



HAL
open science

A methanogen hosted the origin of the genetic code

Massimo Di Giulio

► **To cite this version:**

Massimo Di Giulio. A methanogen hosted the origin of the genetic code. *Journal of Theoretical Biology*, 2009, 260 (1), pp.77. 10.1016/j.jtbi.2009.05.030 . hal-00554620

HAL Id: hal-00554620

<https://hal.science/hal-00554620>

Submitted on 11 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

A methanogen hosted the origin of the genetic code

Massimo Di Giulio

PII: S0022-5193(09)00257-4
DOI: doi:10.1016/j.jtbi.2009.05.030
Reference: YJTBI5587

To appear in: *Journal of Theoretical Biology*

Received date: 20 April 2009
Revised date: 26 May 2009
Accepted date: 29 May 2009



www.elsevier.com/locate/jtbi

Cite this article as: Massimo Di Giulio, A methanogen hosted the origin of the genetic code, *Journal of Theoretical Biology*, doi:[10.1016/j.jtbi.2009.05.030](https://doi.org/10.1016/j.jtbi.2009.05.030)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A methanogen hosted the origin of the genetic code

Massimo Di Giulio

Laboratory for Molecular Evolution, Institute of Genetics and Biophysics 'Adriano Buzzati Traverso', CNR, Via P. Castellino, 111, 80131 Naples, Napoli, Italy

Address for correspondence: Dr. M. Di Giulio, Laboratory for Molecular Evolution, Institute of Genetics and Biophysics 'Adriano Buzzati Traverso', CNR, Via P. Castellino, 111, 80131 Naples, Napoli, Italy

e-mail: digiulio@igb.cnr.it

Fax Number: +39 081 6132706

Telephone Number: +39 081 6132369

Abstract

A comparison is made between orthologous proteins from a methanogen (*Methanopyrus kandleri*) and from a non-methanogen (*Pyrococcus abyssi*) in order to determine the amino acid substitution pattern. This analysis makes it possible to establish which amino acids are significantly and asymmetrically utilised by these two organisms. A methanophily index (MI) based on this asymmetry makes it possible for any protein to be associated with a numerical value which, when calculated for the same orthologous protein from methanogenic and non-methanogenic organisms, turns out to have the power to discriminate between these two groups of organisms, even if only for about 20% of the analysed proteins. The MI can also be associated to the genetic code under the assumption that the frequency of synonymous codons specifying the amino acids in the genetic code also reflects the frequency with which amino acids appeared in ancestral proteins. Finally a t test shows that the MI value associated to the genetic code is not different from the mean value of the MI deriving from methanogen proteins, but it differs from the mean MI of non-methanogen proteins. This might indicate that the genetic code evolved in a methanogenic 'organism'.

Keywords: amino acid substitution pattern – LUCA – timing of methanogenesis – biological dating

1 Introduction

1.1. Geological and biological methods aiming to establish the antiquity of biological processes: the example of methanogenesis

Geological evidence suggest that methanogenesis was one of the earliest biological processes to take place on earth (Brocks et al., 1999; Ueno et al., 2006). Nevertheless, it is unclear how ‘early’ this origin was because, as geological dating indicates methanogenesis taking place approximately 3.5 billion years ago (Ueno et al., 2006), methanogenesis might actually be biologically later in the sense that the very first forms of life, and in particular the last universal common ancestor (LUCA), might be non-methanogenic ‘organisms’. It is therefore clear that geological fossils need to be accompanied by biological evidence if we are to more accurately define the timing of biological processes.

Phylogenetic methods, for instance, have been used in an attempt to define the (hyper)thermophilic or mesophilic nature of the LUCA by exploiting the correlation between optimal growth temperature and the G+C content of ribosomal RNA and some protein indices (Galtier et al., 1999; Di Giulio, 2001a, 2000b, 2001, 2003a, 2003b; Boussau et al., 2008). Subsequently, by phylogenetically reconstructing the ancestral sequences of the LUCA, it was determined whether these were more typical of mesophiles or (hyper)thermophiles (Galtier et al., 1999; Di Giulio, 2000a, 2000b, 2001, 2003a, 2003b; Boussau et al., 2008).

