International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-3, January 2020

Prediction of Genome Sequences in Terms of Cellular Automata Expansion of Rule Based Logics

Rama Naga Kiran Kumar.K, Ramesh Babu. I

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. The analysis of genome sequences with predictive analytics gives a scope of getting the inherent properties of the genome. The predictive model suits the Nucleotide concentration and 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 that whether it is possible to categorize a set of genomes like the human genome repository. The usage of predictive analytics such as recommender systems provides the possible sequence of genome and leads to the evolution of similar models in the processing of genomes.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 thispaper 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 Brucella Suis 1330 Genome Sequence.

Revised Manuscript Received on January 30, 2020.
* Correspondence Author

Rama Naga Kiran Kumar. K*, Research Scholar, Dept of Computer Science & Engineering, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India. E-Mail: karasala_be@hotmail.com

Prof. Ramesh Babu. I, Professor, Dept of Computer Science & Engineering, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.E-Mail: rinampudi@outlook.com.

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 ADJOINTS OF 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 Brucella Suis 1330 for a case study. The actual length of the genome sequence of Brucella Suis 1330 is 5806. 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 ξ_{i-1}, ξ_i, and ξ_{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 ξ^t_{i-1}, ξ^t_i, and ξ^t_{i+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 basically 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
$\overline{(\xi_i}\bar{\xi}_{i+1})$	17
$(\xi_{i-1}\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\bar{\xi}_i\xi_{i+1})$	18
$\overline{(\xi_i}\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\bar{\xi}_i)$	19
$(\xi_{i-1}\bar{\xi}_i\bar{\xi}_{i+1}) + (\bar{\xi}_{i-1}\xi_i\bar{\xi}_{i+1})$	20

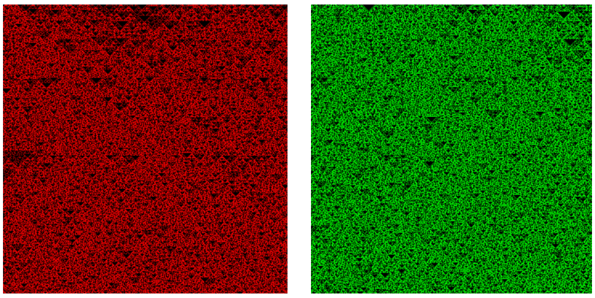
For the case study rule number 90 is applied to the adjoints of Brucella Suis 1330 genome sequence and 500 evolutions generated. Rule 90 is shown below.

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

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

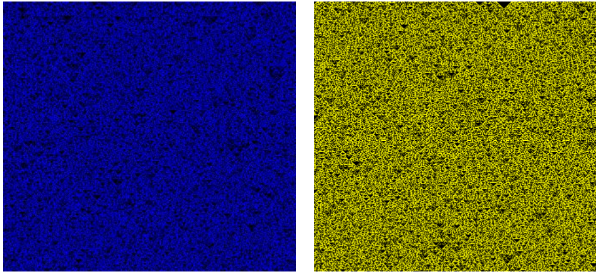
IV. CONCENTRATION OF NUCLEOTIDES IN ADJOINTS OF BRUCELLA SUIS 1330 GENOME SEQUENCE

The values of pA, pT, pG and pC of the Brucella Suis 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 Brucella Suis 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 Brucella Suis 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 Brucella Suis 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 Brucella Suis 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

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	47.96762	46	49.56941	66	41.54323	86	49.63831
7	41.52601	27	48.62212	47	50.22391	67	41.78436	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
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
101	48.67379	121	49.03548	141	48.96659	161	43.04168	181	50.32725
102	49.62108	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	42.19773	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	49.68998	128	48.72546	148	49.32828	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	49.06993	199	49.22494
120	50.36169	140	50.17224	160	52.06683	180	49.37995	200	49.3455

