Short Communication

GC-content of tRNA genes classifies archaea into two groups

Yoshio Kawai^{1,*} and Yoshimi Maeda^{2,†}

¹ Faculty of Life Sciences, Toyo University, Gunma 374-0193, Japan

² College of Science and Engineering, Iwaki Meisei University, Iwaki, Fukushima 970-8551, Japan

(Received January 13, 2009; Accepted May 7, 2009)

Key Words——archaea; evolution; GC-content; tRNA gene

Over 30 years have passed since Woese and Fox proposed the concept of archaea (Woese and Fox, 1977). With improving culture techniques and sequencing techniques, abundant data on archaea have been accumulated. These were represented by recent publications of works on archaea (Blum, 2008; Cavicchioli, 2007; Garrett and Klenk, 2007). In these works we can find about 50 complete genome sequences of archaea that are available for further investigation into their functions. One of the current interesting topics is that *Nanoarchaeum equitans* creates functional tRNAs from a combination of separate genes (Randau et al., 2005). We need further and extensive investigation by various approaches. There are a few areas overlooked and to be studied. The GC-content analysis started to be utilized to analyze the relation between DNA GC-content and its melting temperature (Marmur and Doty, 1962). Since then, it has been widely applied to various fields including genetic engineering techniques like Southern hybridization. Although classification of bacteria systematically by using the GC-content analysis is not used as a vital tool, it is unquestionable that GC-content is one of the basic parameters reflecting characteristics of genome DNA. It may be worth re-checking, especially if it is focused on archaeal tRNA

genes by not only recently developed polygenetic analysis based on the genome sequences but also the traditional and simple GC-content analysis. So we expect the GC-content analysis to offer basic details of archaea genome, and hopefully insights into archaea evolution.

We listed 48 species belong to archaea of which the complete genome had been sequenced, and two eubacteria, *Bacillus subtilis* and *Escherichia coli* with some parameters (Table 1-a and 1-b). All the sequence data and related information were downloaded from GenBank. The GC-content of tRNA gene, GC_{tRNA}, is the weighted average value of GC molar percents of each tRNA genes. GC_{tRNA} value and GC-content of each tRNA genes were rounded off, so in some cases GC_{tRNA} values do not agree with a value obtained by direct calculation for all nucleotides composed of all tRNA genes of an archaeon. In this estimation, we regarded an intro region(s), if it existed, as a part of the tRNA gene. Each optimum growth temperature, T_{opt} , was obtained mainly from Stetter (1999), Kletzin (2007) and Sowers (1995), and also from original papers (Amo et al., 2002; Bräure et al., 2006; Elkins et al., 2008; Erkel et al., 2005; Huber et al., 2000; Kamagata et al., 1992; Kendall et al., 2006). If the optimum temperature was given as a temperature range (optimum, a – b °C), we simply adopted a value rounded off $(a+b)/2$ from the first decimal place.

At first, we tried to find any correlation between the complete genome DNA GC-contents, GC_{CG}, and the optimum growth temperatures, but could not (figure

* Address reprint requests to: Dr. Yoshio Kawai, Faculty of Life Sciences, Toyo University, Itakura-machi, Oura-gun, Gunma 374-0193, Japan.

† Present address: 4-33-8-105, Seijo, Setagaya-ku, Tokyo 157-0066, Japan.

Table 1-a. Characteristic parameters of 30 completely sequenced Euryarchaeota.

