

# Evidence for the independent divergence of vertebrate and high C/G ratio invertebrate mitochondria from the same origin

Kenji Sorimachi<sup>1\*</sup>, Teiji Okayasu<sup>2</sup>, Shuji Ohhira<sup>3</sup>, Ichio Fukasawa<sup>1</sup>, Nobuhide Masawa<sup>4</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, Dokkyo Medical University, Tochigi, Japan;

\*Corresponding Author: [kenjis@dokkyomed.ac.jp](mailto:kenjis@dokkyomed.ac.jp)

<sup>2</sup>Center of Medical Informatics, Dokkyo Medical University, Tochigi, Japan

<sup>3</sup>Laboratory for International Environmental Health, Dokkyo Medical University, Tochigi, Japan

<sup>4</sup>Department of Anatomic Diagnostic Pathology, Dokkyo Medical University, Tochigi, Japan

Received 28 April 2012; revised 30 May 2012; accepted 12 June 2012

## ABSTRACT

The examination of organelle nucleotide contents across a range of groups can provide insight into evolutionary history. We compared the nucleotide contents of a large number of vertebrate and high C/G ratio invertebrate mitochondria in both coding/non-coding regions and complete strands. For most nucleotides tested, high regression coefficients were obtained; when plotted with lines of best fit, G and T contents formed a cuneiform, and C and A contents over-lapped. These findings suggest that vertebrate and high C/G ratio invertebrate mitochondria descended independently from the same origin.

**Keywords:** Evolution; Genome; Mitochondria; Nucleotide Contents; Vertebrate; Invertebrate

## 1. INTRODUCTION

Assuming that nucleotide or amino acid replacement rates are constant in a gene during biological evolution [1], evolutionary phylogenetic trees have been drawn based on differences in nucleotide (or amino acid) substitutions in target genes [2-6]. However, gene stability is known to vary across genes within an organisms and between organisms. In addition, evolutionary phylogenetic trees are assumed to have a single origin. However, Darwin did not explicitly state this in his seminal text "On the Origin of Species". Using cluster analyses, we previously quantified the nucleotide contents of a range of organisms (bacteria, archaea and eukaryotes) using data from complete genomes [7]. Eventually, all organisms were classified into two codon patterns, and analysis suggested that these two groups diverged from a sin-

gle origin [7]. However, these results were based on mathematical analyses and not direct evidence. Furthermore, the addition and removal of taxa have large effects on tree topology in phylogenetic trees.

The fact that evolution can be expressed using a linear formula suggests that evolution proceeded linearly. Thus, the expression of two evolutionary processes using lines of best fit can be informative. Crossing of the two lines suggests that the two evolutionary processes diverged at one stage, and a formation of a cuneiform at the end of the two lines suggests that both evolved from a single origin [8]. In fact, invertebrate mitochondria were divided into two groups based on two regression lines; high C/G and low C/G ratios [9,10]. A species divergence later in evolutionary history would be indicated by separation into two lines. The degree of difference in two evolutionary processes is expressed by the angle of the cuneiform; small and large differences in evolution are expressed by sharp and dull angles, respectively, and the same evolutionary processes by an overlapping single line. Prior analyses have shown regression lines representing nucleotide relationships in coding regions were almost identical in chromosomal DNA among bacteria, archaea and eukaryotes [11]. In a previous study [12], two regression lines representing the homonucleotide contents in chloroplast and plant mitochondria formed a cuneiform in both coding and non-coding regions. Using a similar approach a recent study used genomic data to provide simple evidence for the single origin of life [10]. The present study was designed to determine the evolutionary history of vertebrate and invertebrate mitochondria, since both were grouped together in a previous study [12].

