ACTA SCIENTIFIC MICROBIOLOGY (ISSN: 2581-3226)

Volume 2 Issue 5 May 2019

Research Article

Proteome Diversities Among 19 Archaebacterial Species

Jung Hwan Kim[1,2] and Maurice HT Ling[1,2,3]*

[1]*Department of Applied Sciences, Northumbria University, United Kingdom*

[2]*School of Life Sciences, Management Development Institute of Singapore, Singapore*

[3]*HOHY PTE LTD, Singapore*

***Corresponding Author:** Maurice HT Ling, Department of Applied Sciences, Northumbria University, United Kingdom and School of Life Sciences, Management Development Institute of Singapore, Singapore and HOHY PTE LTD, Singapore.

Received: April 2, 2019; **Published:** April 09, 2019

Abstract

Archaebacteria is known for its presence in varied extreme environments, suggesting potential applications and an on-going need study its diversity. This led to increasing emphasis on archaeal genomic and proteomic studies. However, there is no work to-date examining the overall proteomic diversity in archaebacteria. In this study, we examine the proteomic diversities among 19 sequenced archaebacterial species and found significant differences (p-value < 2×10^{-43}) in average peptide lengths, isoelectric points, aromaticity, instability, and hydropathy. Majority of the peptides in each species are stable. Predominantly consistent correlations, though widely varied, were observed between peptide physical properties except between peptide length and hydropathy. This study provides a cursory view highlighting the diversity of archaeal proteomes; thus, re-iterating the call for further studies into these organisms.

Keywords: Peptide; Hydropathy; Archaebacteria

Introduction

Archaebacteria is one of the three kingdoms of life [1], which is known for its presence in a wide range of hostile environments [2,3] and challenges our understanding on the limits of life [4-6]. A recent review had illustrated the applications of archaebacteria in bioremediation [7]. For example, applications of archaebacteria have been suggested in hypersaline [8], hydrocarbon-rich [9,10], acidic [11], and radioactive [12,13] environments. However, there is on-going needs to study the diversity in archaebacteria, which may lead to more applications in diverse environments [7]. Moreover, the diversity of archaebacteria calls for studies into the basic biology of these organisms [2].

There has been substantial studies on the genomic diversity of archaebacteria [2,14-17] from whole genome level down to genetic features. Reza., *et al*. [18] used metagenomic analysis and found 3 archaea phyla in the surface water collected at Ofunato Bay in Japan, and several studies have shown large diversity in DNA replication [19,20]. Comparatively, there is less work examining the proteomic diversity of archaebacteria. Chowdhury and Dutta [21] found 22 Cluster of Orthologous Groups that are exclusive to methanogens, suggesting environmental specificity in archaeal proteomes. Sun., *et al*. [22] examined protein domains across all three kingdoms and found a cluster of 92 protein domains common in archaea and eukarya, supporting the possibility that archaea and eukarya lineages may originate from a common ancestor – last eukaryotic common ancestor [23,24]. However, there is no work to-date examining the overall proteomic diversity in archaebacteria.

In this study, we examine the proteomic diversities among 19 sequenced archaebacterial species across 7 taxonomical groups. Our results suggest significant differences (p-value < 2×10^{-43}) in average peptide lengths, isoelectric points, aromaticity, instability, and hydropathy. Majority of the peptides in each species are stable. Correlation analyses show large variations in proteomic properties across the 19 species. Hence, this study highlights the diversity of proteomes in archaebacteria, which re-iterates the call [2] for further basic work into these organisms.