Species	GC _{CG} (%)	GC _{tRNA} (%)	Δ_N	T_{opt} (°C)
<i>Archaeoglobus fulgidus</i>	50	68	0.36	83
<i>Candidatus Methanoregula boonei</i>	54	61	0.13	37
<i>Haloarcula marismortui</i>	64	62	-0.03	45
<i>Halobacterium salinarum</i>	70	62	-0.11	43
<i>Halobacterium</i> sp. NRC-1	67	62	-0.07	43
<i>Haloquadratum walsbyi</i>	48	61	0.27	40
<i>Methanobrevibacter smithii</i>	31	59	0.9	37
<i>Methanocaldococcus jannaschii</i>	32	64	1	85
<i>Methanococcoides burtonii</i>	43	60	0.4	23
<i>Methanococcus aeolicus</i>	30	60	1	46
<i>Methanococcus maripaludis</i>	34	61	0.79	38
<i>Methanococcus vanniellii</i>	31	61	0.97	38
<i>Methanocorpusculum labreanum</i>	50	60	0.2	37
<i>Methanopyrus kandleri</i>	62	73	0.18	98
<i>Methanosaeta thermophila</i>	55	62	0.13	58
<i>Methanosarcina acetivorans</i>	46	61	0.33	38
<i>Methanosarcina barkeri</i>	42	60	0.43	45
<i>Methanosarcina mazei</i>	45	61	0.36	41
<i>Methanosphaera stadtmanae</i>	29	61	1.1	35
<i>Methanospirillum hungatei</i>	46	60	0.3	34
<i>Methanothermobacter thermoautotrophicus</i>	52	62	0.19	68
<i>Natronomonas pharaonis</i>	65	63	-0.03	45
<i>Picrophilus torridus</i>	37	54	0.46	60
<i>Pyrococcus abyssi</i>	53	71	0.34	96
<i>Pyrococcus furiosus</i>	41	71	0.73	100
<i>Pyrococcus horikoshii</i>	42	71	0.69	98
<i>Thermococcus kodakaraensis</i>	54	69	0.28	96
<i>Thermoplasma acidophilum</i>	48	61	0.27	59
<i>Thermoplasma volcanium</i>	41	61	0.49	59
uncultured methanogenic archaeon RC-1	54	60	0.11	50

not shown). It does not seem that there exists a correlation between the genome DNA GC-contents and the optimum growth temperatures. Archaeal genome DNA GC-contents are widely distributed from about 30% up to 70%, and are independent of the optimum growth temperatures, indicating no new interesting information. Therefore, we only imagine that the GC-content of genome DNA is not very meaningful for archaea survival even if they live in an extreme environment. But it should be noted that *Nanoarchaeum equitans*, *Staphylothermus marinus*, *Pyrococcus furiosus* and *Pyrococcus horikoshii* do survive in a higher temperature environment than the calculated melting temperature of their genome DNA. This means at least more than half the region of their genome DNA is always in a "melting" state if they have no specific

mechanism to maintain the genome DNA in double helix. Some genera of archaea have histone-like proteins and condense DNA into compact structures (Fahner et al., 2001), or have reverse gyrase which will give and improve thermal stability to the DNA double helix (de la Tour et al., 1990). But it is ambiguous whether these two facts give a clear explanation for maintaining two DNA strands in helix in the cases of the above mentioned four species. Further intensive studies may yield a something new concept related to thermal resistance of the DNA double helix. There may be no physiological need to maintain the genome DNA in double helix.

On the other hand, it seems that GC-contents of archaeal tRNA genes are somehow restricted by environmental factor(s). There exists a strong linear corre-

Table 1-b. Characteristic parameters of 16 completely sequenced Crenarchaeota, *Candidatus Korarchaeum cryptofilum*, *Nanoarchaeum equitans*, *B. subtilis* and *E. coli*.

Species	GC _{CG} (%)	GC _{tRNA} (%)	ΔN	T_{opt} (°C)
<i>Aeropyrum pernix</i>	57	72	0.26	93
<i>Caldivirga maquilingensis</i>	43	67	0.56	85
<i>Hyperthermus butylicus</i>	53	72	0.36	101
<i>Ignicoccus hospitalis</i>	56	71	0.27	90
<i>Metallosphaera sedula</i>	46	67	0.46	75
<i>Nitrosopumilus maritimus</i>	34	58	0.71	28
<i>Pyrobaculum aerophilum</i>	53	70	0.32	100
<i>Pyrobaculum arsenaticum</i>	55	70	0.27	95
<i>Pyrobaculum calidifontis</i>	57	70	0.28	93
<i>Pyrobaculum islandicum</i>	49	70	0.43	100
<i>Staphylothermus marinus</i>	35	68	0.94	92
<i>Sulfolobus acidocaldarius</i>	37	68	0.83	75
<i>Sulfolobus solfataricus</i>	36	67	0.86	82
<i>Sulfolobus tokodaii</i>	33	69	1.1	80
<i>Thermofilum pendens</i>	57	70	0.23	88
<i>Thermoproteus neutrophilus</i>	59	71	0.2	85
<i>Candidatus Korarchaeum cryptofilum</i>	49	71	0.45	85
<i>Nanoarchaeum equitans</i>	30	73	1.4	90
<i>Bacillus subtilis</i>	43	59	0.37	37
<i>Escherichia coli</i>	50	59	0.18	37