## 2. METHODS

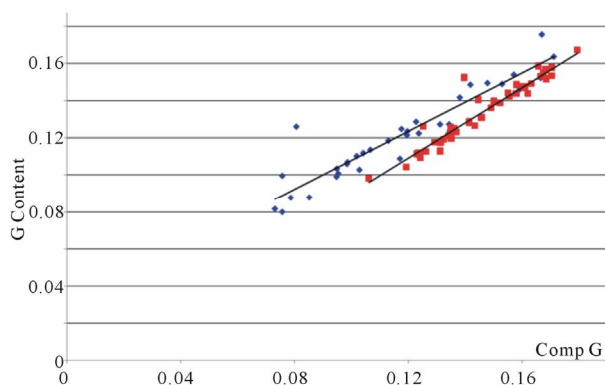
Genome data were obtained from the National Center

for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/sites>), and the list of organelles examined has been described previously [12]. The nucleotide contents of coding and non-coding regions were compared in relation to the complete single DNA strand [10]. When plotted, the nucleotide contents were normalized to 1 ( $G + C + A + T = 1$ ). Calculations were performed using Microsoft Excel (version 2003).

### 3. RESULTS AND DISCUSSION

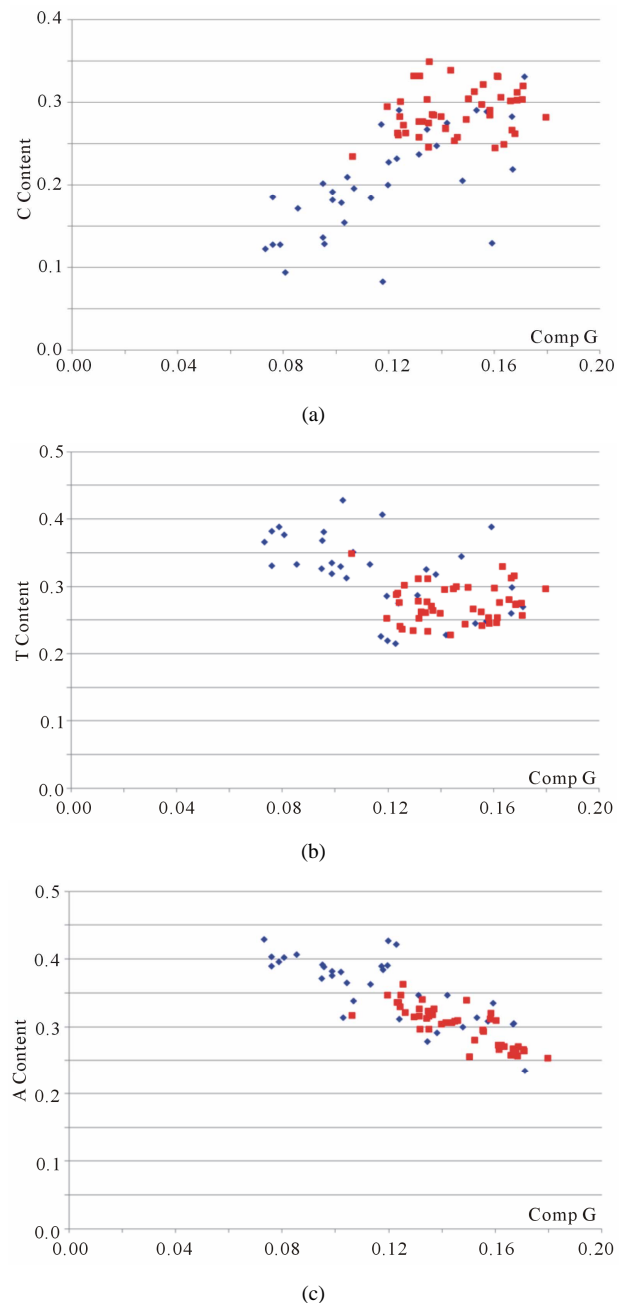
In a previous study [9], when the relationship between four nucleotide contents were examined, linear regression lines with high regression coefficients were obtained in nuclear, chloroplast and plant mitochondria in the complete genome. Plotting four nucleotide relationships in animal mitochondria, the relationships were heteroskedastic. After classifying invertebrate mitochondria into two groups, the relationships between homonucleotides and their analogs were approximately linear in invertebrate mitochondria [10]. However, a group of vertebrate mitochondria was incorporated into one of the invertebrate groups, and nucleotide alterations were limited to homonucleotides [12].