Whereas, by exploiting the invariance and the antiquity of the genetic code, methods and ideas were introduced to enable an investigation into the physical environment in which the genetic code was structured (Di Giulio, 2000, 2005b, 2005c; Archetti and Di Giulio, 2007). This was essentially based on the assumption that the frequency with which synonymous codons specifying the amino acids appear in the genetic code also reflects the frequency with which these were used in ancestral proteins. By subsequently constructing amino acid indices derived from the comparison of orthologous proteins from two organisms living in environments with a different characteristic, it was possible to furnish evidence in

favour of a hyperthermophilic, barophilic, anaerobic and low pH primordial setting (Di Giulio, 2000, 2005b, 2005c; Archetti and Di Giulio, 2007).

Here, these methods (Di Giulio, 2000, 2005a, 2005b, 2005c; Archetti and Di Giulio, 2007) are used to attempt to establish whether the genetic code originated in a methanogenic or non-methanogenic organism.

2. Materials and Methods

All the proteins used in the analysis were taken from the NCBI using BLASTP (Altschul et al., 1997). Two or more proteins were aligned using CLUSTALX (Thompson et al., 1997). Only highly conserved regions were used in the analysis, while poorly conserved regions or regions containing gaps were eliminated from this alignment.

For each amino acid (Tab. 2) or for each pair of amino acids (Tab. 3) the significant deviation from the expected theoretical ratio of 50:50 was determined by calculating the exact binomial probability.

When not otherwise specified, the methods and ideas referred in equivalent analyses hold (Haney et al., 1999; McDonald et al., 1999; Di Giulio, 2000, 2005a, 2005b, 2005c; Archetti and Di Giulio, 2007).

3. Results and Discussion

3.1 The amino acid substitution pattern in the presence/absence of methane

In order to obtain information on the amino acid substitution pattern between a methanogen and a non-methanogen, I have compared orthologous proteins from *Pyrococcus abyssi*, a non-methanogenic archaeon and *Methanopyrus kandleri*, a methanogenic archaeon. These two organisms were chosen because they seem to possess the majority of equivalent physicochemical variables (temperature, pressure, etc.) but differ primarily in the

absence/presence of methane; therefore, the amino acid substitution pattern deriving from the comparison of their proteins should be subject to the effects of this molecule (McDonald et al., 1999; Di Giulio, 2005a).

I then compared 140 proteins from *P. abyssi* and *M. kandleri* for a total of 35,095 amino acids (Tab. 1). This sample seems to be representative of the amino acid substitution pattern because it presents a total number of variable amino acid positions equal to 12,461 (Tab. 1) and with an identity percentage of 64.5%.

Table 2 shows how the total amino acid substitutions for a single amino acid are distributed over the two compared organisms. Table 3, on the other hand, only reports the statistically significant deviations from the expected theoretical ratio of 50:50 of the single amino acid substitutions in the sample of all the amino acid substitutions (Tab. 1).

Equivalent analyses have already been conducted in a similar way for other variables (Haney et al., 1999; McDonald et al., 1999; Di Giulio, 2000, 2005a, 2005b, 2005c; Archetti and Di Giulio, 2007).

3.2 The construction of a methanophily index

The comparison between the proteins of a methanogen and a non-methanogen (Tab. 1) makes it possible to establish which amino acids are statistically and significantly preferred by the methanogen and which are not (Tab. 2). Then, by associating every amino acid with a rank established simply on the basis of the probability of deviation from the expected theoretical ratio of 50:50 (Tab. 2), we can define a methanophily index (MI) as follows:

$$MI = \sum_{j=1}^N R_j/N,$$

where R_j is the methanophily rank (Tab. 2) of the j -th amino acid and N is the protein's amino acid length. This has already been done for other variables (Di Giulio, 2000a, 2005b, 2005c; Archetti and Di Giulio, 2007).