lation ($r=0.902$) between the GC-contents of tRNA genes and the optimum growth temperatures. Archaea grown under 70°C have tRNA genes of which the GC-contents are up to about 63% and archaea grown above 70°C have those of which the GC-contents are higher than 65% (Fig. 1). Crenarchaeota have high GC-contents of tRNA genes with the one exception of *Nitrosopumilus maritimus*. Mesophilic bacteria, *Escherichia coli* and *Bacillus subtilis*, seem not so different from Euryarchaeota already sequenced. Recently, Kimura et al. reported that there was a strong correlation between the optimum growth temperatures of archaeal species and the GC-contents of the 16S rRNA genes, and also that, with the exception of some species of halophilic archaea, 16S rRNA genes of mesophilic archaea had low GC-contents, and the 16S rRNA genes of thermophilic and hyperthermophilic archaea had much higher GC-contents than the 16S rRNA genes of mesophilic archaea (Kimura et al., 2007). Their result partially supports our inference: tRNA molecules which must take higher order structures to function should have higher GC-contents if they act in a higher temperature environment. Compared with their result, the GC-contents of tRNA genes, in the

Fig. 1. Correlation between the averaged GC-contents of tRNA genes and the optimum growth temperatures.

Each averaged GC-content value of tRNA genes of 48 archaeal strains completely sequenced, *B. subtilis* and *E. coli* was plotted against the corresponding optimum growth temperature.

temperature range of 30–120°C, are higher than those of 16S rRNA genes, since our obtained equation for the regression curve is $y = 0.18x + 53.1$ and their equation is $y = 0.20x + 47.8$, where y is the GC-con-

tent and x is the optimum growth temperature. This may be explained by the fact that the higher intramolecular structure of 16S rRNA is stabilized by many ribosomal proteins but tRNA molecule, of which GC-content is the same as that of the tRNA gene, has no support from other molecules to stabilize its own structure, and also the additional fact that there exist differences in length of duplex regions between the 16S rRNA molecule and tRNA molecule. Since the duplex region of the tRNA molecule is much too short, the tRNA molecule requires more GC pairs to stabilize and maintain its own structure. It is rational to have a high GC-content for the tRNA gene if archaea grow in a high temperature environment. As for this point, we will discuss it again later.

Figure 2 shows, at least, the following three points. (1) Most archaea have higher tRNA gene GC-contents than the complete genome DNA GC-contents, which seems also true in the case of mesophilic bacteria. (2) tRNA gene GC-contents scarcely depend on genome DNA GC-contents. (3) Archaea contain two groups distinguished from each other by GC-content of tRNA genes. One group (tentatively named group N) is mainly composed of Euryarchaeota and have "normal" GC-content of the tRNA gene near 60%, whereas the other (named group H) is composed of both the rest of Euryarchaeota sequenced and almost all Crenarchaeota sequenced, and have "high" GC-content of tRNA genes near 70%. The third point, which is the most critical, is worth discussing further. Although the

genome DNA GC-contents are distributed widely in the range from near 30% up to 70%, tRNA gene GC-contents range narrowly and exist only either almost at 60% or 70%. This feature reminds us of something like homeostasis. We already saw that the tRNA gene GC-contents have a positive linear correlation to the optimum growth temperatures; in other words, archaea surviving in a high temperature environment should have high GC-contents of tRNA genes. Although high GC-contents of tRNA genes need not correspond rigorously to high thermal stabilities of tRNA molecules, high GC-content of the tRNA gene is surely advantageous to the tRNA molecule's formation of a more stable three-dimensional structure since the tRNA nucleotide sequence reflects directly GC-content of the tRNA gene. To be recognized by a specific aminoacyl-tRNA synthetase, the tRNA molecule must maintain its unique structure even in a seriously severe hot environment but the tRNA molecule has loops which make its intramolecular duplex structure less stable, and in addition, will be affected to an unnegligible extent by "end effect" which acts also on the unstable duplex stem structure. Considering that even a single G-C pair produces a marked contribution to the thermal stability of the duplex structure (Cantor and Schimmel, 1980), the high GC-contents of group H certainly reflect a severe environment in which their ancestral organisms lived, and must have been a necessary condition to survive.

