GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: G
INTERDISCIPLINARY
Volume 20 Issue 2 Version 1.0 Year 2020
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

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

By 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.

Keywords: cellular automata, evolutions of adjoints, linear boolean functions, nucleotide concentration in a genome.

GJCST-G Classification: F.1.1

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

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

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.

Keywords: cellular automata, evolutions of adjoints, linear boolean functions, nucleotide concentration in a genome.

I. INTRODUCTION

The 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.

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.

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.

III. 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

---

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-mails: prashanthi89@gmail.com, dr.rajaneg@gmail.com

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 twenty 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

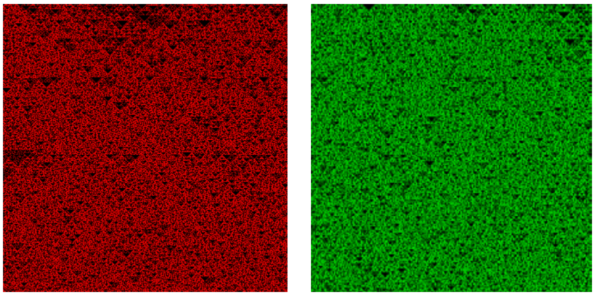
For the case study rule number 90 is applied to the adjoints of BrucellaSuis 1330 genome sequence and 500 evolutions generated. Rule 90 is shown below.

$$(\xi_{i-1}\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\xi_{i+1}) \qquad 90$$

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

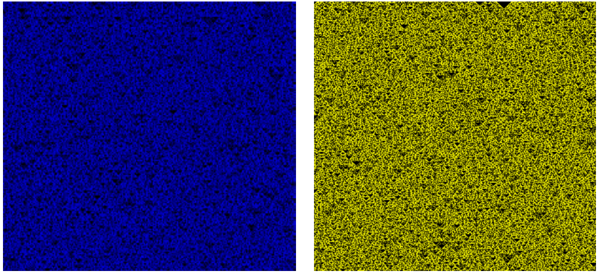
IV. Concentration of Nucleotidesin Adjoints of BrucellaSuis 1330 Genome Sequence

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. 1 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. 2 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. 4and 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. 4 shows the graph of variations of pA values of all generations. Table 4 shows the pC values of C(n) of BrucellaSuis 1330 genome sequence and 500 generations of C(n) using rule 90 of 123CA. Fig. 5 shows the graph of the variations of pC values of all generations.



500 evolutions of A(n) using 123CA rule 90

500 evolutions of T(n) using 123CA rule 90

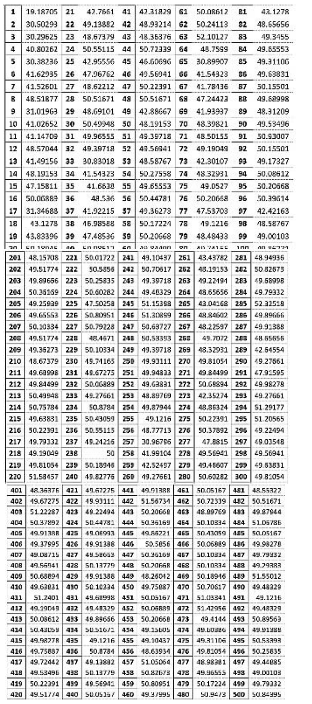


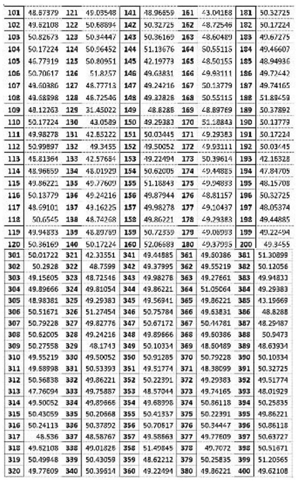
500 evolutions of G(n) using 123CA rule 90

500 evolutions of C(n) using 123CA rule 90

Fig. 1: Evolutions of the adjoints of A(n), T(n). G(n) and C(n).

Table 1: pA values of A(n) and its 500 evolutions





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.

