Dokkyo Journal of Medical Sciences
37(2) : 139～142, 2010

Short Communication

Evaluation of Chargaff's Parity Rules Using Simulation Analysis

Yoshifumi Ebara[1], Takashi Koge[2] and Kenji Sorimachi[3]

[1]*Terao Child Clinic, Terao, Takasaki, Gunma 370-0865, Japan,*
[2]*Department of Pharmacology, and* [3]*Educational Support Center, Dokkyo Medical University School of Medicine, Mibu, Tochigi 321-0293, Japan*

Key Words : Chargaff's second parity rule

INTRODUCTION

Evolutionary history is embedded in the DNA content of genomes. Considering the structure of double-stranded DNA[1], Chargaff's first parity rule is always followed, and makes intuitive sense : G＝C and T＝A and (G＋A) ＝ (T＋C)[2]. However, Chargaff's second parity rule[3]—in which similar nucleotide relationships are retained across single DNA strands—is not so easily understood. Even though the rule was reported 40 years ago, the biological significance of Chargaff's second parity rule has yet to be elucidated—presumably because of its unclear biological basis. Therefore, the evolutionary significance of this rule is not understood. However, a recent publication claims to have solved this historic puzzle[4]. The solution is based on the fact that genome structure is homogeneous regarding nucleotide composition[5,6], and that both the forward and reverse strands are almost the same[7]. Deviations from Chargaff's second parity rule were reported[8,9], and animal mitochondrial deviations differed fundamentally from the rule[10]. The first parity rule is due to the physicochemical characteristics of nucleotides that form double-stranded DNA, and the second parity rule is due to similarities of nucleotide composition between the forward and reverse strands. These two rules represent different phenomena : the former is mathematically definitive and independent of biological significance, and the latter is less definite, and may or may not have biological significance. The recent study used regression lines (derived from Chargaff's second parity rule) and showed that all advanced forms of life descended from a single origin[11]. The present study was designed to evaluate Chargaff's second parity rule using a simulation of the random choice of nucleotides.

Received February 12, 2010 ; accepted April 19, 2010
Reprint requests to : Yoshifumi Ebara
Terao Child Clinic, Terao, Takasaki, Gunma 370-0865, Japan

METHODS, RESULTS AND DISCUSSION

We tested whether nucleotide alternations in the genome are based on random nucleotide mutations (via biological evolution) using a simulation analysis with random nucleotide choice. We chose four nucleotide contents of virtual genome segments at random (from 1–100) and the nucleotide numbers were plotted each other among four nucleotides. We assumed that genome segments were replaceable with the DNA of other organisms and then evaluated evolutionary genome changes across organisms. When plotted, the relationships were highly heteroskedastic, as shown in Fig. 1, upper panel. Other nucleotide relationships were also heteroskedastic (data not shown). Although there were no clear correlations, a linear equation with a regression coefficient of nearly 0 was calculated (Table 1). The slope was almost 0 and the line parallel