Table 1 continues…

201	48.15708	221	50.01722	241	49.10437	261	43.43782	281	48.94936
202	49.51774	222	50.5856	242	50.70617	262	48.19153	282	50.82673
203	49.89666	223	50.25835	243	49.39718	263	49.22494	283	49.68998
204	50.36169	224	50.60282	244	49.48329	264	48.65656	284	49.79332
205	49.25939	225	47.50258	245	51.15398	265	43.04168	285	52.32518
206	49.65553	226	50.80951	246	51.30899	266	48.84602	286	49.89666
207	50.10334	227	50.79228	247	50.63727	267	48.22597	287	49.91388
208	49.51774	228	48.4671	248	50.53393	268	49.7072	288	48.65656
209	49.36273	229	50.10334	249	49.39718	269	48.32931	289	42.64554
210	48.67379	230	49.74165	250	49.93111	270	49.81054	290	49.27661
211	49.68998	231	49.67275	251	49.94833	271	49.84499	291	47.91595
212	49.84499	232	50.06889	252	49.63831	272	50.68894	292	49.98278
213	50.49948	233	49.27661	253	48.89769	273	42.35274	293	49.27661
214	50.75784	234	50.8784	254	49.87944	274	48.86324	294	51.29177
215	49.63831	235	50.43059	255	49.1216	275	50.22391	295	51.20565
216	50.22391	236	50.55115	256	48.77713	276	50.37892	296	49.22494
217	49.79332	237	49.24216	257	30.96796	277	47.8815	297	49.03548
218	49.19049	238	50	258	41.99104	278	49.56941	298	49.56941
219	49.81054	239	50.18946	259	42.52497	279	49.46607	299	49.63831
220	51.58457	240	49.82776	260	49.27661	280	50.60282	300	49.81054
301	50.01722	321	42.33551	341	49.44885	361	49.60386	381	51.30899
302	50.2928	322	48.7599	342	49.37995	362	49.55219	382	50.12056
303	49.15605	323	48.72546	343	49.98278	363	49.27661	383	49.94833
304	49.89666	324	49.81054	344	49.86221	364	51.05064	384	49.29383
305	48.98381	325	49.29383	345	49.56941	365	49.86221	385	43.19569
306	50.51671	326	51.27454	346	50.75784	366	49.63831	386	48.8288
307	50.79228	327	49.82776	347	50.67172	367	50.44781	387	48.29487
308	50.62005	328	49.24216	348	49.89666	368	49.60386	388	50.9473
309	50.27558	329	48.1743	349	50.10334	369	48.60489	389	48.63934
310	49.55219	330	49.50052	350	50.91285	370	50.79228	390	50.10334
311	49.68998	331	50.53393	351	49.51774	371	48.38099	391	50.32725
312	50.56838	332	49.86221	352	50.22391	372	49.29383	392	49.51774
313	47.76094	333	49.75887	353	48.57044	373	49.74165	393	48.01929
314	49.50052	334	49.89666	354	49.68998	374	50.86118	394	50.25835
315	50.43059	335	50.20668	355	50.41337	375	50.22391	395	49.86221
316	50.24113	336	50.37892	356	50.70617	376	50.34447	396	50.86118
317	48.536	337	48.58767	357	49.58663	377	49.77609	397	50.63727
318	49.62108	338	49.01826	358	51.49845	378	49.7072	398	50.51671
319	50.49948	339	50.43059	359	48.62212	379	50.25835	399	51.20565
320	49.77609	340	50.39614	360	49.22494	380	49.86221	400	49.62108

Table 1 continues…

401	48.36376	421	49.67275	441	49.91388	461	50.05167	481	48.55322
402	49.67275	422	49.93111	442	51.56734	462	50.72339	482	50.51671
403	51.22287	423	49.22494	443	50.20668	463	48.89769	483	49.87944
404	50.37892	424	50.44781	444	50.36169	464	50.10334	484	51.06786
405	49.91388	425	49.06993	445	49.86221	465	50.43059	485	50.05167
406	49.37995	426	49.91388	446	50.5856	466	50.06889	486	49.98278
407	49.08715	427	49.58663	447	50.36169	467	50.10334	487	49.79332
408	49.56941	428	50.13779	448	50.20668	468	50.10334	488	49.29383
409	50.68894	429	49.91388	449	48.26042	469	50.18946	489	51.55012
410	49.63831	430	50.10334	450	49.75887	470	50.70617	490	49.48329
411	51.2401	431	49.68998	451	50.05167	471	51.03341	491	49.1216
412	49.19049	432	49.48329	452	50.06889	472	51.42956	492	49.48329
413	50.08612	433	49.89666	453	50.20668	473	49.4144	493	50.89569
414	50.43059	434	50.51671	454	49.15605	474	49.60386	494	49.91388
415	49.98278	435	49.1216	455	49.10437	475	49.31106	495	50.53393
416	49.75887	436	50.8784	456	48.63934	476	49.81054	496	50.25835
417	49.72442	437	49.13882	457	51.05064	477	48.98381	497	49.44885
418	49.53496	438	50.13779	458	50.82673	478	49.96555	498	49.00103
419	50.22391	439	49.56941	459	50.80951	479	50.17224	499	49.79332
420	49.51774	440	50.05167	460	49.37995	480	50.9473	500	50.84395