Fig. 2: pA values of A(n) and of its evolutions

$A_e(n)$	pA
e = 1	30.50293
e = 2	30.29625
e = 4	30.38236
e = 8	31.01963
e = 16	31.34688
e = 32	30.83018
e = 64	30.89907
e = 128	31.45022
e = 256	30.96796

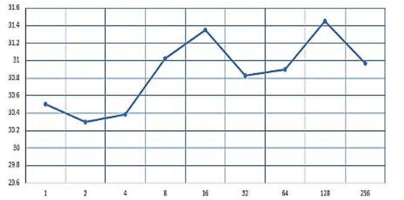
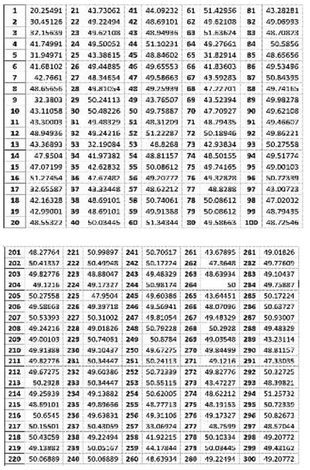
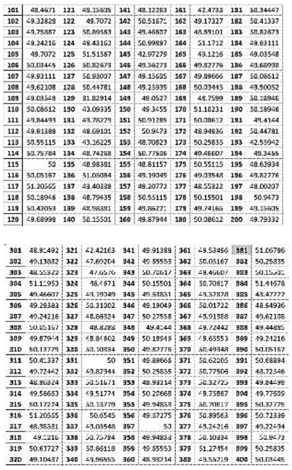
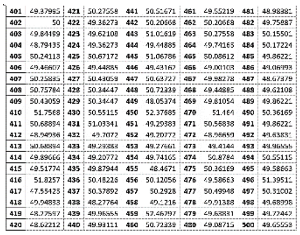


Fig. 3: Minimum pA values of A(n) and of its evolutions

Table 2: pT values of T(n) and its 500 evolutions







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.

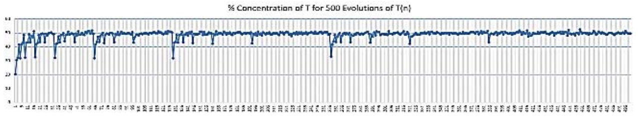


Fig. 4: pT values of T(n) and of its evolutions

$T_e(n)$	pT
e = 1	30.45126
e = 2	32.15639
e = 4	31.94971
e = 8	32.3803
e = 16	32.65587
e = 32	32.19084
e = 64	31.82914
e = 128	31.82914
e = 256	33.06924

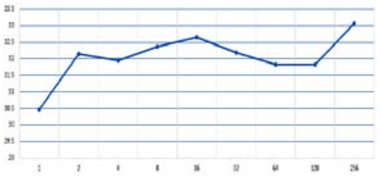
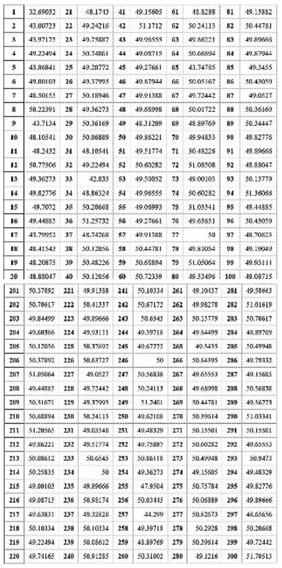
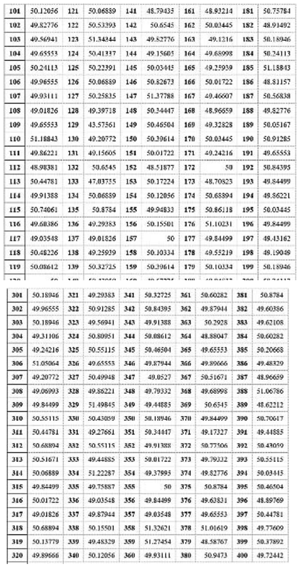
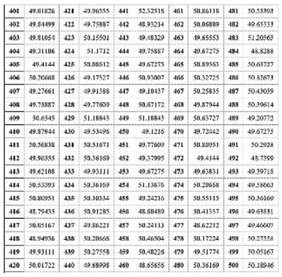


Fig. 5: Minimum pT values of T(n) and of its evolutions

Table 3: pG values of G(n) and its 500 evolutions







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.