Materials and Methods

Proteomic Sequences

Nineteen fully sequenced and assembled archaebacterial species were selected from seven different groups from NCBI Genome Database; namely, (1) A*cidilobus saccharovorans* 345-15 (Accession no. NC_014374.1), (2) *Aciduliprofundum boonei* T469 (Accession no. NC_013926.1), (3) *Archaeoglobus fulgidus* DSM 4304 (Accession no. NC_000917.1), (4) *Desulfurococcus mucosus* DSM 2162 (Accession no. NC_014961.1), (5) *Ferroglobus placidus* DSM 10642 (Accession no. NC_013849.1), (6) *Geoglobus ahangari* 234 (Accession no. NZ_CP011267.1), (7) *Halomicrobium mukohataei* DSM 12286 (Accession no. NC_013202.1), (8) *Metallosphaera cuprina* Ar-4 (Accession no. NC_015435.1), (9) *Methanobrevibacter millerae* SM9 (Accession no. NZ_CP011266.1), (10) *Methanocaldococcus vulcanius* M7 (Accession no. NC_013407.1), (11) *Methanococcus aeolicus* Nankai-3 (Accession no. NC_009635.1), (12) *Methanomethylovorans hollandica* DSM 15978 (Accession no. NC_019977.1), (13) *Natronomonas moolapensis* 8.8.11 (Accession no. NC_020388.1), (14) *Nitrososphaera viennensis* EN76 (Accession no. NZ_CP007536.1), (15) *Palaeococcus pacificus* DY20341 (Accession no. NZ_CP006019.1), (16) *Picrophilus torridus* DSM 9790 (Accession no. NC_005877.1), (17) *Pyrococcus abyssi* (Accession no. NC_000868.1), (18) *Thermococcus barossii* SHCK-94 (Accession no. NZ_CP015101.1), and (19) *Thermoplasma volcanium* GSS1 (Accession no. NC_002689.2). A proteome is defined as the set of downloadable coding sequences features as protein from each record using accession number.

Analyzing physical properties of peptides

Physical properties of peptides (aromaticity, hydropathy, isoelectric point, and instability) were calculated using methods in Biopython library [25]. Aromaticity refers to the relative abundance of aromatic amino acids in a peptide [26]. Hydropathy refers to the overall hydrophobic/hydrophilic properties of a peptide [27]. Isoelectric point (pI) is the pH where a peptide is electrical neutrality [28]. Instability index refers to the stability of the peptide where high instability score suggests shorter half-life [29]. Differences in between sample means and correlation between two sets of data were performed using ANOVA and Pearson's correlation respectively.

Results and Discussion

The physical properties of each proteome were analysed individually and found to be substantially varied. Correlation analyses of properties showed variations between each archaebacterial species.

Significant proteome diversities within archaebacterial taxons

Genomic size analysis (Table 1) shows that the genome size of the 19 examined species ranges from 3,110,487 bp (*Halomicrobium mukohataei* DSM 12286) to 1,314,639 bp (*Desulfurococcus mucosus* DSM 2162) with an average genome size of 1,989,466 bp. This is consistent with Koonin and Wolf [30] suggesting that the size of archaebacterial genome centres around 2 Mb. The number of coding genes ranges from 3,232 to 1,345 with an average of 2,036. There is high correlation (r = 0.96) between genome size and the number of coding genes. Hence, the average genome length for each coding gene is 981 bp with a standard deviation of 65.25 bp. These results are consistent (p-value = 0.22) with the rule of thumb of 1 gene per 1000 bases in bacterial genomes [31].