3.3 For some proteins, the methanophily index can distinguish methanogen proteins from non-methanogen proteins

It is not easy to check whether the methanophily index (MI) which can be associated to each protein has the power to distinguish methanogen proteins from non-methanogen proteins because it is not known which amino acids are preferentially used by these two groups of organisms. The only exception regards the amino acid cysteine, for which there are indications that it is particularly used in methanogens (Klipcan et al., 2008), which is compatible with the high rank that cysteine has in Tab. 2.

Therefore, the only means we have of checking whether the MI can distinguish between methanogen and non-methanogen proteins is to calculate the MI values for a sample of the same orthologous protein from these two groups of organisms. I have therefore conducted this analysis for 31 different orthologous proteins (Tab. 4), that is, for every orthologous protein I have built a multiple alignment and calculated the MI values for the groups of methanogens and non-methanogens (Tab. 4). It emerged that, for 6 out of 31 proteins, an unpaired t test furnishes statistically highly significant results (top of Tab. 4), that is, the MI can distinguish between proteins from methanogens and those from non-methanogens. In other words, the comparison of proteins from a methanogenic and a non-methanogenic organism can produce an index capable of discriminating between these organisms, albeit only for about 20% of the analysed proteins (Tab. 4).

Another limitation was identified in this analysis. For four observations, an 'inverted' significance was detected, that is to say that the mean of the MI values of the non-methanogen group is statistically and significantly higher than that of methanogens (bottom

of Tab. 4). In these cases, the MI evidently measures the opposite of what it is required to measure.

In order to clarify this point, I have broken down the significance into its components, introducing into the unpaired t test not two groups (methanogens and non-methanogens) but four (methanogenic Archaea, non-methanogenic Archaea, Bacteria and Eukarya) in order to understand if there is an effect linked to the domains of life. Under this new condition, the unpaired t test gives, for the elongation factor-Tu, a high mean MI value (=11.962) for the sequences of Bacteria such as to invert the test significance (Tab. 5). Whereas, the t test is no longer significant for histidyl-tRNA synthetase (data not shown) while, for seryl-tRNA synthetase, the inverted significance of the t test still depends on the very high mean MI value (=11.455) for the sequences of Bacteria (data not shown). Also for phosphoribosylamine-glycine ligase a similar behaviour is observed with very high mean MI values for the sequences of Bacteria and Eukarya, and a low mean MI value for methanogenic Archaea (data not shown). Therefore, the MI is also subject to an effect due to the three domains of life.

This urges us to conduct more thorough investigations into the six observations that seem to give the MI the power to discriminate between methanogen and non-methanogen groups (top of Tab. 4). As far as glycyl tRNA synthetase is concerned, Tab. 6 clearly shows that the significance of the test is primarily due to the high mean value of the methanogens' MI, even though the mean MI values for the methanogenic Archaea and the non-methanogenic Archaea are different only at the level of 15% significance (Tab. 6). Whereas, the mean MI values in the latter two groups are significantly different for the thermosome sequences (Tab. 7), thus indicating that, although there is an effect due to the domains of life (see the statistical significance between the sequences of Bacteria and Eukarya (Tab. 7); note also that although the mean MI value for Bacteria is high and lowers the overall significance of the t test between methanogens and non-methanogens, it cannot jeopardise it (Tab. 4 and 7)), the overall significance of the test is, however, dependent upon the MI's discriminatory power (Tab. 7). Equivalent considerations can also be made for the remaining four proteins,

which have a power to discriminate between the sequences of methanogens and those of non-methanogens (data not shown; top of Tab. 4).