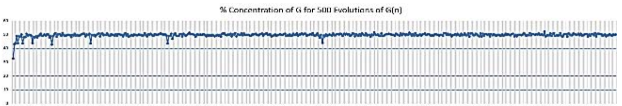


Fig. 6: pG values of G(n) and of its evolutions

$G_e(n)$	pG
e = 1	43.00723
e = 2	43.97175
e = 4	43.86841
e = 8	43.7134
e = 16	43.79952
e = 32	42.835
e = 64	43.74785
e = 128	43.57561
e = 256	44.299

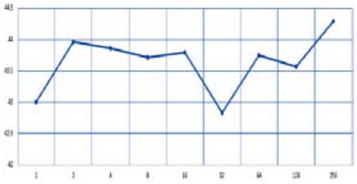


Fig. 7: Minimum pG values of G(n) and of its evolutions

Table 4: pC values of C(n) and its 500 evolutions

1	27.86772	21	46.69308	41	48.65656	61	48.31209	81	46.95143
2	40.7165	22	49.67275	42	50.86118	62	50.34447	82	50.31002
3	41.31932	23	49.08715	43	50.41337	63	50.9473	83	50.67172
4	48.55322	24	50.48226	44	49.87944	64	49.96555	84	49.53496
5	39.37306	25	47.22701	45	50	65	40.47537	85	49.03548
6	48.81157	26	49.37995	46	48.38099	66	47.20978	86	50.32725
7	48.50155	27	49.74165	47	50.34447	67	47.77816	87	49.29383
8	50.15501	28	50.79228	48	50.84395	68	49.10437	88	50.55115
9	40.69928	29	49.84499	49	47.57148	69	47.55425	89	50.63727
10	47.9504	30	49.94833	50	49.98278	70	50.12056	90	49.46607
11	46.98588	31	49.20772	51	50.49948	71	50.08612	91	50.80951
12	48.01929	32	49.3455	52	51.22287	72	50.01722	92	50.20668
13	47.96762	33	39.88977	53	50.36169	73	48.60489	93	50.6545
14	50.03445	34	48.01929	54	49.81054	74	50.36169	94	51.30899
15	49.68998	35	48.1743	55	49.94833	75	50.13779	95	50.41337
16	49.44885	36	50.08612	56	49.93111	76	49.3455	96	50.27558
17	39.32139	37	48.08818	57	49.3455	77	49.98278	97	47.2959
18	47.48536	38	50.20668	58	50.46504	78	48.94936	98	49.24216
19	47.12367	39	49.93111	59	50.70617	79	49.79332	99	49.86221
20	49.51774	40	49.48329	60	50.56838	80	50.17224	100	49.39718

101	49.81054	121	50.46504	141	49.77609	161	46.77919	181	51.01619
102	49.94833	122	50.13779	142	49.43162	162	49.68998	182	50.25835
103	50.15501	123	49.37995	143	50	163	50.37892	183	49.96555
104	50.62005	124	51.11953	144	49.29383	164	49.55219	184	50.05167
105	48.8288	125	49.31106	145	46.02136	165	49.94833	185	49.89666
106	50.05167	126	50	146	49.62108	166	51.22287	186	49.60386
107	49.74165	127	48.89769	147	49.98278	167	50.80951	187	49.43162
108	49.08715	128	50.43059	148	48.65656	168	50.17224	188	50.05167
109	50.32725	129	39.57975	149	49.96555	169	50.32725	189	48.7599
110	50.51671	130	48.58767	150	50.06889	170	51.63624	190	49.10437
111	50.96452	131	48.77713	151	49.51774	171	51.06786	191	50.06889
112	51.20565	132	50.82673	152	50.55115	172	50.12056	192	50.84395
113	50.39614	133	45.93524	153	50.75784	173	50.49948	193	48.65656
114	49.22494	134	49.44885	154	50.20668	174	50.49948	194	50.74061
115	49.29383	135	50.15501	155	50.22391	175	49.74165	195	50.06889
116	49.20772	136	49.3455	156	51.08508	176	49.43162	196	50.75784
117	49.79332	137	48.81157	157	50.77506	177	49.22494	197	49.31106
118	49.29383	138	49.89666	158	50.31002	178	48.03651	198	49.55219
119	50.74061	139	51.44678	159	49.81054	179	50.2928	199	50.74061
120	50.05167	140	50.79228	160	50.24113	180	50.03445	200	49.48329

