92215	57	49.36273	77	47.53703	97	42.42163
18	43.1278	38	46.98588	58	50.17224	78	49.1216	98	48.58767
19	43.83396	39	47.48536	59	50.20668	79	48.48433	99	49.00103
20	50.18945	40	50.08612	60	49.84499	80	49.74155	100	49.86221

8

Figure 6: Fig. 8 :

101	48.57379	121	49.03548	141	48.96659	161	43.04168	181	50.32725
102	49.52108	122	50.68894	142	50.32725	162	48.72546	182	50.17224
103	50.82673	123	50.34447	143	50.36169	163	48.60489	183	49.67275
104	50.17224	124	50.96452	144	51.13676	164	50.55115	184	49.46607
105	46.77919	125	50.80951	145	42.19779	165	48.50155	185	48.94936
106	50.70617	126	51.8257	146	49.63831	166	49.93111	186	49.72442
107	49.60386	127	48.77713	147	49.24216	167	50.13779	187	49.74165
108	49.68598	128	48.72546	148	49.32826	168	50.55115	188	51.89459
109	48.12263	129	31.45022	149	48.8288	169	48.89769	189	50.37892
110	50.17224	130	43.0589	150	49.29383	170	51.18843	190	50.13779
111	49.98278	131	42.85222	151	50.03445	171	49.29383	191	50.17224
112	50.99897	132	49.3455	152	49.50052	172	49.93111	192	50.03445
113	45.81364	133	42.57654	153	49.22494	173	50.39614	193	42.16328
114	48.96659	134	48.01929	154	50.62005	174	49.44885	194	47.84705
115	49.86221	135	49.77609	155	51.18843	175	49.94833	195	48.15708
116	50.13779	136	49.24216	156	49.87944	176	48.81157	196	50.32725
117	48.69101	137	49.16225	157	49.98278	177	49.10437	197	48.05374
118	50.6545	138	48.74268	158	49.86221	178	49.29383	198	49.44885
119	49.94833	139	48.89769	159	50.72339	179	49.05993	199	49.22494
120	50.36169	140	50.17224	160	52.06683	180	49.37935	200	49.3455

9

Figure 7: Fig. 9 :

1

Rule number 90 is applied to A(n) and its 500 generations. It is observed that the pA value becomes minimum at regular intervals of 1, 2, 4, 8, 16, 32, 64, 128 and 256. This indicates a fractal behavior of the evolution. Min(A(n))=30.2965 and Max(A(n))=31.4502. The deviation is 1.15.

Figure 8: Table 1 :

2	Rule number 90 is applied to T(n) and its 500 generations. It is observed that the pT value becomes minimum at regular intervals of 1, 2, 4, 8, 16, 32, 64, 128 and 256. This indicates a fractal behavior of the evolution. Min(A(n))=30.45126; Max(A(n))=33.06924. The deviation is 2.61.

Figure 9: Table 2 :

3	Rule number 90 is applied to G(n) and its 500 generations. It is observed that the pG value becomes minimum at regular intervals of 1, 2, 4, 8, 16, 32, 64, 128 and 256. This indicates a fractal behavior of the evolution. Min(A(n))=43.00723 and Max(A(n))=44.29900 The deviation is 1.46.

Figure 10: Table 3 :

.1 Acknowledgements

[Ponting] , C P Ponting .

[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] , 10.1101/gr.5586307.PMC1891343. 17568002. *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 .

[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 , Gertonlunter . 10.1371/journal.pgen.1004525.PMC4109858. 25057982. *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'. 22955616. PMC3439153. *Bibcode: 2012Natur.489...57T*, 2012. 489 p. .

[Kellis ()] 'Defining functional DNA elements in the human genome'. M Kellis . 10.1073/pnas.1318948111.PMC4035993. 24753594. *Bibcode: 2014PNAS..111.6131K*, 2014. 111 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. *PNAS..9112283M*, December 1994. 1994. 91 p. .

[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. .

[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 . 10.1038/nature08451.PMC2923221. 19741700. *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* 2009. 6 (4) p. .

[Nelson et al. (2004)] 'Human endogenous retroviruses: transposable elements with potential?'. P N Nelson , P Hooley , D Roden , H Davariejtehadi , P Rylance , P Warren , J Martin , P G Murray . 10.1111/j.1365-2249.2004.02592.x.PMC1809191. 15373898. *Clin ExpImmunol* Oct 2004. 138 (1) p. .

[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. .

[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*, 2013. 110 p. .

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

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

[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 , Eranelhaik . 10.1093/gbe/evt028.PMC3622293. 23431001. *Genome Biology* 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. .

[Hartl ()] 'Pseudogene evolution and natural selection for a compact genome'. Hartl . 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.

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

[Eddy ()] 'The C-value paradox, junk DNA, and ENCODE'. Sean Eddy . *CurrBiol* 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* 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. .

[Elgar et al. (2008)] 'Tuning in to the signals: noncoding sequence conservation in vertebrate genomes'. G Elgar , T; Vavouri , 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* 2011. 21 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)