The conclusion is that these six proteins (top of Tab. 4) make it possible to estimate the mean value of the MI characterising the proteins of both methanogens and non-methanogens because, unlike the four in which significance is inverted (bottom of Tab. 4), in these six proteins the behaviour of MI can be associated to that of methanogens and non-methanogens (Tab. 6 and 7) while the domains of life might be responsible for the inverted behaviour of the four proteins (Tab. 5).

3.4 The genetic code might have originated in a methanogen

The mean protein that can be associated to the genetic code on the basis of the number of synonymous codons that the code attributes to amino acids has a methanophily index (MI_{code}) equal to 11.328. Obviously, in order to calculate this value, Met for instance, which has a single codon in the genetic code, has been attributed with a frequency in ancestral proteins of 1/61, while Ser, which has six codons in the code, has been attributed with a frequency of 6/61 (for a justification of this assumption, see Di Giulio (2000a)). Therefore, it is possible to test whether the value of $MI_{code} = 11.328$ is typical of proteins of methanogens or non-methanogens. In order to do this, we have to estimate the mean MI for the proteins in these two groups of organisms. This has been done using only the six observations in the top part of Tab. 4. The mean MI value of methanogens is equal to $MI_{mean} = 11.256$, and that of non-methanogens is $MI_{mean} = 10.914$, which are clearly seen to be different in a paired t test (mean diff. = +0.342, $df=5$, $t=+7.459$, $P=0.0007$), while in the more relevant unpaired t test, the difference between the two groups is only marginally significant (mean diff. = +0.342, $df=10$, $t=+1.982$, $P=0.076$). However, the crucial test (Blaam, 1972; Di Giulio, 2000a, 2005b, 2005c; Archetti and Di Giulio, 2007) for establishing whether or not these two means are different from the value $MI_{code} = 11.328$ of the mean ancestral protein has determined that, while the $MI_{mean} = 11.256$ of methanogens is not different from the MI_{code} ($t=-0.6729$, $df=5$,

0.50<P<0.60), that of non-methanogens ($MI_{\text{mean}} = 10.914$) is different from MI_{code} ($t = -3.067$, $df = 5$, $0.02 < P < 0.05$). This indicates that the genetic code might have originated in a methanogen because the mean of the MI values for methanogen proteins is not different from that associated to the genetic code, whereas the mean of the MI values for non-methanogen proteins is different from that derived from the genetic code.

4. Conclusions

The comparison of proteins from a methanogenic and a non-methanogenic organism is such as to produce a methanophily index capable of discriminating between these two groups of organisms, even if only 20% of the analysed proteins are sensitive to this index. This shows that methane influenced the amino acid substitution pattern in these two organisms (Tab. 1, 2, and 3).

The use of this finding in order to establish whether the genetic code evolved in a methanogen or a non-methanogen furnishes evidence in favour of the hypothesis that methanogenesis is an extremely ancient pathway because the genetic code seems to have been structured in a methanogen. This is compatible with the suggestion that methanogenesis is a very early pathway in the history of life (Brocks et al., 1999; Battistuzzi et al., 2004; Baptiste et al., 2005; Ueno et al., 2006).

Finally, this observation also corroborates the hypothesis that the LUCA was a methanogen (Xue et al., 2005; Wong et al., 2007).

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acid. Res.* 25, 3389-3402.

Archetti, M., Di Giulio, M. 2007. The evolution of the genetic code took place in an anaerobic environment. *J. Theor. Biol.* 245, 169-174.

Bapteste, E., Brochier, C., Boucher, Y., 2005. Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea* 1, 353-363.

Battistuzzi, F.U., Feljao, A., Hedges, S. B., 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol. Biol.* 4, 44.

Balaam, L. N. 1972. *Fundamentals of biometry*. George Allen & Unwin. London, pp. 120-142.

Boussau, B., Blanquart, S., Necsulea, A., Lartillot, N., Gouy, M., 2008. Parallel adaptations to high temperature in the Archaean eon. *Nature* 456, 942-945.

Brocks, J.J., Logan, G.A., Buick, R., Summons, R.E. 1999. Archean molecular fossils and the early rise of eukaryotes. *Science* 285, 1033-1036.