Figure 3 shows the difference between genome DNA GC-content and tRNA gene GC-content divided by genome DNA GC-content, $\Delta_N = (GC_{tRNA} - GC_{CG}) / GC_{CG}$, plotted against the optimum growth temperature. It is possible to point out a tendency: the lower the optimum growth temperature, the smaller the Δ_N value, namely, the ratio of GC-content of the tRNA gene to that of genome DNA. It seems that Fig. 3 displays an outline of an archaeal evolution pattern including at least two types: "diagonal" type and "vertical" type. For example, the plots of *Methanosarcina barkeri* (numbered 1), *Methanosarcina mazei* (2), *Methanosarcina acetivorans* (3) and *Methanospirillum hungatei* (4) array themselves on one diagonal, which is rational because they share a close relation with each other if based on Archaeal phylogeny (Elkins et al., 2008). The plots of *Hyperthermus butylicus* (5), *Ignicoccus hospitalis* (6), *Pyrobaculum calidifontis* (7), *Thermofilum pendens* (8) and *Thermoproteus neutrophilus* (9) are arrayed on another diagonal and the

Fig. 2. Independence of tRNA gene GC-contents of genome DNA GC-contents.

Each averaged GC-content value of tRNA genes of 48 archaeal strains, *B. subtilis* and *E. coli*, was plotted against the corresponding genome DNA GC-content value.

Fig. 3. Scattergram of the GC-content difference Δ_N ($= (GC_{tRNA} - GC_{CG}) / GC_{CG}$) against the optimum growth temperature.

As for numbers shown, see the text.

plots of *Aeropyrum pernix* (10), *Pyrobaculum aerophilum* (11), *Pyrobaculum arsenaticum* (12) and *Pyrobaculum islandicum* (13) locate near the line. Note that these archaea are all of Crenarchaeota. We now speculate that the former 4 members of methanogen and these 9 members of Crenarchaeota changed their genome in a rather moderate way(s) and evolved out successively. tRNA genes, conversely, tRNA molecules got affinity with aminoacyl-tRNA synthetase instead of losing their high GC-content which stabilized their higher order structure, since too much rigid structure is not favorable to an appropriate and accurate interaction. As for enzymes, it may be said that enzymes in general lost their thermal stability; they acquired affinity for substrates instead. Considering that both enzymes and tRNA molecules have to act precisely as functional molecules, it also can be said that the tRNA molecules have similar thermal characteristics evolutionally. We can also find nearly vertical lines on Fig. 3. Some members of methanogen array themselves on one of these lines located near 38°C. These genera must have evolved in a different manner from the former genera arraying themselves on lines having a slope. Their ancestral genome DNAs might have been seriously and variously damaged by a certain enormous power(s) of the environment since there exists a considerable difference in resulting Δ_N value. Then, some variants among damaged ancestors suddenly abandoned in a specific circumstance managed to survive in cases that they happened to acquire the ability as Methanogen. Hyperthermophilic archaea,

Methanopyrus kandleri (14), *Pyrococcus abyssi* (15), *Pyrococcus furiosus* (16), *Pyrococcus horikoshii* (17) and *Thermococcus kodakaraensis* (18) are arrayed on another vertical line and are of close relation in the phylogenetic tree. Finally, we can easily link the secluded plots of *Staphylothermus marinus* (19), *Sulfolobus acidocaldarius* (20) and *Sulfolobus solfataricus* (21) to the plots of the major group of Crenarchaeota on Fig. 3, assuming a “vertical” relation between *Aeropyrum pernix* (10) and *Staphylothermus marinus* (19) (rational, because the difference between their optimum growth temperatures is a negligible value of 1°C). Both our assumption and the resulting linkage are in good agreement with the archaeal phylogenetic tree. Such a combination of a diagonal relation and vertical relation is also found between the above-mentioned methanogen genera, which means that archaeal evolution might have proceeded through combinations of the diagonal type evolutions and the perpendicular type evolutions.

Our results described here provide fundamental insights and the use of GC-content analysis in studies to reveal evolution and phylogeny of archaea in combination with other modern analysis methods. In addition, this study suggests that further accumulated data including complete genome sequences and GC-content analysis will provide clues to highlight an interesting but hidden feature of archaea in the future.