In this study, the homonucleotide content relationships (between coding/non-coding regions and the complete strand) were expressed by linear regression and lines of best fit. The nucleotide contents in the coding regions were plotted against G nucleotide contents in the complete strand (**Figure 1**). A correlation was found between G content in the coding region and complete single DNA strand, and a clear separation was found between vertebrate and high C/G ratio invertebrate mitochondria (**Figure 1**). The two regression lines formed a cuneiform suggesting that both descended independently from the same single origin. However, it might not be denied that both lines converge in mitochondrial evolution.



**Figure 1.** Nucleotide relationships in normalized vertebrate (red squares) and high C/G ratio invertebrate (blue square) mitochondrial values. The G nucleotide content in the coding region (vertical axis) is plotted against G content in the complete single DNA strand (horizontal axis).

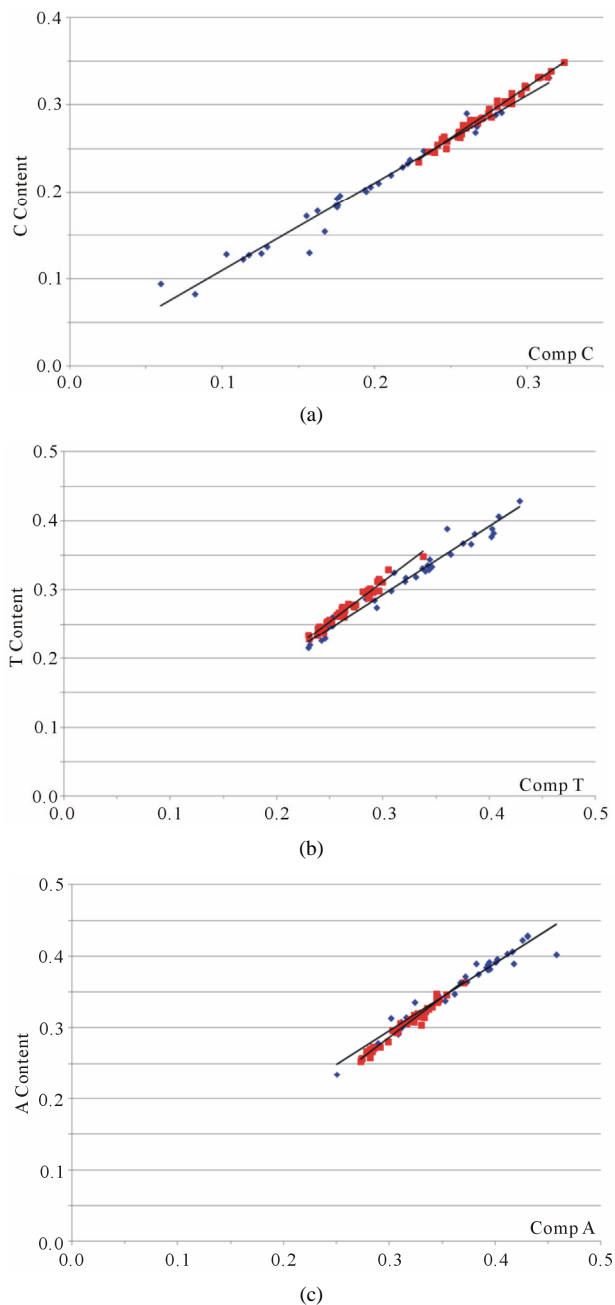
Classification of hagfish (*Eptatretus burgeri*), which is basal to vertebrate and is living fossil, has been controversial [13]. Hagfish whose total G content (0.106) was lowest among vertebrate examined completely fitted the vertebrate line rather than the invertebrate line (**Figure 1**). On the other hand, bony fish (*Oryzias latipes*) G content (0.179) was the top of the line, while human being (*Homo*



**Figure 2.** Nucleotide relationships in normalized vertebrate (red squares) and high C/G ratio invertebrate (blue squares) mitochondrial values. The nucleotide content in the coding region (vertical axis) is plotted against the G content in the complete single DNA strand (horizontal axis). (a) C; (b) T; (c) A.