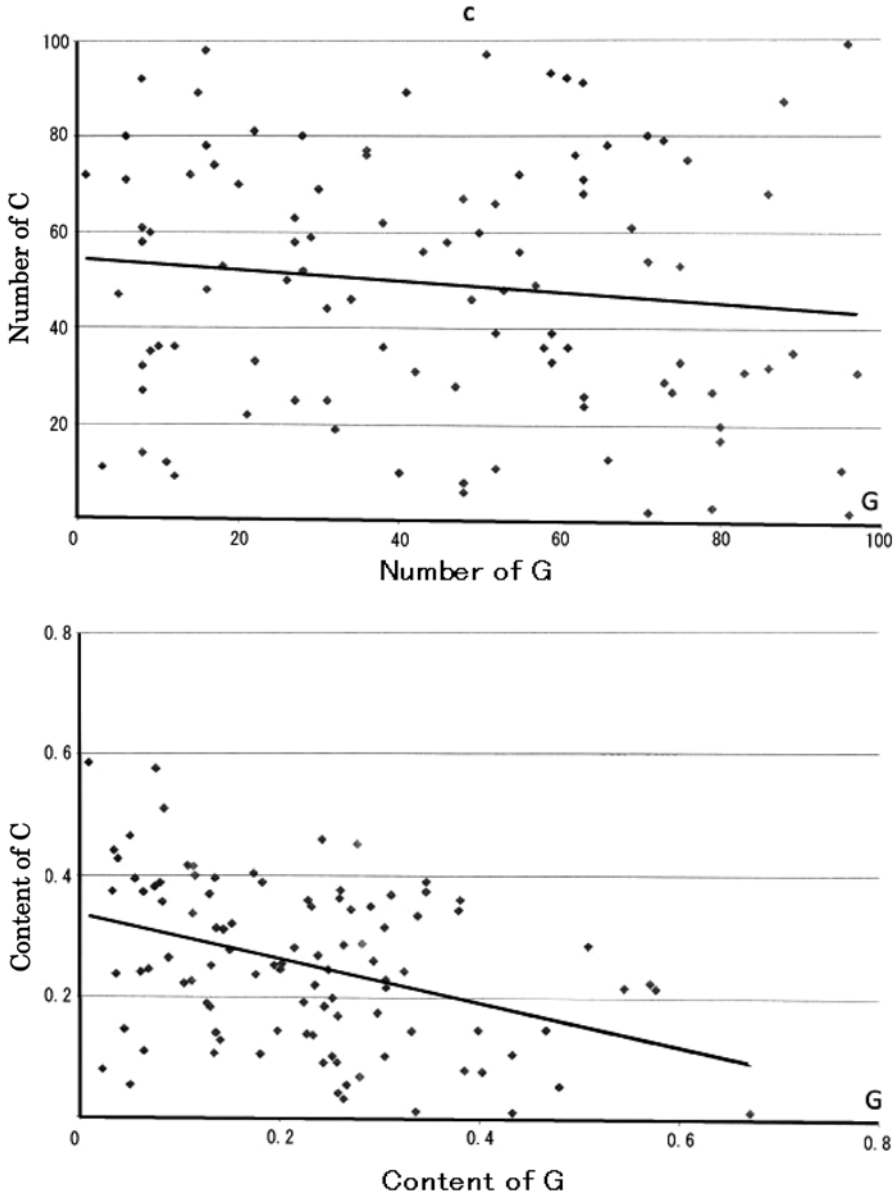

Fig. 1 Relationships between the contents of G and C nucleotides. The G and C nucleotides, chosen at random numbers from 1 to 100, were plotted against each other. Upper panel : C versus G in nucleotide content numbers. Lower panel : C versus G in nucleotide contents normalized to 1.

with the horizontal axis. The constant value at the vertical intercept was around 50 in each equation—the average of 1–100, an expected result.

When the nucleotide content numbers were normalized to 1, the sample distribution changed (Fig. 1, lower panel) : each value shifted towards the left. Computational calculations showed a linear equation with a small regression coefficient (Table 2). The slope was negative regarding the relationship between different nucleotide numbers, and the slope was 1 for the relationship between the content numbers of the same nucleotides (Table 2). Using normalized values of the four equations, as the summation of the four nucleotides is 1, the summation of the four equation slopes is 0 and that of the constant values at the vertical intercept is 1.0 in all cases[7], but not in the original values (Table 1). The consistent result, based on various organisms, was a product of codon evolution[7]. Based on Chargaff's first parity rule and $G+C+T+A=1$, $2G+2T=1$ or $2G+2A=1$. Finally, $T=0.5-G$ or $A=0.5$

Table 1 Regression lines representing nucleotide contents based on random choice of four nucleotide contents.

		R			R
G=	G+0	1	G=	−0.120C+50.8	0.12
C=	−0.115G+54.6	0.12	C=	C+0	1
T=	−0.036G+56.6	0.03	T=	−0.040C+57.0	0.04
A=	0.082G+47.7	0.07	A=	−0.007C+51.7	0.02
G=	−0.030T+46.5	0.03	G=	0.065A+41.5	0.07
C=	−0.033T+51.2	0.04	C=	−0.005A+49.7	0.02
T=	T+0	1	T=	0.046A+52.6	0.05
A=	0.049T+48.6	0.05	A=	A+0	1

R : regression coefficient.

