

Chaotic Analysis of DNA Codes

Jiguo Dong¹, Takako Yamada² and Katsufusa Shono

¹ Graduate School of Information Systems, The University of Electro-Communications

1-5-1, Chofu-ga-oka, Chofu-shi, Tokyo, 182-8585 Japan

² School of Policy Studies, KWANSEI GAKUIN University

2-1, Gakuen, Sanda-shi, Hyougo, 669-1337 Japan

E-mail: ¹ jiguo@tanaka.is.uec.ac.jp, ² takakoyamada@ksc.kwansei.ac.jp

Abstract: By synthesizing the logistic map $f(x_t) = 4x_t(1-x_t)$, $x_{t+1} = f(f(x_t))$ produces chaos having $L = 4$ where Lyapunov exponent is $\lambda = \ln L$. The successive backward calculation $f^{-1}(x_{t+1}) = (1 \pm \sqrt{1-x_{t+1}})/2$, $x_t = f^{-1}(f^{-1}(x_{t+1}))$ can accept external four bit codes such as DNA(A,G,T,C) for the sign determination, the internal state x_t obtained gives us the Lyapunov exponent L' along the external codes and can be compared with $L = 4$ chaos. The Lyapunov exponent L' obtained is a measure of evolution of a gene, and the entropy G_2 gives characteristic distributions.

In the nature, we can find many time series, such as blood pressure and pulse rate, weather data, language files, et al. DNA(A,G,C,T) codes of a gene are an example of such data, where we want to find map functions representing the time series in the nature.

By synthesizing the logistic map $f(x_t) = 4x_t(1-x_t)$, the internal state $x_{t+1} = f(f(x_t))$ produces chaos having $L = 4$ where Lyapunov exponent is $\lambda = \ln L$. For the observation of the internal state x_t , we employ the isomorphic transform and quantization that can be represented as $Y_{t,n} = [(2/\pi)\arcsin(\sqrt{x_t})4^n]$, where $[]$ eliminates the small value and n is a quantization resolution[1]. The integer time series $Y_{t,1}$ for $n=1$ has a random combination of 0,1,2,3, whose distributions are kept to be uniform. The two bit codes of $Y_{t,1}$ acts as a sign determination for the successive backward calculation

$$f^{-1}(x_{t+1}) = (1 \pm \sqrt{1-x_{t+1}})/2, x_t = f^{-1}(f^{-1}(x_{t+1})).$$

The forward and backward calculation of the synthesized logistic map was performed under a fixed point calculation of 128 bit. A part of the results from $t = 0$ to $t = 79$, where t is a discrete time, is shown in Table 1. Starting from an initial value x_0 , at a top on the left hand side of Table 1, successive forward calculations give us the internal state $\{x_t\}$ of 128 bit. The results of the isomorphic transform and

quantization observation employing $Y_{t,n} = [(2/\pi)\arcsin(\sqrt{x_t})4^n]$ are shown for $Y_{t,4}$ ($0 \sim 255$, $n = 4$), $Y_{t,3}$ ($0 \sim 63$, $n = 3$), $Y_{t,2}$ ($0 \sim 16$, $n = 2$), $Y_{t,1}$ ($0 \sim 3$, $n = 1$). Starting from the internal state x_{79} as an initial value for the backward calculation, from the bottom on the right hand side of Table 1, the internal state $\{x_t\}$ was successively obtained along with $Y_{t,1}$ (0,1,2,3) for the sign determination, that was obtained in the forward calculation in advance, and the quantization observation was also given as $Y_{t,4}$, $Y_{t,3}$, $Y_{t,2}$, $Y_{t,1}$. Looking at Table 1 of the right and left hand side, we found that the forward and backward calculation is reversible and the closed loop was found at any range of the discrete time. Chaos states, that can be determined starting from an initial value x_0 to just before falling down into a periodic state, are as long as about $2^{64} \approx 10^{19}$ steps of the discrete time in the fixed point calculation of 128 bit. The initial value x_0 of 128 bit can be designated as a compression code for the chaos time series $\{x_t\}$. This is a chaos where such a compression can be always achieved within the chaotic time series $\{x_t\}$ realized in a digital computer.