*sapiens*) G content (0.132) was the middle of the line. These results suggest that the G content apparently increases along evolution. In fact, G or C content of the origin of life was  $\sim 0.03$  based on mathematical calculations, while that of T or A was  $\sim 0.47$  [10].

High C/G ratio invertebrate plus vertebrate mitochondria heteronucleotides (the C/T coding regions) showed



**Figure 3.** Homonucleotide relationships in normalized vertebrate (red squares) and high C/G ratio invertebrate (blue square) mitochondrial values. The nucleotide content in the coding region (vertical axis) is plotted against homonucleotide content in the complete single DNA strand (horizontal axis). (a) C; (b) T; (c) A.

weak correlation with G content (in the complete strand), and the regression coefficients were relatively small (**Figures 2(a) and (b)**). The relationships between vertebrate-only mitochondria T/C and G contents were heteroskedastic, and their distributions did not overlap [10]. The correlation between A and G nucleotides was comparatively higher than with the pyrimidines C and T (**Figure 2(c) and Table 1**).

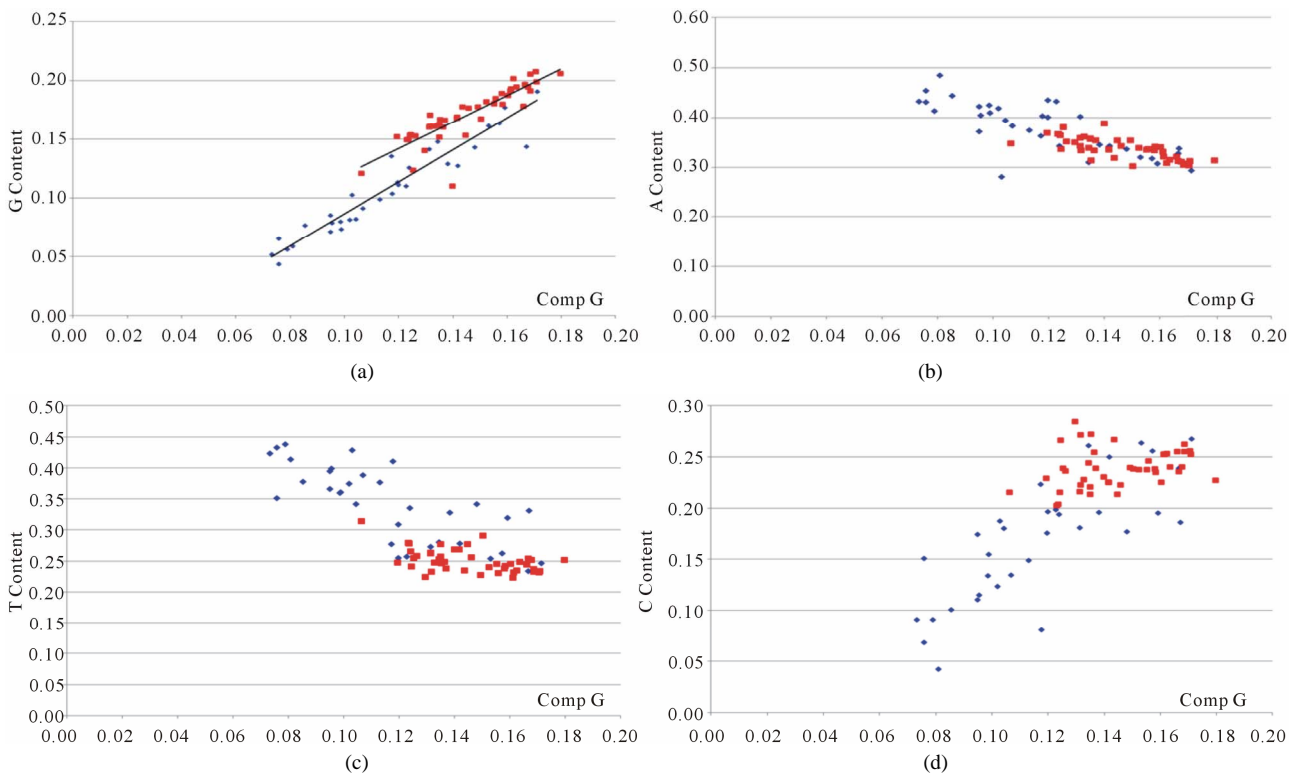
Two high regression coefficients with overlapping lines of best fit were found when plotting C or A contents (the coding region and complete single DNA strand; **Figure 3**). This suggests that C and A alternations were similar in the evolutionary history of vertebrate and high C/G ratio invertebrate mitochondria, while the formation of a cuneiform in T content (vertebrate and high C/G ratio invertebrate mitochondria) suggests that both groups descended independently from the same origin.