−G, and each nucleotide content is expressed by G content as follows : C=G, A=0.5−G, and T=0.5−G. Eventually, Chargaff's parity rules are expressed by linear formulas. However, these formulas were not obtained by our simulation analysis using a random choice of nucleotides.

Using a substantial amount of genomic data, we showed that only double-stranded DNA obeys Chargaff's second parity rule, and that single-stranded DNA (such as some viral genomes) does not. To solve the puzzle of Chargaff's second parity rule, both double-stranded DNA and homogeneity of the nucleotides are necessary[4]. This is consistent with Mitchell and Bridge's results[12]. These facts indicate that the structure of double-stranded DNA is important in elucidating biological evolution[4].

In the present study, single DNA strands were simulated by random choice of nucleotides, and Chargaff's second parity rule was not obtained (Fig. 1 and Tables 1 and 2). Simulation analysis (with random nucleotide choice) applies to primitive forms of life and helps determine whether initial genomes were constructed randomly—without the codon rule. We previously showed that a random choice of nucleotides for initial nucleotide polymers can not produce functional proteins in the absence of the codon rule (i.e. surprisingly, protein formation preceded codon formation)[13], although we currently have no explanation for this inverse relationship. The present result is consistent with our previously proposed model for the formation of primitive life[13].

Table 2 Regression lines representing nucleotide contents based on random choice of nucleotide contents after normalization.

			R				R
G=	G+0		1	G=	−0.394C+0.324		0.38
C=	−0.358G+0.334		0.38	C=	C+0		1
T=	−0.383G+0.361		0.41	T=	−0.211C+0.329		0.21
A=	−0.259G+0.305		0.27	A=	−0.395C+0.347		0.39
Σ	0	1		Σ	0	1	
G=	−0.433T+0.344		0.41	G=	−0.280A+0.294		0.27
C=	−0.217T+0.313		0.21	C=	−0.387A+0.349		0.39
T=	T+0		1	T=	−0.333A+0.358		0.34
A=	−0.350T+0.344		0.34	A=	A+0		1
Σ	0	1		Σ	0	1	

R : regression coefficient.

REFERENCES

1) Watson JD, Crick FHC : Genetical implications of the structure of deoxyribonucleic acid. Nature **171** : 964-967, 1953.
2) Chargaff E : Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. Experimentia, VI : 201-209, 1950.
3) Rudner R, Karkas JD, Chargaff E : Separation of B. subtilis DNA into complementary strands. 3. Direct analysis. Proc. Natl. Acad. Sci. USA **60** : 921-922, 1968.
4) Sorimachi K : A proposed solution to the historic puzzle of Chargaff's second parity rule. Open Genom. J **2** : 12-14, 2009.
5) Sorimachi K, Okayasu T : Gene assembly consisting of small units with similar amino acid composition in the Saccharomuces cerevisiae genome. Mycoscience **44** : 415-417, 2003.
6) Sorimachi K, Okayasu T : Genomic structure is homogeneous based on codon usages. Curr. Top. Pep. Protein Res **8** : 19-24, 2007.
7) Sorimachi K, Okayasu T : Codon evolution is governed by linear formulas. Amino Acids **34** : 661-668, 2008.
8) Nikolaou C, Almirantis Y : Deviations from Chargaff's second parity rule in organelle DNA insights into the evolution of organelle genomes. Gene **381** : 34-41, 2006.
9) Bell SJ, Forsdyke DR : Deviations from Chargaff's

second parity rule with direction of transcription. J. Theor. Biol **197** : 63-76, 1999.

10) Sorimachi K, Okayasu T : Universal rules governing genome evolution expressed by linear formulas. Open Genom. J **1** : 33-43, 2008.

11) Sorimachi K : Genomic data provides simple evidence for a single origin of life. Natural Scinece **2** : 519-525, 2010.

12) Mitchell D, Bridge R : A test of Chargaff's second rule. Biochem. Biophys. Res. Commun **340** : 90-94, 2006.

13) Sorimachi K, Okayasu T : Mathematical proof of the chronological precedence of protein formation over codon formation. Curr. Top. Pep. Protein Res **8** : 25-34, 2007.