In the step of backward calculation on the synthesized logistic map that produces $L = 4$ chaos, instead of employing $Y_{t,1}$ codes for the sign

determination, we can use four codes of DNA[2]. Along the DNA codes we can determine the internal state $\{x_t\}$ characterized by the DNA codes and we can calculate Lyapunov exponent $L' = \sum \log_2 \{\Delta x_{t+1} / \Delta x_t\}$,

$\Delta x_{t+1} = 2^{-20}$ that is very close to but a little bit smaller than $L = 4$ (integer). The Lyapunov exponent L' can be designated as an evolution degree of DNA. On the other hand, it is well known that the unique distribution of the DNA codes are required by the characteristic chemical behavior of the codes. They are numerically determined by calculating the entropy

$$G_2 = -p(A)\log_2 p(A) - p(G)\log_2 p(G) - p(C)\log_2 p(C) - p(T)\log_2 p(T)$$

In the physics(thermo dynamics), it is well known that the entropy always increases, but in the field of

information the entropy decreases in the case of the evolution of DNA and the compression of a digital file. The relation between the Lyapunov exponent L' and the entropy G_2 is shown in Fig. 1 in the case of bacteria. At a left upper corner, an advanced example(AJ628144) is located, where the Lyapunov exponent L' is very close to 4(integer) and the entropy G_2 is kept to be small, and at a lower right corner we found a primitive DNA(AY522431), where the Lyapunov exponent L' is even small as $L'=3.962$ and the entropy G_2 is very close to $G_2=1.999...$ showing random distribution. In addition to them, $L=4$ chaos is given at a upper right corner. To succeed the simulation of DNA codes, the synthesized map function must satisfy the Lyapunov exponent L' , that is a little bit smaller than 4(integer), and the entropy G_2 , that represents the characteristic distribution of the four DNA codes. The map function, which must be normalized in the range of $0 \sim 1$ in the internal state x_i , will be a combination of the extended logistic map of $x_{i+1} = ((m+1)^{m+1} / m^m)x_i(1-x_i)^m$ and the three dimensional map of $x_{i+1} = P_3x_i^3 + P_2x_i^2 + P_1x_i + P_0$, where $P_3 = 16P^3 / 27$, $P_2 = -8P^3 / 9$, $P_1 = 2P^3 / 3 - P$, $P_0 = P / 2 - 2P^3 / 27 - 0.5$ and $P = 3\sqrt{3} / 2 + \alpha$ are parameters. The parameter m and α must be determined as to satisfy the Lyapunov exponent L' and the entropy G_2 [3, 4].

DNA has four codes of A,G,C,T, but their functional actions are performed through three combinations of A,G,C,T and correlation between the three combinations and amino acids are known as DNA cipher codes. Table 2 gives the relationship between the quantized state $Y_{t,3}$ (0~63) and the amino acid codes(0~20) are characteristic for the evolution degree L' . Bacteria AY522431 has uniform distribution, where the entropy is $G_2=1.999...$, and, on the other hand, bacteria AJ628149 has strongly distorted one for amino acid codes of 3, 8, 9, and 12, where the entropy is $G_2=1.88...$. These distortion would be a result of evolution.

By accepting DNA codes in the backward calculation of $L=4$ chaos and by the isomorphic transform and quantization, the quantized state $Y_{t,3}$ (0~63) and DNA code $\{Y_{t,1}\}_3$ give fractal relationship, as shown in Fig. 2. Nonoverlapping and one-to-one correspondence are observed. Local behavior of DNA horizontal transition can be

$Y_{t,3}$	$\{Y_{t,1}\}_3$	$\{Y_{t,1}\}_3$ code	DNAcode	amino acid	
0	0	000	AAA	Lys	0
1	1	001	AAG	Lys	0
2	2	002	AAC	Asp	1
3	3	003	AAT	Asp	1
4	7	013	AGT	Ser	2
5	6	012	AGC	Ser	2
6	5	011	AGG	Arg	3
7	4	010	AGA	Arg	3
8	8	020	ACA	Thr	4
9	9	021	ACG	Thr	4
10	10	022	ACC	Thr	4
11	11	023	ACT	Thr	4
12	15	033	ATT	Ile	5
13	14	032	ATC	Ile	5
14	13	031	ATG	Met	6
15	12	030	ATA	Ile	5
16	28	130	GTA	Val	7
17	29	131	GTG	Val	7
18	30	132	GTC	Val	7
19	31	133	GTT	Val	7
20	27	123	GCT	Ala	8
21	26	122	GCC	Ala	8
22	25	121	GCG	Ala	8
23	24	120	GCA	Ala	8
24	20	110	GGA	Gly	9
25	21	111	GGG	Gly	9
26	22	112	GGC	Gly	9
27	23	113	GGT	Gly	9
28	19	103	GAT	Asp	1
29	18	102	GAC	Asp	1
30	17	101	GAG	Glu	10
31	16	100	GAA	Glu	10
32	32	200	CAA	Glu	10
33	33	201	CAG	Glu	10
34	34	202	CAC	His	11
35	35	203	CAT	His	11
36	39	213	CGT	Arg	3
37	38	212	CGC	Arg	3
38	37	211	CGG	Arg	3
39	36	210	CGA	Arg	3
40	40	220	CCA	Pro	12
41	41	221	CCG	Pro	12
42	42	222	CCC	Pro	12
43	43	223	CCT	Pro	12
44	47	233	CTT	Leu	13
45	46	232	CTC	Leu	13
46	45	231	CTG	Leu	13
47	44	230	CTA	Leu	13
48	60	330	TTA	Leu	13
49	61	331	TTG	Leu	13
50	62	332	TTC	Phe	14
51	63	333	TTT	Phe	14
52	59	323	TCT	Ser	2
53	58	322	TCC	Ser	2
54	57	321	TCG	Ser	2
55	56	320	TCA	Ser	2
56	52	310	TGA	---	18
57	53	311	TGG	Trp	15
58	54	312	TGC	Cys	16
59	55	313	TGT	Cys	16
60	51	303	TAT	Tyr	17
61	50	302	TAC	Tyr	17
62	49	301	TAG	---	19
63	48	300	TAA	---	20