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

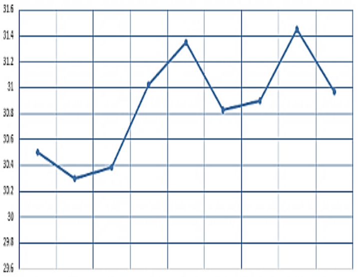
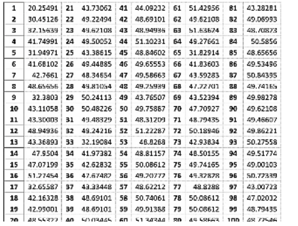


Fig. 3: Minimum pA values of A(n) and of its evolutions

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



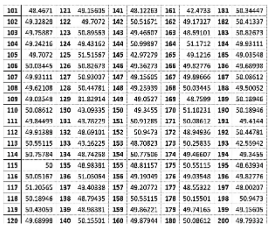
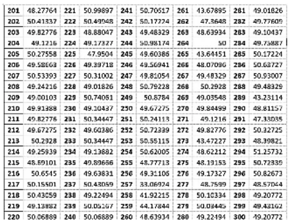


Table 2 continues…



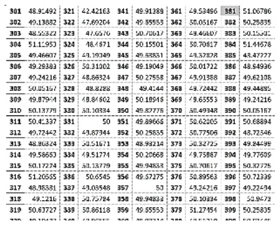
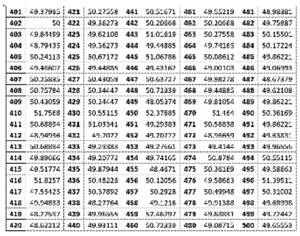


Table 2 continues…



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.

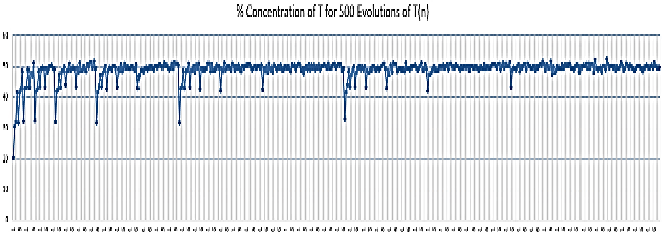


Fig. 4: pT values of T(n) and of its evolutions

$T_e(n)$	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

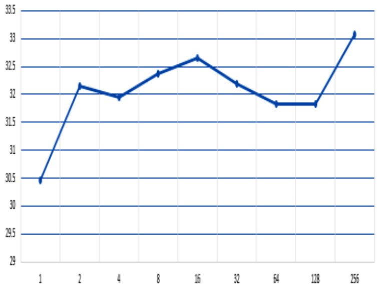
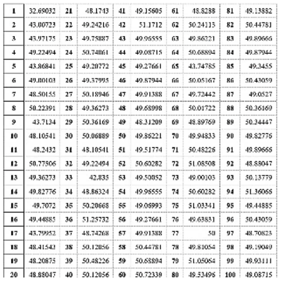


Fig. 5: Minimum pT values of T(n) and of its evolutions

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



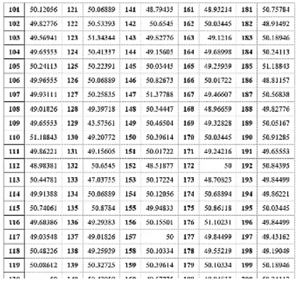
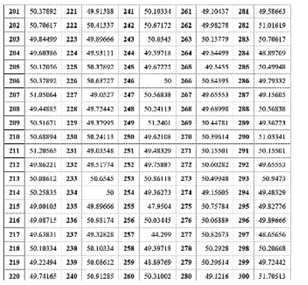


Table 3 continues…



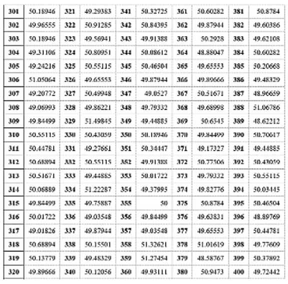
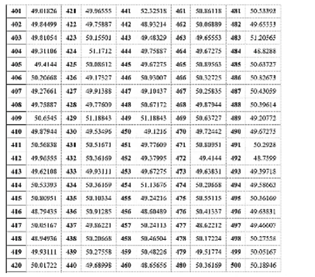


Table 3 continues…



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.