Taxon	Species	Accession Number	Genome Length (base-pairs)	Number of Coding Genes
Archaeoglobi	*Archaeoglobus fulgidus* DSM 4304	NC_000917.1	2,178,400	2,369
	Ferroglobus placidus DSM 10642	NC_013849.1	2,196,266	2,479
	Geoglobus ahangari 234	NZ_CP011267.1	1,770,093	1,985
Crenarchaeota	*Acidilobus saccharovorans* 345-15	NC_014374.1	1,496,453	1,487
	Desulfurococcus mucosus DSM 2162	NC_014961.1	1,314,639	1,345
	Metallosphaera cuprina Ar-4	NC_015435.1	1,840,348	1,894
Diaforarchaea	*Aciduliprofundum boonei* T469	NC_013926.1	1,486,778	1,521
	Picrophilus torridus DSM 9790	NC_005877.1	1,545,895	1,563
	Thermoplasma volcanium GSS1	NC_002689.2	1,584,804	1,545
Stenosarchaea	*Halomicrobium mukohataei* DSM 12286	NC_013202.1	3,110,487	3,232
	Methanomethylovorans hollandica DSM 15978	NC_019977.1	2,428,904	2,525
	Natronomonas moolapensis 8.8.11	NC_020388.1	2,912,573	2,733
Methanomada	*Methanobrevibacter millerae* SM9	NZ_CP011266.1	2,543,538	2,209
	Methanocaldococcus vulcanius M7	NC_013407.1	1,746,329	1,695
	Methanococcus aeolicus Nankai-3	NC_009635.1	1,569,500	1,489
Thaumarchaeota	*Nitrososphaera viennensis* EN76	NZ_CP007536.1	2,527,938	2,801
Thermococci	*Palaeococcus pacificus* DY20341	NZ_CP006019.1	1,859,370	1,950
	Pyrococcus abyssi	NC_000868.1	1,765,118	1,862
	Thermococcus barossii SHCK-94	NZ_CP015101.1	1,922,421	1,996

Table 1: Genome Length and Number of Coding Genes.

Average peptide lengths range from 250 to 328 amino acids (Figure 1A), with a grand mean and standard error of 290 amino acids and 16.5 amino acids respectively. This is consistent (p-value = 0.24) with Zhang [32] whom examined the average protein length across all three kingdoms and found that archaebacteria have shorter average protein length (270 ± 9 amino acids) compared to eubacteria (averaging 330 ± 5 amino acids). However, there is significant diversity in peptide length (Table 2) across all 19 species (p-value = 2×10^{-43}) and within 3 of the 6 taxonomical classifications.

Taxon	Accession Numbers	ANOVA p-values				
		Peptide Length	pI	Aromaticity	Instability	Hydropathy
19 species	All 19 Accession Numbers	2×10^{-43}	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$	3×10^{-94}	8×10^{-138}
Crenarchaeota (n = 3)	NC_014374.1 NC_014961.1 NC_015435.1	0.001	4×10^{-6}	$<1 \times 10^{-300}$	1×10^{-16}	3×10^{-8}
Thermococci (n = 3)	NZ_CP006019.1 NC_000868.1 NZ_CP015101.1	0.721	6×10^{-15}	0.009	0.001	0.184
Stenosarchaea (n = 3)	NC_013202.1 NC_019977.1 NC_020388.1	0.046	4×10^{-244}	$<1 \times 10^{-300}$	0.004	$<1 \times 10^{-300}$
Methanomada (n = 3)	NZ_CP011266.1 NC_013407.1 NC_009635.1	2×10^{-5}	4×10^{-133}	2×10^{-16}	0.020	0.018
Diaforarchaea (n = 3)	NC_013926.1 NC_005877.1 NC_002689.2	0.097	2×10^{-5}	1×10^{-12}	0.118	0.015
Archaeoglobi (n = 3)	NC_000917.1 NC_013849.1 NZ_CP011267.1	0.147	$< 1 \times 10^{-300}$	1×10^{-9}	8×10^{-6}	4×10^{-4}

Table 2: Differences Within Taxon and Between Each of the 19 Archaebacterial Species.