In summary, both the G and T homonucleotide relationships suggest that the two organelle groups descended from the same origin. When non-coding and the complete nucleotide contents were investigated, similar results were obtained (**Figures 4 and 5**), a finding consistent with a previous report [10]. However, although G and T homonucleotide regression lines formed a cuneiform,

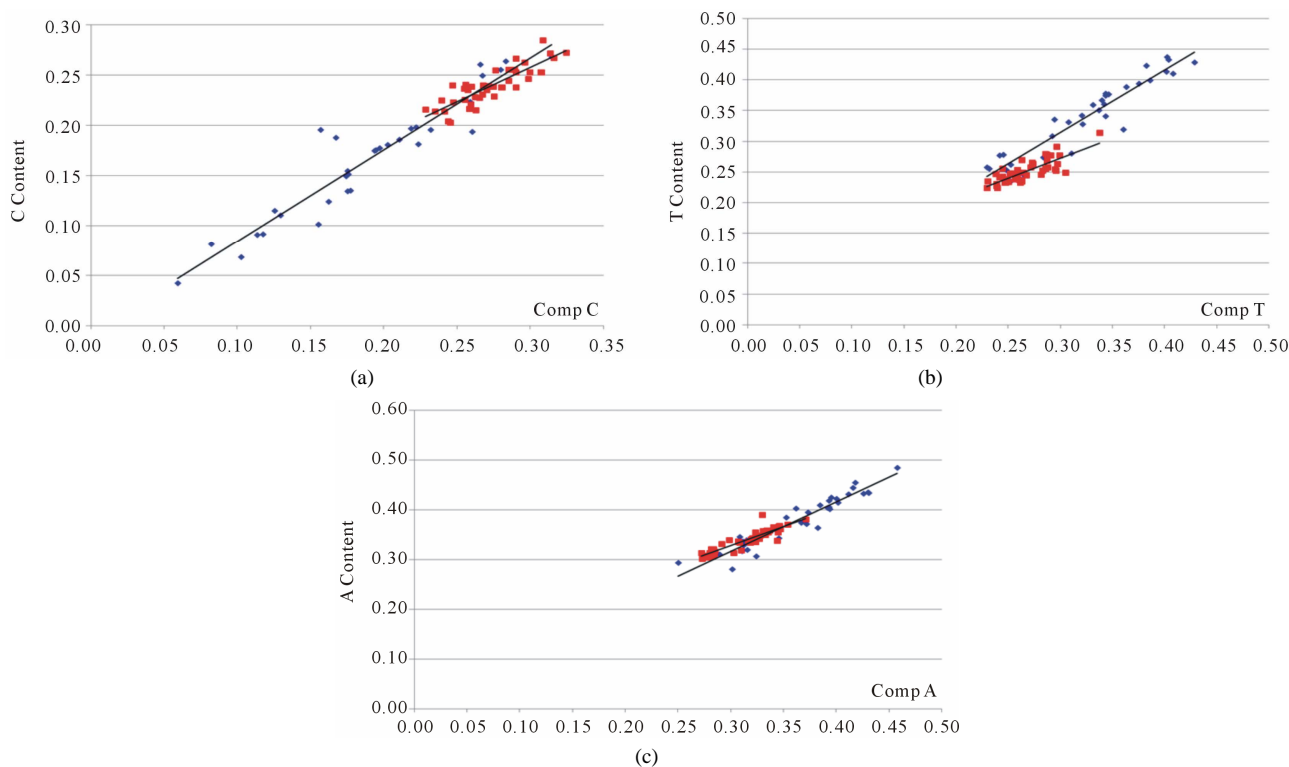
**Table 1.** Regression lines representing nucleotide contents in the coding and non-coding regions against the nucleotide contents in the complete single strand DNA based on vertebrate plus high C/G ratio invertebrate mitochondria.

