Mini review of an archaeon

Mini review of *Methanothermococcus okinawensis* genome

Suvorova A. A.[1]

[1]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University

Abstract

Methanothermococcus okinawensis is a thermophilic methane-producing archaeon. In this review there is an attempt to retrieve and analyze some information from its genome such as nucleotide composition, statistical data about its proteome and ribosomal genes.

Contact: cringebutfree@fbb.ru

1 Introduction

Methanothermococcus okinawensis is a thermophilic, methane-producing archaeon first isolated from a deep-sea hydrothermal vent chimney at the Iheya Ridge, in the Okinawa Trough, Japan, in 2000. It appeared to be strictly an anaerobic, mesophilic autotroph that uses hydrogen as a source of electrons, carbon dioxide or formate as a source of carbon and electron acceptor and ammonium as a nitrogen source.[1]

Such organisms with unique biochemical pathways are important in fundamental research, including the ones about extraterrestrial life,[2] and can be applied in biotechnology.[3]

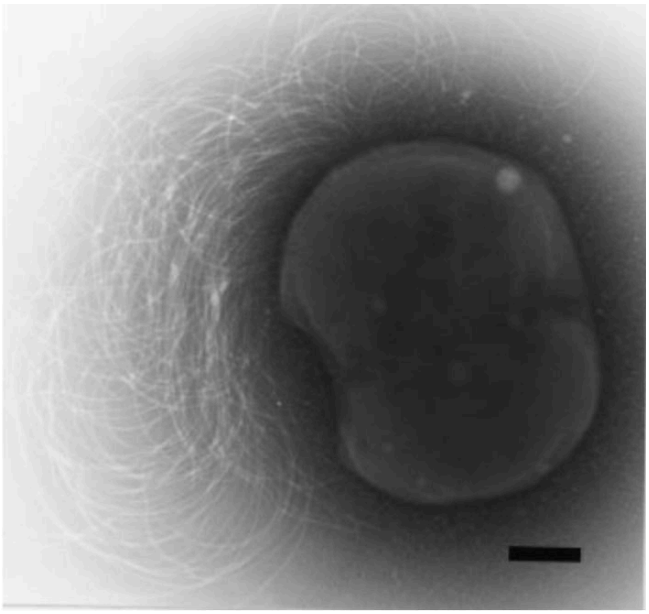

Fig. 1. Electron micrograph of negatively stained cells. The polar bundle of flagella is observed. Bar, 400 nm *(Ken Takai et al. 2002)*

2 Methods

Python programming: Google Colab notebook with python codes used here.

Excel: table with histogram showing distribution of proteins by length.

3 Results

3.1 Standard data about *Metanothermococcus okinawensis* genome

Metanothermococcus okinawensis has 2 DNA sequences: a chromosome and a plasmid pMETOK01. Frequencies of nucleotides were counted and are shown in Table 1. According to them one can see that Chargaff's rules work but are more accurate for longer sequences.

Also, frequencies of nucleotides and GC content were counted separately for CDS (coding sequence) and not CDS – see Table 2. GC content is much higher in CDS, but still less than 0.5, which would be expected if nucleotides appeared randomly with equal probability 0.25. One of the reasons is different probability of mutations from one nucleotide to another. Quite often occurs such mutation as methylcytosine deamination that results in forming thymine. Cytosine can be modified in CpG dinucleotides including methylation and this mutation is difficult for DNA repair systems.[4] As a result, frequencies of cytosine and guanine are lower than those of adenine and thymine and GC content is less than 0.5. However, in CDS stabilizing selection is more present than in not CDS, so GC content is higher in CDS.

3.2 Statistical data about *M. okinawensis* proteome

3.2.1 Distribution of proteins by length

As can be seen from Fig. 2 most of proteins have length between 50 and 450 amino acid residues.

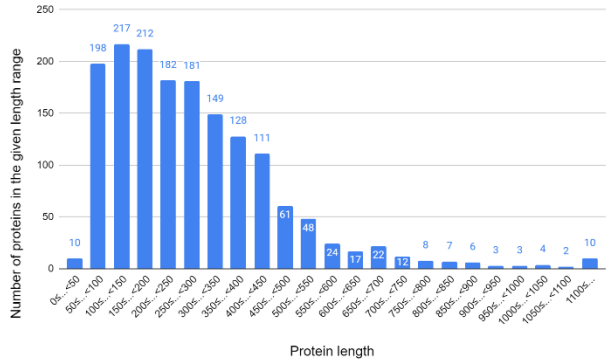

Fig. 2. Distribution of proteins by length

3.2.2 Number of genes, coded on direct and complementary? DNA chains

The null hypothesis H_0 is that genes are randomly distributed between direct and complementary chains with equal probabilities. Chi square equals 0.1402 with p-value 0.7081 that is much more than 0.05, so it is very likely that the null hypothesis is correct.

Table 3. Number of genes coded on direct (+) and complementary (-) DNA chains of *M. okinawensis*

Sequence name	+	-
Methanothermococcus okinawensis IH1, complete sequence	795	810
Methanothermococcus okinawensis IH1 plasmid pMETOK01, complete sequence	4	6

Genes are approximately evenly distributed between + and – DNA chains.

