

## Chaotic Analysis of DNA Codes

Jiguo Dong<sup>1</sup>, Takako Yamada<sup>2</sup> and Katsufusa Shono

<sup>1</sup> Graduate School of Information Systems, The University of Electro-Communications

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

<sup>2</sup> School of Policy Studies, KWANSEI GAKUIN University

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

E-mail: <sup>1</sup> jiguo@tanaka.is.uec.ac.jp, <sup>2</sup> 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

### forward calculation

### backward calculation

t	$x_t$				$Y_{t,4}$	$Y_{t,3}$	$Y_{t,2}$	$Y_{t,1}$
0	145c8ae2	dd03acea	dd1a5886	9c17f73b	46	11	2	0
1	d40e6dad	6da84adb	164fe288	297aa309	186	46	11	2
2	fb28ad90	ed08d5e9	7b484f1f	bd899480	233	58	14	3
3	465a7177	a8430a39	43b570aa	e4650c10	89	22	5	1
4	a591d536	bf5fba9b	b322bc5d	8e1e1412	152	38	9	2
5	509e16d8	274d23a8	ec29b6ff	ac80b622	97	24	6	1
6	79173cd0	d8397ab9	837797fc	fa0c6538	123	30	7	1
7	02f992d1	cb5c9bc0	7cf33dcd	2ecbf843	17	4	1	0
8	2ce2434e	1c540c79	855e2533	461f323e	70	17	4	1
9	f9b6ba19	1f015b95	167b60d7	d8332be5	230	57	14	3
10	58b5b89c	9d2db4ed	bd4b4add	fd8cbf4a	102	25	6	1
11	5764400e	d3710748	0933aff3	d0bec3a6	101	25	6	1
12	5cb110fe	c2e1ede6	c263121d	b3df8ee8	105	26	6	1
13	47fd133e	cf8d6279	457e2587	199c9722	91	22	5	1
14	9e88a2c6	4ca5567d	91ac7bcc	afd406ce	147	36	9	2
15	36f41d45	0efc2a40	b1f382d7	b6eb2077	78	19	4	1
16	e0e136d3	768542a4	3991bb6e	b75cf363	197	49	12	3
17	fa908349	57bb0eb1	1614934f	503309c6	232	58	14	3
18	4e0baacc	bc126a11	77b00dfc	e6e8410d	95	23	5	1
19	8435d23e	fb0fdb7e	c5bd20b8	714e892d	130	32	8	2
20	011b4f9d	6433aaa3	ac571dd7	70090045	10	2	0	0
21	1153ac1a	b338932e	acbcc059	86284276	42	10	2	0
22	c13a1a0f	ae123003	ddac3333	88fec7fa	171	42	10	2
23	ca4dbe57	2f866baa	12e8e5d8	bf64ff7f	174	43	10	2
24	d2957f44	d4f2113b	4f1c84fb	665ae59d	185	46	11	2
25	f8d20943	474ad953	05727b39	71a3e2ec	228	57	14	3
26	637a3199	6607326d	8c588435	dbe32a54	109	27	6	1
27	30525a3a	b54292c7	5ef87457	5548f049	73	18	4	1
28	f309a3b5	398e0a07	1453b925	66a80fde	219	54	13	3
29	9f096357	9f701554	d0b3bc17	47beee8b	147	36	9	2
30	38aa3fc5	77c2e8d7	ac2a650c	e199f486	79	19	4	1
31	db431ad2	5b0858d9	435d4cad	3b910c64	192	48	12	3
32	ffedbb14	ace1884c	23cc9c04	454f6caa	253	63	15	3
33	0123e673	f68cea2a	3f63073d	9d43128c	10	2	0	0
34	11d721ca	0c0e2fea	cfabd257	cc02bb07	43	10	2	0
35	c4b0030d	de7be4cc	3701f78f	0f9aa460	174	43	10	2
36	d1f60a9f	cdb473fc	276f155c	663f9e01	184	46	11	2
37	f7b52844	7c69fafd	662cb812	f208367c	226	56	14	3
38	7048be2e	db110538	54fa4633	9cf4b4a0	117	29	7	1
39	0f33e5dd	e721ee15	b313f18b	6223d6a5	40	10	2	0
40	b1ad3812	967098d1	441d7940	188c9041	160	40	10	2
41	830124b6	e6a39cc0	075a6399	79549589	129	32	8	2
42	0090597a	90a43878	43352b2c	527a9d1e	7	1	0	0
43	08ec3f16	473e9782	71dbd1ac	083fc4ce	30	7	1	0
44	773f9854	d7e50771	8e3d7bac	7ad0782f	122	30	7	1
45	04c3b6e2	36e0d913	71a459db	b82042fa	22	5	1	0
46	4558f627	d5d4292f	d2482a29	f8e191ed	89	22	5	1
47	a9dcd159	e0031e8a	246e6ea4	f5fb80d9	155	38	9	2
48	61d0688b	2227fd1a	531ab201	a6f9bc80	108	27	6	1
49	35c8373e	2ad6156a	53289659	cd592360	77	19	4	1
50	e4868b87	8f970afd	1ba8fbce	4d43adee	201	50	12	3
51	f208bf20	47e8b9ec	0c6b32b3	ff8db619	217	54	13	3
52	a7ad529c	d6670cb3	644f61a7	b3cc51c7	153	38	9	2
53	58f00ca0	a6f65449	6e2ab68c	111960fb	102	25	6	1
54	567c2137	97ba9950	8f5fac00	1afa7d38	101	25	6	1
55	60633b38	9345eb26	a02ae3ad	52e3b53e	107	26	6	1
56	3aa6333a	38d5e70b	11f53e98	9e0c5686	81	20	5	1
57	d45b0d87	f1a9e68c	7298513a	a65449f4	186	46	11	2
58	fb951b67	a37a91c1	7593b0f7	afa219d7	234	58	14	3
59	40bfd6b8	360eda60	223f3931	9d35f30a	85	21	5	1
60	bccf41e8	bad50e13	3aadf712	4c32f6c2	168	42	10	2
61	b3af184b	64a0d3f0	0268cabd	4e62d954	161	40	10	2
62	8bb78ac	9c5fadae	200d9434	f76cb70e	135	33	8	2
63	0887d9d7	4f1b6b36	5b6e8ad7	8fb3317e	29	7	1	0
64	72f10ee1	dc62d97f	32268b0d	5ea9c1a4	119	29	7	1
65	0a8bf06c	1abb0728	681b299d	13887ad9	33	8	2	0
66	883afa8b	93133d82	e261140e	8a48a912	133	33	8	2
67	043758cc	9b2dbb60	2cf93eff	2f27971e	20	5	1	0
68	3e0c9bee	30f91fea	3aef6173	141e2d12	83	20	5	1
69	c7ad4015	90a043e6	b647f00e	1f783c9e	176	44	11	2
70	dc690bb4	27e72343	a8959177	69a9a9e9	193	48	12	3
71	ff9f565	3ffae34c	7f84882d	b38a094d	249	62	15	3
72	074fb3e7	51375305	2442a7b5	0ccbaff3	27	6	1	0
73	65076a67	92cdf82c	3d712d26	14ef338f	110	27	6	1
74	2b723951	adb6fa8a	e51a7610	b4252d9f	69	17	4	1
75	fbda5ad2	91bb909d	e3a52376	3cc771ea	235	58	14	3
76	3d1dda32	46af6239	429e5eb5	2ccb0c46	83	20	5	1
77	cb400a4c	f2096db6	c6e59cdd	3a8f2399	179	44	11	2
78	e797f455	cf5140ab	5aea9289	4a481519	204	51	12	3
79	e7657433	3e680c75	b02bcc1b	2b874984	204	51	12	3