Di Giulio, M. 2000a. The late stage of genetic code structuring took place at a high temperature. *Gene* 261, 189-195.

Di Giulio, M., 2000b. The universal ancestor lived in a thermophilic or hyperthermophilic environment. *J. Theor. Biol.* 203, 203-213.

Di Giulio, M. 2001. The universal ancestor was a thermophile or a hyperthermophile. *Gene* 281, 11-17.

Di Giulio, M. 2003a. The universal ancestor was a thermophile or a hyperthermophile: tests and further evidence. *J. Theor. Biol.* 221, 425-436.

Di Giulio, M. 2003b. The universal ancestor and the ancestor of Bacteria were hyperthermophiles. *J. Mol. Evol.* 57, 721-730.

Di Giulio, M. 2005a. A comparison of protein from *Pyrococcus furiosus* and *Pyrococcus abyssi*: bariphily in the physicochemical properties of amino acids and in the genetic code. *Gene* 346, 1-6.

Di Giulio, M. 2005b. The ocean abysses witnessed the origin of the genetic code. *Gene* 346, 7-12.

Di Giulio, M. 2005c. Structuring of the genetic code took place at acidic pH. *J. Theor. Biol.* 237, 219-226.

Galtier, N., Tourasse, N., Gouy, M. 1999. A nonhyperthermophilic common ancestor to extant life forms. *Science* 283, 981-987.

- Haney, P. J., Badger, J. H., Buldak, G. L., Reich, C. I., Woese, C. R., Olsen, G. J. 1999. Thermal adaptation analyzed by comparison of proteins sequences from mesophilic and extremely thermophilic Methanococcus species Proc. Natl. Acad. Sci. USA 96, 3578-3583.
- Klipcan, L., Frenkel-Morgenstern, M., Safro, M. G., 2008. Presence of tRNA-dependent pathways correlates with high cysteine content in methanogenic Archaea. Trends Genet. 24, 59-63.
- McDonald, J. H., Grasso, A. M., Rejto, L. K. 1999. Patterns of temperature adaptation in proteins from Methanococcus and Bacillus. Mol. Biol. Evol. 16, 1785-1790.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., Higgins, D. G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876-4882.
- Ueno, Y., Yamada, K., Yoshida, S., Maruyama, S., Isozaki, Y., 2006. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. Nature 440, 516-519.
- Wong, J. T., Chen, J., Mat, W. K., Xue, H. 2007. Polyphasic evidence delineating the root of life and roots of biological domains. Gene 403, 39-52.
- Xue, H. , Ng, S. K., Tong, K. L., Wong, J. T. 2005. Congruence of evidence for Methanopyrus-proximal root based on transfer RNA and aminoacyl-tRNA synthetase genes. Gene 360, 120-130.

Legend to the Tables

Table 1

The matrix shows amino acid substitutions between *Pyrococcus abyssi* (non-methanogen) and *Methanopyrus kandleri* (methanogen). For every row and column, the table also shows the relative sum but not including the diagonal element. See text for further information.

Table 2

Sum of all the amino acid substitutions involving a single amino acid as identified in Tab. 1. The substitution direction is: non-methanogenic amino acid -> methanogenic amino acid. The highest ranks refer to 'methanogenic' amino acids. See text for further information.

Table 3

Deviation from the theoretical expected ratio 50:50 of the single pairs of amino acids observed in Tab. 1. The first amino acid refers to the one present in the non-methanogenic organism, while the second to the methanogenic organism. For instance, AC = 65 indicates that 65 alanines (A) in the non-methanogen have been replaced in the methanogen in the same number of cysteines (C).

Table 4

This shows: (i) the alignment length (aln); (ii) the number of proteins used (n); (iii) the mean and standard deviation of the methanophily index value (MI) of the proteins from methanogens and non-methanogens; (iv) the difference between the mean value of the MI of methanogens and non-methanogens (mean diff.); (v) the t test value and the relative probability. The proteins are arranged in decreasing order of significance of the discriminatory power of MI. See text for further information.