3.2.3 Number of ribosomal, hypothetical and transport proteins

Table 4. Number of different types of proteins in *M. okinawensis* proteome

Type of proteins	Number	Proportion in total number of protein
Ribosomal proteins	63	0.0382
Hypothetical proteins (in chromosome)	287	0.1752
Hypothetical proteins (in plasmid)	5	0.5
Transport proteins	67	0.0409

3.3 Statistical data about ribosomal genes

The number of RNA coding genes is much lower than the number of protein coding genes. Except for tRNA coding genes there are two other non-coding RNA (ncRNA) genes: ribonuclease P (RNase P) – a ribozyme that cleaves tRNA precursor molecules,[5] and signal recognition particle RNA (SRP RNA) that recognizes signal peptide of membrane or secretory proteins and then associates with SRP receptor anchored to endoplasmic reticulum membrane.[6]

Table 5. Number of RNA coding genes in *M. okinawensis* compared to the number of protein coding genes

Type of gene product	Number of genes
Protein coding genes	1615
All RNA coding genes	47
rRNA coding genes	7
tRNA coding genes	38

3.4 Cumulative GC skew

Cumulative GC skew for the chromosome and for the plasmid are shown in Fig. 3 and Fig. 4 respectively.
Minimum in cumulative GC skew for the plasmid corresponds with minimum in GC content that is shown in Fig. 5. Peaks on this graph must be referring to genes, while the lowest GC content probably marks origin of replication.

References

1. Ken Takai et al. (2002) Methanothermococcus okinawensis sp. nov., a thermophilic, methane-producing archaeon isolated from a Western Pacific deep-sea hydrothermal vent system doi: 10.1099/ijs.0.02106-0
2. Ruth-Sophie Taubner et al. (2018) Biological methane production under putative Enceladus-like conditions doi: 10.1038/s41467-018-02876-y
3. Lisa-Maria Mauerhofer et al. (2021) Hyperthermophilic methanogenic archaea act as high-pressure CH4 cell factories doi: 10.1038/s42003-021-01828-5
4. Akira Sassa et al. (2016) M. Mutagenic consequences of cytosine alterations site-specifically embedded in the human genome doi: 10.1186/s41021-016-0045-9
5. CeciliaGuerrier-Takada et al. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme doi:10.1016/0092-8674(83)90117-4

6. Kiyoshi Nagai (2003) Structure, function and evolution of the signal recognition particle https://doi.org/10.1093/emboj/cdg337

Sequence name	Sequence length, bp	A
Methanothermococcus okinawensis IH1, complete sequence	1662525	0.35
Methanothermococcus okinawensis IH1 plasmid pMETOK01, complete sequence	14930	0.39699	0.334

Table 1. Frequencies of nucleotides and GC content in *M. okinawensis* chromosome and in plasmid pMETOK01

Table 2. Frequencies of nucleotides and GC content in and out of CDS

Sequence name	CDS/not CDS	Nucleotide frequencies				GC content
		A	T	G	C	
Methanothermococcus okinawensis IH1, complete sequence	CDS	0.34268	0.34416	0.15585	0.15731	0.31316
	Not CDS	0.40748	0.40349	0.09494	0.09409	0.18903
Methanothermococcus okinawensis IH1 plasmid pMETOK01, complete sequence	CDS	0.38823	0.33424	0.15203	0.12550	0.27753
	Not CDS	0.35295	0.35395	0.14591	0.14719	0.29310

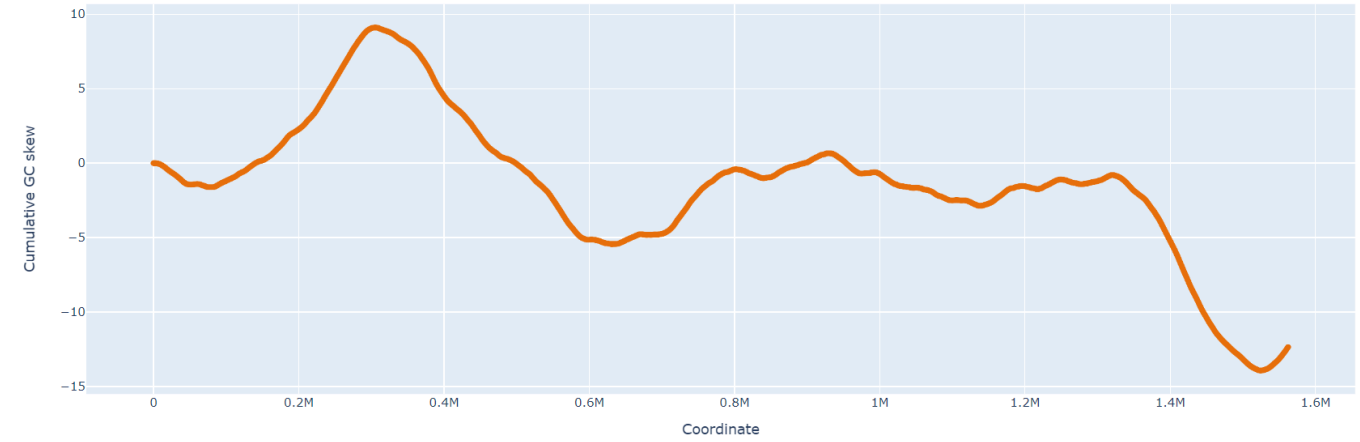

Fig. 3. Cumulative GC skew in *Methanothermococcus okinawensis* chromosome. Window length – 100000 nucleotides, step – 1000 nucleotides. Minimal value is -13.92554, in coordinate 1524000, maximal value is 9.09971, in coordinate 305000.