Coding	R	Non-coding	R
$G_c = 0.725 G + 0.032$	0.92	$G_n = 1.532 G - 0.059$	0.93
$G_c(i) = 0.781 G + 0.299$	0.94	$G_n(i) = 1.364 G - 0.050$	0.96
$G_c(v) = 0.948 G - 0.004$	0.96	$G_n(v) = 1.133 G + 0.007$	0.87
$C_c = 1.732 G + 0.023$	0.75	$C_n = 1.585 G - 0.002$	0.78
$T_c = -0.966 G + 0.421$	0.55	$T_n = -1.819 G + 0.531$	0.79
$A_c = -1.491 G + 0.524$	0.85	$A_n = -1.299 G + 0.530$	0.81
<hr/>			
$G_c = 0.195 C + 0.081$	0.55	$G_n = 0.630 C - 0.006$	0.85
$C_c = 1.037 C + 0.005$	0.99	$C_n = 0.893 C - 0.005$	0.97
$T_c = -0.649 C + 0.448$	0.82	$T_n = -0.959 C + 0.519$	0.92
$A_c = -0.583 C + 0.466$	0.74	$A_n = -0.564 C + 0.492$	0.78
<hr/>			
$G_c = -0.211 T + 0.189$	0.49	$G_n = -0.641 T + 0.331$	0.72
$C_c = -1.119 T + 0.580$	0.90	$C_n = -0.936 T + 0.482$	0.86
$T_c = 0.920 T + 0.025$	0.98	$T_n = 1.159 T - 0.049$	0.94
$T_c(i) = 0.984 T - 0.002$	0.98	$T_n(i) = 1.016 T + 0.010$	0.94
$T_c(v) = 1.167 T - 0.037$	0.98	$T_n(v) = 0.647 T + 0.078$	0.80
$A_c = 0.410 T + 0.207$	0.44	$A_n = 0.419 T + 0.236$	0.48
<hr/>			
$G_c = -0.329 A + 0.239$	0.71	$G_n = -0.829 A + 0.433$	0.86
$C_c = -1.017 A + 0.596$	0.75	$C_n = -0.940 A + 0.525$	0.79
$T_c = 0.334 A + 0.180$	0.33	$T_n = 0.867 A - 0.003$	0.64
$A_c = 1.012 A - 0.014$	0.98	$A_n = 0.902 A + 0.054$	0.96

Xc and Xn mean the nucleotide content in the coding and non-coding regions, respectively. (i) and (v) represent 31 high C/G ratio invertebrate and 45 vertebrate mitochondria, respectively.



**Figure 4.** Nucleotide relationships in normalized vertebrate (red squares) and high C/G ratio invertebrate (blue squares) mitochondrial values. The nucleotide content in the non-coding region (vertical axis) is plotted against the G content in the complete single DNA strand (horizontal axis). (a) G; (b) A; (c) T; (d) C.



**Figure 5.** Homonucleotide relationships in normalized vertebrate (red squares) and high C/G ratio invertebrate (blue square) mitochondrial values. The nucleotide content in the non-coding region (vertical axis) is plotted against homonucleotide content in the complete single DNA strand (horizontal axis). (a) C; (b) T; (c) A.