Table 5

Results of the unpaired t test of the elongation factor-Tu for the differences between the means of MIs for the four groups: a= Eukarya; b = non-methanogenic Archaea; c= methanogenic Archaea;
d = Bacteria

Table 6

Results of the unpaired t test of the glycyl-tRNA synthetase for the differences between the means of MIs for the four groups: a= non-methanogenic Archaea; b = Bacteria; c = non-methanogenic Archaea; d = Eukarya.

Table 7

Results of the unpaired t test of the thermosome proteins for the differences between the means of MIs for the four groups: a= non-methanogenic Archaea; b = Bacteria; c = non-methanogenic Archaea; d = Eukarya.

P. abyssi

M. kandleri

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y		
A	2091	65	24	76	10	114	11	33	30	37	19	13	53	9	45	112	79	158	3	14	905	
C	7	162	0	0	1	3	0	0	0	3	0	0	0	1	0	0	3	3	7	0	0	28
D	20	7	1343	236	3	38	22	3	24	5	3	42	12	13	38	23	18	2	2	2	4	515
E	56	3	260	2132	4	39	21	9	133	25	6	23	43	56	147	47	42	25	3	13	955	
F	20	6	6	14	706	6	18	42	7	127	26	6	15	5	14	2	7	51	18	114	504	
G	126	12	62	40	1	2579	5	4	10	5	2	21	12	2	28	38	20	7	2	2	399	
H	6	2	17	25	4	11	507	1	19	7	2	12	8	15	32	12	8	4	4	20	209	
I	65	11	5	46	44	5	7	1366	27	341	48	4	22	9	36	11	42	671	10	30	1434	
K	56	4	117	413	5	63	33	13	924	40	14	27	46	49	403	57	43	46	1	20	1450	
L	68	11	5	50	81	10	18	238	33	1978	108	9	18	21	45	8	45	264	6	33	1071	
M	32	11	5	22	15	6	5	46	16	106	410	1	4	12	27	9	18	56	4	7	402	
N	18	8	130	73	6	49	31	5	34	10	4	527	12	16	38	44	26	9	1	8	522	
P	34	2	25	64	5	12	6	10	13	10	0	3	1499	6	19	20	9	22	2	4	266	
Q	14	2	22	85	0	4	27	1	39	18	11	5	6	412	49	12	11	11	0	3	320	
R	26	3	48	156	6	35	12	10	194	19	8	24	28	32	1465	29	30	28	2	11	701	
S	144	20	40	70	3	50	5	8	21	10	5	25	25	11	40	689	117	18	1	1	614	
T	76	18	28	61	7	13	10	26	32	25	9	17	27	9	34	80	1032	87	3	7	569	
V	168	30	9	49	33	10	9	263	24	184	36	5	28	9	36	19	85	1945	2	23	1022	
W	8	0	7	8	21	1	9	7	3	15	3	1	1	2	12	1	3	4	199	21	127	
Y	31	4	13	30	89	12	52	18	20	44	13	7	10	2	31	13	12	33	14	668	448	
	975	219	823	1518	338	481	301	737	679	1031	317	245	371	278	1074	540	618	1503	78	335		