References

- Amo, T., Paje, M. L. F., Inagaki, A., Ezaki, S., Atomi, H., and Imanaka, T. (2002) *Pyrobaculum calidifontis* sp., a novel hyperthermophilic archaeon that grows in atmospheric air. *Archaea*, **1**, 113–121.
- Blum, P. (2008) *Archaea: New Models for Prokaryotic Biology*, Caister Academic Press, Norfolk, 247 pp.
- Bräuer, S., Cadillo-Quiroz, H., Yashiro, E., Yavitt, J., and Zinder, S. H. (2006) Isolation of a novel acidiphilic methanogen from an acidic peat bog. *Nature*, **442**, 192–194.
- Cantor, C. R. and Schimmel, P. R. (1980) *Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules*, W. H. Freeman and Company, San Francisco, 522 pp.
- Cavicchioli, R. (2007) *Archaea: Molecular and Cellular Biology*, ASM Press, Washington, D.C., 523 pp. + 16 pls.
- de la Tour, C. B., Portemer, C., Nadal, M., Stetter, K. O., and Forterre, P. (1990) Reverse gyrase, a hallmark of the hyperthermophilic archaeobacteria. *J. Bacteriol.*, **172**, 6803–6808.
- Elkins, J. E., Podar, M., Graham, D., Makarova, K. S., Wolf, Y., Randau, L., Hedlund, B. P., Brochier-Armanet, C., Kunin,

- V., Anderson, I., Lapidus, A., Goltsman, E., Barry, K., Koonin, E. V., Hugenholtz, P., Kyrpides, N., Wanner, G., Richardson, P., Keller, M., and Stetter, K. O. (2008) A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl. Acad. Sci. USA*, **105**, 8102–8107.
- Erkel, C., Kemnitz, D., Kube, M., Rieke, P., Chin, K. J., Dedysh, S., Reinhardt, R., Conrad, R., and Liesack, W. (2005) Retrieval of first genome data for rice cluster I methanogens by a combination of cultivation and molecular techniques. *FEMS Microbiol. Ecol.*, **53**, 187–204.
- Fahrner, R. L., Casio, D., Lake, J. A., and Slesarev, A. (2001) An ancestral nuclear protein assembly: Crystal structure of the *Methanopyrus kandleri* histone. *Protein Sci.*, **10**, 2002–2007.
- Garrett, R. A. and Klenk, H.-P. (2007) Archaea: Evolution, Physiology, and Molecular Biology, Blackwell Publishing, Garsington Road, Oxford, 388 pp.
- Huber, R., Sacher, M., Vollmann, A., Huber, H., and Rose, D. (2000) Respiration of arsenate and selenate by hyperthermophilic archaea. *System. Appl. Microbiol.*, **23**, 305–314.
- Kamagata, Y., Kawasaki, H., Oyaizu, H., Nakamura, K., Mikami, E., Endo, G., Koga, Y., and Yamasato, K. (1992) Characterization of three thermophilic strains of *Methanotherix* ("Methanosaeta") *thermophila* sp. nov. and rejection of *Methanotherix* ("Methanosaeta") *thermapacetophila*. *Int. J. Syst. Bacteriol.*, **42**, 469–473.
- Kendall, M. M., Liu, Y., Sieprawaska-Lupa, M., Stetter, K. O., Whitman, W. B., and Boone, D. R. (2006) *Methanococcus aeolicus* sp. nov., a mesophilic, methanogenic archaeon from shallow and deep marine sediments. *Int. J. Syst. Evol. Microbiol.*, **56**, 1525–1529.
- Kimura, H., Ishibashi, J., Masuda, H., Kato, K., and Hanada, S. (2007) Selective phylogenetic analysis targeting 16S rRNA genes of hyperthermophilic archaea in the deep-subsurface hot biosphere. *Appl. Environ. Microbiol.*, **73**, 2110–2117.
- Kletzin, A. (2007) General characteristics and important model organisms. In *Archaea: Molecular and Cellular Biology*, ed. by Cavicchioli, R., ASM Press, Washington, D.C., pp. 14–92.
- Marmur, J. and Doty, P. (1962) Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J. Mol. Biol.*, **5**, 109–118.
- Randau, L., Münch, R., Hohn, M. J., Jahn, D., and Söll, D. (2005) *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5'- and 3'-halves. *Nature*, **433**, 537–541.
- Sowers, K. R. (1995) Methanogenic archaea: An overview. In *Archaea: A Laboratory Manual*, ed. by Sowers, K. R. and Schreier, H. J., Cold Spring Harbor Laboratory Press, Plainville, New York, pp. 3–14.
- Stetter, K. O. (1999) Extremophiles and their adaptation to hot environments. *FEBS Lett.*, **452**, 22–25.
- Woese, C. and Fox, G. (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. USA*, **74**, 5088–5090.