Table 2. Functional relation between the time series $Y_{t,3}$ (0 ~ 63) and amino acid codes (0 ~ 20).

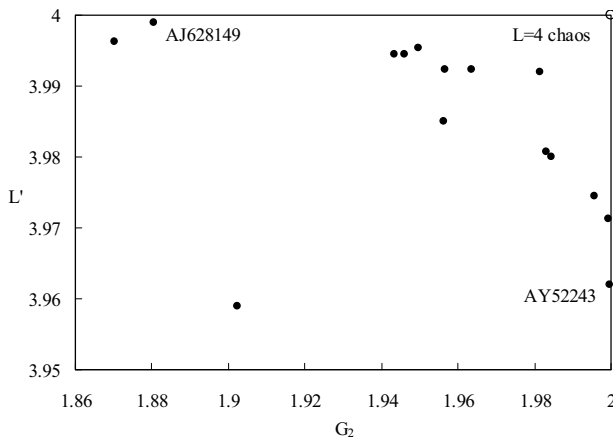


Fig. 1. Relation between evolution degree L' and entropy G_2 of bacteria.

visualized on the fractal figure. Relation of degenerated fractal points, as was shown in Table 2, of 0~20 amino acid codes are marked on the figure. Horizontal transition of DNA codes can be visualized by superimposing the trajectories. As an example, O157 and O168 DNA codes are superimposed from end(20) to next end(20). By repeating such transition and reversible processes, the step-by-step transition would happen in DNA.

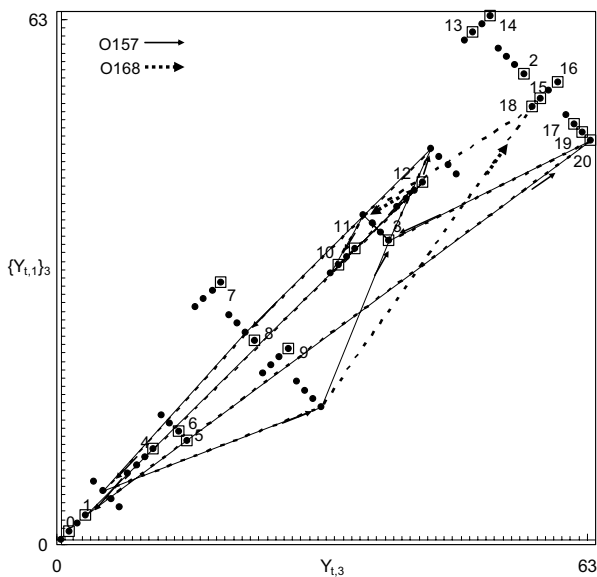


Fig. 2. Quantized fractal relation between the state $Y_{t,3}$ and the code $\{Y_{t,1}\}_3$ that are specified for amino acid codes(0~20). Superimposed trajectories of a part of O157 and O168 are shown.

On an extension of our study, we have observed languages of a digital file, that are stored in the memory of digital computer as a combination of four bit codes(0~15). English(New York Times) gives us

the Lyapunov exponent $L'=16.000\dots$ and the entropy $G_4=3.32\dots$. Japanese(Asahi Shinbun) gives us the Lyapunov exponent $L'=16.000\dots$ and the entropy $G_4=3.66\dots$. English is more in order than Japanese is. Human language is showing a good result of the evolution.

References

[1] K.Shono, "Chaos Engineering", Springer Verlag Tokyo, 2002.11.
 [2] J.Dong, T.Yamada, K.Shono, "Analysis of a Gene by the Use of Chaos having $L=4$ ", IEICE Technical Report NLP 2008, pp.27-32, 2008.3.
 [3] J.Dong, T.Yamada, K.Shono, "Nonlinear Quantization Analysis of Chaos produced by the Third Order Mapping Functions and their Predictability", IEICE Technical Report CAS 2007, pp.7-12, 2007.10.
 [4] J.Dong, T.Yamada, K.Shono, "Nonlinear Quantization Analysis of Chaos produced by the (m+1) Order Mapping Functions and their Predictability", IEICE Technical Report NLP 2007, pp.21-26, 2007.12.