201	49.39718	221	49.3455	241	48.88047	261	47.82983	281	48.86324
202	49.84499	222	51.25732	242	50.93007	262	50.67172	282	49.31106
203	49.7072	223	50.06889	243	50.80951	263	49.72442	283	50.9473
204	49.50052	224	48.81157	244	50.20668	264	50.43059	284	51.1712
205	50.68894	225	50.27558	245	51.67069	265	48.36376	285	49.82776
206	50.44781	226	49.08715	246	50.70617	266	51.2401	286	51.11953
207	49.00103	227	49.7072	247	50.8784	267	50.36169	287	49.68998
208	48.93214	228	50.10334	248	50.18946	268	49.48329	288	50.22391
209	48.91492	229	50.80951	249	50.79228	269	50.2928	289	46.53806
210	49.86221	230	50.36169	250	48.77713	270	50.25835	290	50.01722
211	50.20668	231	49.7072	251	50.70617	271	49.53496	291	50.80951
212	50.6545	232	49.56941	252	49.81054	272	50.39614	292	51.68791
213	49.53496	233	49.62108	253	48.58767	273	47.76094	293	48.31209
214	49.72442	234	49.15605	254	50	274	49.7072	294	50.03445
215	49.43162	235	50.53393	255	50.46504	275	49.68998	295	50.44781
216	49.55219	236	50	256	50.18946	276	49.51774	296	49.98278
217	50.03445	237	49.65553	257	40.95763	277	49.98278	297	49.46607
218	49.51774	238	49.53496	258	48.36376	278	49.72442	298	49.87944
219	50.75784	239	50.56838	259	48.31209	279	50.82673	299	49.31106
220	49.98278	240	49.94833	260	49.32828	280	49.74165	300	50.6545

301	49.22494	321	48.62212	341	49.27661	361	49.98278	381	50.56838
302	49.56941	322	50.05167	342	49.56941	362	49.65553	382	48.91492
303	50.20668	323	50.67172	343	50.77506	363	49.96555	383	49.96555
304	50.46504	324	50.98174	344	50.44781	364	49.67275	384	49.55219
305	48.7599	325	49.86221	345	49.56941	365	50.46504	385	47.10644
306	49.72442	326	50.24113	346	50.5856	366	49.48329	386	50.01722
307	49.86221	327	50.9473	347	50.86118	367	49.53496	387	49.81054
308	51.15398	328	49.82776	348	50.13779	368	48.96381	388	51.05064
309	49.00103	329	50.24113	349	49.08715	369	50.01722	389	49.36273
310	49.37995	330	49.96555	350	50.17224	370	49.44885	390	49.79332
311	49.96555	331	50.32725	351	51.20565	371	49.31106	391	50.67172
312	49.87944	332	50.9473	352	50.25835	372	49.84499	392	49.74165
313	50.13779	333	50.6545	353	50.5856	373	50.37892	393	49.4144
314	49.29383	334	50.36169	354	48.79435	374	49.65553	394	50.99897
315	49.86221	335	49.31106	355	50.18946	375	49.7072	395	50.75784
316	49.63831	336	50.6545	356	50.15501	376	51.87737	396	50
317	49.94833	337	48.39821	357	50.03445	377	50.49948	397	49.96555
318	51.5329	338	49.68998	358	50.01722	378	49.37995	398	51.37788
319	50.17224	339	50.08612	359	49.10437	379	50.80951	399	50.63727
320	50.74061	340	50.06889	360	50.91285	380	50.25835	400	49.17327