C and A homonucleotide regression lines overlapped. The reason for this is unclear, particularly since G and A nucleotides are both purines, and C and T both pyrimidines.

The findings suggest that vertebrate and high C/G ratio invertebrate mitochondria descended independently from the same origin. It has previously been suggested that mitochondria were derived from proteobacteria [14, 15]; if this is the case, the findings from this study suggest that this occurred once in evolutionary history.

## REFERENCES

- [1] Zuckerkandl, E. and Pauling, L.B. (1962) Molecular disease, evolution, and genetic heterogeneity. In: Kasha, M. and Pullman, B., Eds., *Horizons in Biochemistry*, Academic Press, New York, 189-225.
- [2] Dayhoff, M.O., Park, C.M. and McLaughlin, P.J. (1977) Building a phylogenetic trees: Cytochrome C. In: *Atlas of Protein Sequence and Structure*, National Biomedical Foundation, Washington DC, 5, 7-16.
- [3] Sogin, M.L., Elwood, H.J. and Gunderson, J.H. (1986) Evolutionary diversity of eukaryotic small subunit rRNA genes. *Proceedings of the National Academy of Sciences, USA*, **83**, 1383-1387. [doi:10.1073/pnas.83.5.1383](https://doi.org/10.1073/pnas.83.5.1383)
- [4] de Poupiana, L.R., Turner, R.J., Steer, B.A. and Schimmel P. (1998) Genetic code origins: tRNAs older than their synthetases? *Proceedings of the National Academy of Sciences, USA*, **95**, 11295-11300.
- [5] Doolittle, W.F. and Brown, J.R. (1994) Tempo, mode, the progenote, and the universal root. *Proceedings of the National Academy of Sciences, USA*, **91**, 6721-6728.
- [6] Maizels, N. and Weiner, A.M. (1994) Phylogeny from function: Evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proceedings of the National Academy of Sciences, USA*, **91**, 6729-6734.
- [7] Okayasu, T. and Sorimachi, K. (2008) Organisms can essentially be classified according to two codon patterns. *Amino Acids*, **36**, 261-271. [doi:10.1007/s00726-008-0059-0](https://doi.org/10.1007/s00726-008-0059-0)
- [8] Sorimachi, K. (2010) Evolution based on genome structure: "The diagonal genome universe". *National Sciences*, **2**, 1104-1112.
- [9] Sorimachi, K. (2010) Genome data provides simple evidence for a single origin of life. *National Sciences*, **2**, 519-525.
- [10] Sorimachi, K. (2010) Codon evolution in double-stranded organelle DNA: Strong regulation of homonucleotides and their analog alternations. *National Sciences*, **2**, 846-854.
- [11] Sorimachi, K. and Okayasu, T. (2008) Codon evolution is governed by linear formulas. *Amino Acids*, **34**, 661-668. [doi:10.1007/s00726-007-0024-3](https://doi.org/10.1007/s00726-007-0024-3)
- [12] Sorimachi, K. and Okayasu, T. (2008) Universal rules governing genome evolution expressed by linear formulas. *Open Journal of Genomics*, **1**, 33-43. [doi:10.2174/1875693X00801010033](https://doi.org/10.2174/1875693X00801010033)
- [13] Janvier, P. (2010) microRNAs revive old views about jawless vertebrate divergence and evolution. *Proceedings of the National Academy of Sciences, USA*, **107**, 19137-19138. [doi:10.1073/pnas.1014583107](https://doi.org/10.1073/pnas.1014583107)
- [14] Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G.J. and Woese, C.R. (1985) Mitochondrial Origins. *Proceedings of the National Academy of Sciences, USA*, **82**, 4443-4447. [doi:10.1073/pnas.82.13.4443](https://doi.org/10.1073/pnas.82.13.4443)
- [15] Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, B., Lemieux, C., Sankoff, D., Turmel, M. and Gray, M.W. (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature*, **387**, 493-497. [doi:10.1038/387493a0](https://doi.org/10.1038/387493a0)