**Table 1.** Forward and backward calculation of  $L = 4$  chaos and isomorphic transform and quantization

$Y_{t,4}$ ,  $Y_{t,3}$ ,  $Y_{t,2}$  and  $Y_{t,1}$ .

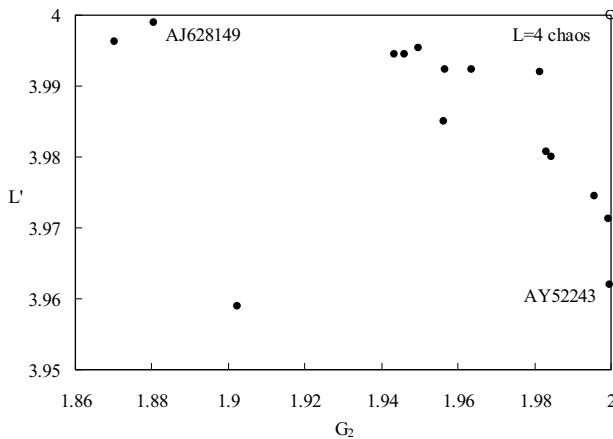
information the entropy decreases in the case of the evolution of DNA and the compression of a digital file. The relation between the Lyapunov exponent  $L'$  and the entropy  $G_2$  is shown in Fig. 1 in the case of bacteria. At a left upper corner, an advanced example(AJ628144) is located, where the 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  $\alpha$  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

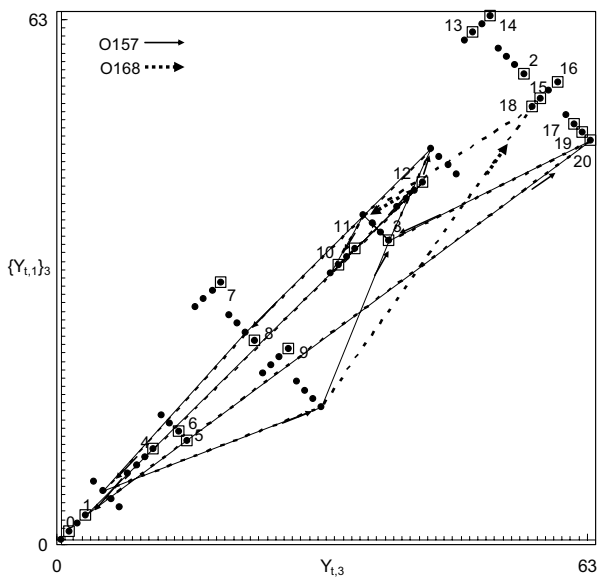
$Y_{t,3}$	$\{Y_{t,1}\}_3$	$\{Y_{t,1}\}_3$ code	DNAcode	amino acid	
0	0	000	AAA	Lys	0
1	1	001	AAG	Lys	0
2	2	002	AAC	Asp	1
3	3	003	AAT	Asp	1
4	7	013	AGT	Ser	2
5	6	012	AGC	Ser	2
6	5	011	AGG	Arg	3
7	4	010	AGA	Arg	3
8	8	020	ACA	Thr	4
9	9	021	ACG	Thr	4
10	10	022	ACC	Thr	4
11	11	023	ACT	Thr	4
12	15	033	ATT	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.