*P. abyssi**M. kandleri*

Substitution Direction: non-methanogenic-AA->methanogenic-AA		P	Rank
AAAs->E=1518	E->AAAs=955	<0.000001	20
AAAs->V=1503	V->AAAs=1022	<0.000001	19
AAAs->R=1074	R->AAAs=701	<0.000001	18
AAAs->D=823	D->AAAs=515	<0.000001	17
AAAs->C=219	C->AAAs=28	<0.000001	16
AAAs->P=371	P->AAAs=266	0.000036	15
AAAs->H=301	H->AAAs=209	0.000054	14
AAAs->G=481	G->AAAs=399	0.0063	13
AAAs->A=975	A->AAAs=905	0.11	10.5
AAAs->T=618	T->AAAs=569	0.16	10.5
AAAs->L=1031	L->AAAs=1071	0.40	10.5
AAAs->Q=278	Q->AAAs=320	0.094	10.5
AAAs->S=540	S->AAAs=614	0.032	8
AAAs->M=317	M->AAAs=402	0.0017	7
AAAs->W=78	W->AAAs=127	0.00076	6
AAAs->Y=335	Y->AAAs=448	0.000061	5
AAAs->F=338	F->AAAs=504	<0.000001	4
AAAs->N=245	N->AAAs=522	<0.000001	3
AAAs->I=737	I->AAAs=1434	<0.000001	2
AAAs->K=679	K->AAAs=1450	<0.000001	1

Table 2

		Probability
AC=65	CA=7	<0.000001
AI=33	IA=65	0.0016
AK=30	KA=56	0.0068
AL=37	LA=68	0.0032
AR=45	RA=26	0.032
AY=14	YA=31	0.016
CG=3	GC=12	0.035
CS=3	SC=20	0.00049
CT=3	TC=18	0.0014
CV=7	VC=30	0.00019
CD=0	DC=7	0.016
CI=0	IC=11	0.00098
CM=0	MC=11	0.00098
CN=0	NC=8	0.0078
ME=22	EM=6	0.0037
MR=27	RM=8	0.0019
MV=56	VM=36	0.047
RH=12	HR=32	0.0036
RI=10	IR=36	0.00016
RK=194	KR=403	<0.000001
RL=19	LR=45	0.0016
RW=2	WR=12	0.013
RY=11	YR=31	0.0029
KD=117	DK=24	<0.000001
KE=413	EK=133	<0.000001
KG=63	GK=10	<0.000001
KI=13	IK=27	0.038

KP=46	PK=13	0.000019
KS=57	SK=21	0.000056
KV=46	VK=24	0.012
GD=62	DG=38	0.021
GN=21	NG=49	0.0011
GY=2	YG=12	0.013
YD=13	DY=4	0.049
YE=30	EY=13	0.014
YH=52	HY=20	0.00021
YS=13	SY=1	0.0018
SD=40	DS=23	0.043
SE=70	ES=47	0.042
SN=25	NS=44	0.029
ST=117	TS=80	0.010
SY=1	YS=13	0.0018
LE=50	EL=25	0.0052
LF=81	FL=127	0.0017
LH=18	HL=7	0.043
LI=238	IL=341	0.000021
LT=45	TL=25	0.022
LV=264	VL=184	0.00018
VE=49	DV=25	0.0071
VI=263	IV=671	<0.000001
IC=11	CI=0	0.00098
IE=46	EI=9	0.000001
IQ=9	QI=1	0.021
FE=14	EF=4	0.034
FH=18	HF=4	0.0043

PN=3	NP=12	0.035
PD=25	DP=12	0.047
HN=12	NH=31	0.0054
NE=73	EN=23	<0.000001
ND=130	DN=42	<0.000001
NQ=16	QN=5	0.027
EQ=56	QE=85	0.018