401	50.20668	421	49.7072	441	50.27558	461	50.32725	481	49.27661
402	50.03445	422	50.32725	442	50.08612	462	49.31106	482	50.79228
403	50.15501	423	49.51774	443	47.67482	463	49.13882	483	50.44781
404	49.75887	424	50.15501	444	50.43059	464	49.3455	484	48.62212
405	50.53393	425	49.89656	445	50.01722	465	50.03445	485	50.48226
406	49.74165	426	51.05064	446	49.22494	466	48.84602	486	48.57044
407	49.55219	427	48.20875	447	49.27661	467	50.62005	487	50.34447
408	49.81054	428	51.36056	448	50.56838	468	49.77609	488	50.03445
409	51.63624	429	49.13882	449	50.36169	469	49.94833	489	49.55219
410	50.13779	430	49.25939	450	49.50052	470	49.00103	490	49.65553
411	50.15501	431	49.62108	451	50.22391	471	48.55322	491	49.31106
412	49.37995	432	51.10231	452	49.06993	472	49.60386	492	50.25835
413	50.25835	433	49.01826	453	48.36376	473	48.67379	493	49.53496
414	49.93111	434	49.84499	454	49.3455	474	49.82776	494	49.39718
415	50.17224	435	49.58663	455	49.32828	475	49.84499	495	50.03445
416	50.27558	436	49.27661	456	48.81157	476	50.60282	496	49.77609
417	49.81054	437	49.06993	457	50.48226	477	50.06889	497	49.15605
418	49.63831	438	50.68894	458	49.3455	478	50.63727	498	49.68998
419	50.86118	439	50.60282	459	48.89769	479	50.53393	499	50.25835
420	49.98278	440	49.1216	460	49.62108	480	49.1216	500	50.13779

Rule number 90 is applied to C(n) and its 500 generations. It is observed that the pC 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))=40.71650 and Max(A(n))=40.95763. The deviation is 1.99

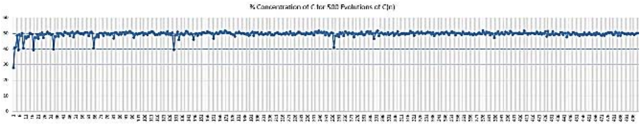


Fig. 8: pC values of C(n) and of its evolutions

$C_e(n)$	pC
e = 1	40.7165
e = 2	41.31932
e = 4	39.37306
e = 8	40.69928
e = 16	39.32139
e = 32	39.88977
e = 64	40.47537
e = 128	39.57975
e = 256	40.95763

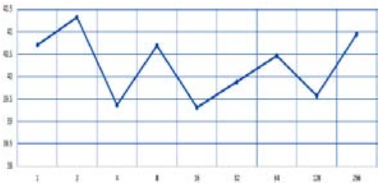


Fig. 9: Minimum pC values of C(n) and of its evolutions

V. CONCLUSIONS

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.

ACKNOWLEDGEMENTS

The authors express their profound gratitude to Mrs. G. Chitra, Managing Director of Pentagram Research Centre Private Limited, Hyderabad, India and Mrs. Prashanthi Govindarajan, CEO of PRC Global Technologies Inc, Ontario, Canada for their boundless support in carrying out research in their laboratories. The authors further put on record the invaluable moral and scientific support extended by Prof. Jean Claude Perez of France, Prof Dr. Michael Patrick Coyle of Brody Medical School, NC, USA, Prof. Dr. Andras Pellionisz of HolGenTech, California, USA, Mr. M. V. Ramanujam, CEO, Propinquity Sciences, Bangalore, India and data scientists of Pentagram Research 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/ nature 11247.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, CurrBiol 22(21):R898–R899.
8. Doolittle, W. Ford (2013). "Is junk DNA bunk? A critique of ENCODE". Proc Natl AcadSci USA110 (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 Genetics10 (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 EranElhaik (2013). "On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE" (PDF). Genome Biology and Evolution5 (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 Research21: 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". PNAS111 (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 GertonLunter (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 Surg33 (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/nature08451. PMC 2923221. PMID 19741700.
22. Nielsen H, Johansen SD; Johansen (2009). "Group I introns: Moving in new directions". RNA Biol6 (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: 1994 PNAS..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.; DavariEjtehadi, H.; Rylance, P.; Warren, P.; Martin, J.; Murray, PG. (Oct 2004). "Human endogenous retroviruses: transposable elements with potential?". Clin ExpImmunol 138 (1): 1–9. doi:10.1111/j.1365-2249.2004.02592.x.PMC 1809191. PMID 15373898.