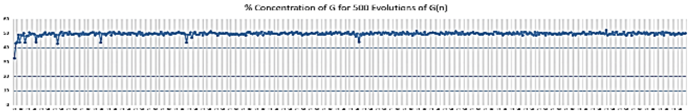


Fig. 6: pG values of G(n) and of its evolutions

$G_e(n)$	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

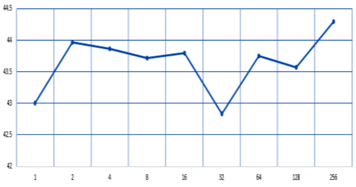
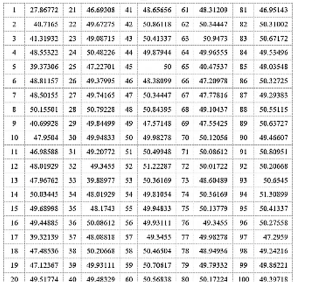


Fig. 7: Minimum pG values of G(n) and of its evolutions

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



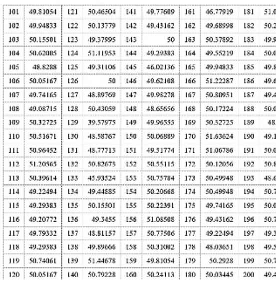
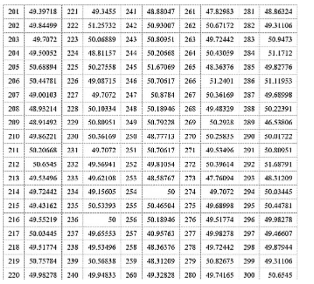
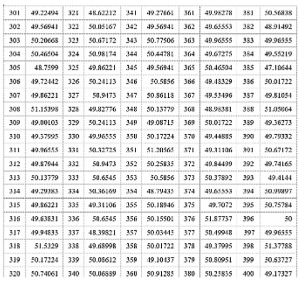


Table 4 continues…

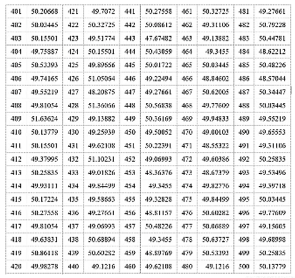


201	49.39718	221	49.3455	241	48.88047	261	47.82983	281	48.86324
202	49.84499	222	51.25732	242	50.93007	262	50.67172	282	49.31106
203	49.7072	223	50.06889	243	50.80951	263	49.72442	283	50.9473
204	49.50052	224	48.81157	244	50.20668	264	50.43059	284	51.1712
205	50.68894	225	50.27558	245	51.67069	265	48.36376	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	48.98381	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

Table 4 continues…



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	51.36066	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

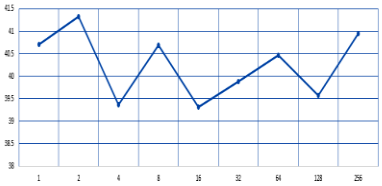


Fig. 9: Minimum pC values of C(n) and of its evolutions

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 the predictive method allows us to guess the pattern of sequences which helps the formulation of genome sequences. In short, the concept of "**Percentage Nucleotide Concentration** (**PNC**)" introduced in this paper seems to show a way to accomplish this task.

REFERENCES

1. Carey, Nessa (2015). Junk DNA: A Journey Through the Dark Matter of the Genome. Columbia University Press. ISBN 9780231170840.
2. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. International Human Genome Sequencing Consortium (February 2001). "Initial sequencing and analysis of the human genome". Nature **409** (6822): 879–888.
11. Bibcode:2001Natur.409..860L. doi:10.1038/35057062. PMID 11237011.

AUTHORS PROFILE

K Rama Naga Kiran Kumar completed his B.E(CSE), M.Tech(CSE) and pursuing Ph.D. (CSE) from Acharya Nagarjuna University. His area of interest is in AI, Data Mining, Machine Learning, Software Engineering.

Dr. Ramesh Babu I, done his B.E (ECE), M.E (Computer Engg), Ph.D. (CSE) in ANU. His area of interest is in Image Processing, Computer Graphics, AI, Data Mining, Machine Learning, Software Engineering