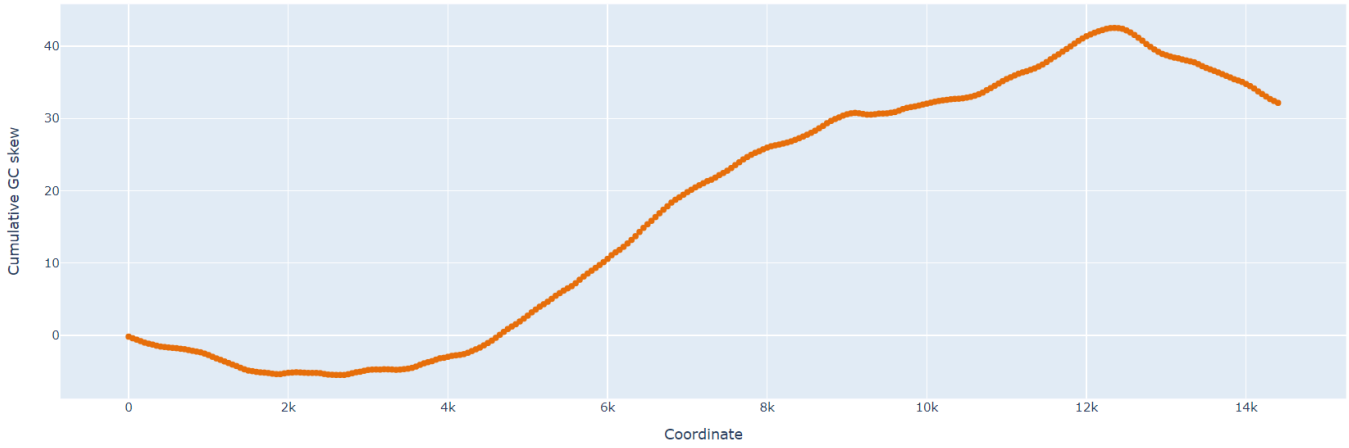

Fig. 4. Cumulative GC skew in *Methanothermococcus okinawensis* plasmid pMETOK01. Window length – 500 nucleotides, step – 50 nucleotides. Minimal value is -5.46870, in coordinate 2600, maximal value is 42.52213, in coordinate 12350.

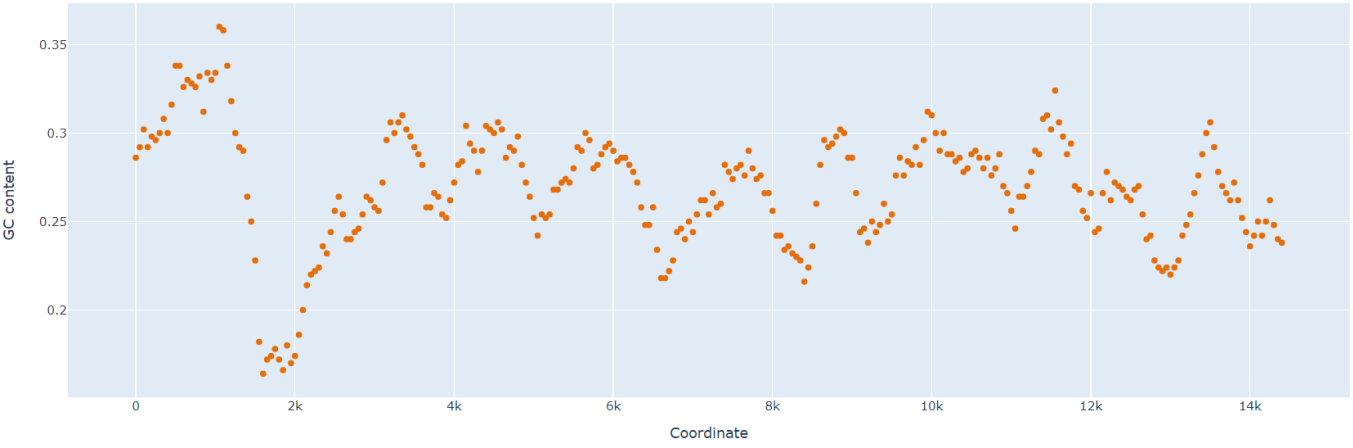

Fig. 4. GC content in *Methanothermococcus okinawensis* plasmid pMETOK01. Window length – 500 nucleotides, step – 50 nucleotides. Minimal value is - 0.164, in coordinate 1600, maximal value is 0.36, in coordinate 1050.