Table 3

Accepted manuscript

5. Tables

Protein	aln	Methanogens		non-Methanogens		t	P
		n	mean	n	mean		
Arginyl-tRNA synthetase	162	17	11.072	58	10.542	+3.721	0.00040
Release Factor	275	17	10.892	40	10.578	+3.332	0.0015
Glycyl-tRNA synthetase	279	16	11.345	43	10.926	+3.000	0.0040
Alanyl-tRNA synthetase	355	16	11.412	57	11.146	+2.361	0.021
Threonyl-tRNA synthetase	223	15	11.182	60	10.889	+2.285	0.025
Thermosome	323	16	11.633	52	11.404	+2.149	0.035
Methionyl-tRNA synthetase	271	14	11.011	43	10.786	+1.597	0.12
Phenylalanyl-tRNA synt. beta	283	17	11.153	68	11.029	+1.451	0.15
Elongation factor 2	447	18	11.635	48	11.518	+1.447	0.15
ATPase subunit A	440	15	11.576	41	11.465	+1.185	0.24
Enolase	337	11	11.085	51	10.957	+1.132	0.26
Inosine-5' monoP dehydrogenase	266	12	11.292	49	11.138	+1.012	0.32
Leucyl-tRNA synthetase	209	17	10.482	57	10.421	+0.735	0.46
Valyl-tRNA synthetase	380	15	10.914	57	10.848	+0.574	0.57
Glutamyl-tRNA synthetase	233	17	11.082	62	11.026	+0.472	0.64
Phenylalanyl-tRNA synt.alpha	184	17	11.226	45	11.208	+0.120	0.90
Tryptophan synthase	324	15	11.242	32	11.225	+0.101	0.92
Isoleucyl-tRNA synthetase	426	17	10.869	58	10.857	+0.092	0.93
S-adenosylHomocysteine	347	14	11.003	38	11.049	-0.385	0.70
Cell division protein	241	12	11.442	53	11.516	-0.641	0.52
Tryptophanyl-tRNA synthetase	128	17	10.479	55	10.625	-0.700	0.49
Succinyl-CoA	272	10	10.803	31	10.953	-0.746	0.46
Signal recognition particle 54D	298	17	10.671	64	10.830	-0.982	0.33
CTP synthase	378	11	11.011	38	11.141	-1.023	0.31
Initiator factor-2	253	16	11.234	57	11.378	-1.129	0.26
Methionine adenosyltransferase	124	17	11.125	59	11.242	-1.238	0.22
Topoisomerase I	204	16	10.649	58	10.861	-1.576	0.12
Histidyl-tRNA synthetase	225	15	10.849	70	11.217	-2.267	0.026
Seryl-tRNA synthetase	298	17	10.837	61	11.172	-2.471	0.016
P-ribosylamine-glycine ligase	208	16	11.002	35	11.481	-2.486	0.016
Elongation factor-Tu	317	16	11.138	47	11.500	-2.948	0.0045

	Mean Diff.	DF	t-Value	P-Value
a, b	-,119	25	-1,015	,3200
a, c	-,047	26	-,625	,5377
a, d	-,870	30	-19,941	<,0001
b, c	,072	29	,623	,5382
b, d	-,752	33	-8,131	<,0001
c, d	-,824	34	-13,422	<,0001

	Count	Mean	Variance	Std. Dev.	Std. Err
a	12	11,091	,013	,115	,033
b	15	11,210	,153	,391	,101
c	16	11,138	,057	,239	,060
d	20	11,962	,015	,122	,027

Table 5

	Mean Diff.	DF	t-Value	P-Value
a, b	,310	29	1,612	,1178
a, d	,237	23	1,490	,1497
a, c	-,223	27	-1,250	,2220
b, d	-,072	28	-,416	,6803
b, c	-,533	32	-2,997	,0052
d, c	-,460	26	-3,026	,0055

	Count	Mean	Variance	Std. Dev.	Std. Err
a	13	11,122	,236	,486	,135
b	18	10,812	,308	,555	,131
d	12	10,884	,073	,271	,078
c	16	11,345	,221	,471	,118

Table 6

	Mean Diff.	DF	t-Value	P-Value
a, b	-,274	43	-2,204	,0329
a, c	-,372	27	-3,166	,0038
a, d	,193	18	1,365	,1890
b, c	-,099	46	-,892	,3771
b, d	,466	37	3,056	,0042
c, d	,565	21	4,662	,0001

	Count	Mean	Variance	Std. Dev.	Std. Err
a	13	11,261	,117	,342	,095
b	32	11,535	,152	,390	,069
c	16	11,633	,085	,291	,073
d	7	11,069	,038	,195	,074

Table 7

Accepted manuscript