Average isoelectric points range from 5.0 to 7.5 (Figure 1B), with a grand mean and standard error of 6.8 and 0.75 respectively. There are significant pI variations (Table 2) across all 19 species (p-value < 1×10^{-300}) and within all 6 taxonomical classifications (p-value < 2×10^{-5}). Average aromaticity indices range from 0.068 to 0.107 (Figure 1C), with a grand mean and standard error of 0.088 and 0.0099 respectively. This suggests that archaeal proteomes are predominantly aromatic (p-value = 5×10^{-8}). There are significant aromaticity variations (Table 2) across all 19 species (p-value < 1×10^{-300}) and within all 6 taxonomical classifications (p-value < 0.009). Average instability indices range from 33.3 to 38.5 (Figure 1D), with a grand mean and standard error of 33.6 and 1.29 respectively. There are significant stability variations (Table 2) across all 19 species (p-value = 3×10^{-94}) and within 5 of the 6 taxonomical classifications. An instability index above 40 suggests unstable peptide and is indicative of short half-life based on the documentation regarding instability index in Biopython library [25]. This suggests that a majority of peptides in each archaeal species are stable (p-value = 0.0001). This is consistent with studies suggesting that archaeal proteins are adapted for stability under extreme environments where they live [33,34]. Average hydropathy indices range from -0.21 to 0.02 (Figure 1E), with a grand mean and standard error of -0.10 and 0.064 respectively. There are significant hydropathy variations (Table 2) across all 19 species (p-value = 8×10^{-138}) and within 5 of the 6 taxonomical classifications. Taken together, the four physical properties of peptides (pI, aromaticity,

instability, and hydropathy) are determined by their amino acid compositions. Moura., *et al*. [35] found that relative amino acid abundance in an organism correlates to the relative amino acid abundance of its environment and archaebacteria is known for its diverse environments [2,3]. This may suggest that the differences in physical properties of peptides may be a result of environmental amino acid abundance.

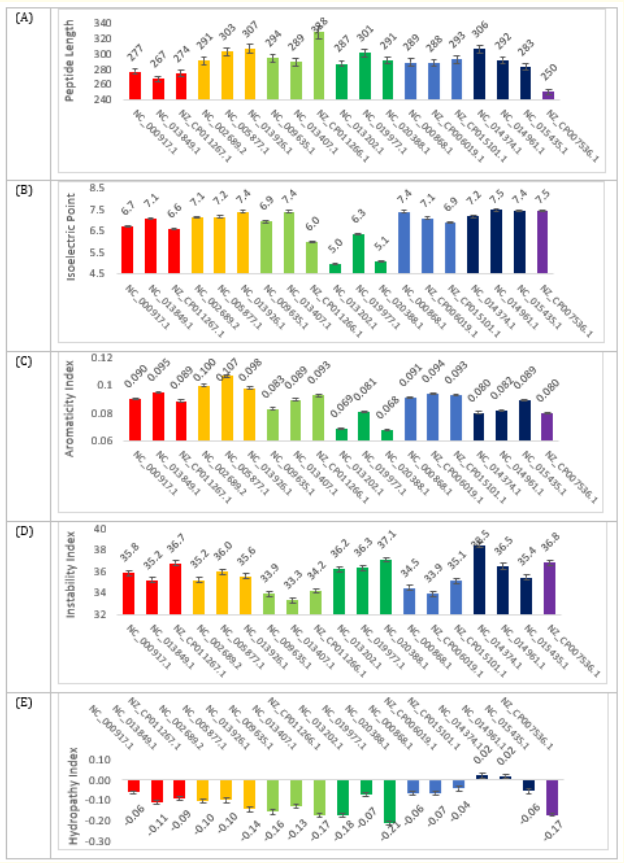

Figure 1: Distributions of Physical Properties of Peptides. 19 archaebacterial species are colour-coded by taxon as (1) Archaeoglobi (red), (2) Diaforarchaea (yellow), (3) Methanomada (yellow green), (4) Stenosarchaea (green), (5) Thermococci (blue), (6) Crenarchaeota (dark blue), and (7) Thaumarchaeota (purple). Panels A to E are peptide length, isoelectric point, aromaticity index, instability index, and hydropathy index respectively. Standard errors are shown as error bars.

Correlations between peptide length and peptide physical properties

Taking all 38,680 peptides across all 19 species (total peptide set), the correlation between peptide length and pI is -0.117 (Figure 2A). Average correlation coefficient between peptide length and pI range from -0.069 to 0.180, with a grand mean and standard error of -0.123 and 0.0251 respectively; which is not significantly different (p-value = 0.80) from the correlation using total peptide set but significantly different (p-value < 0.007) from no correlation. At the individual species level, correlation is significantly different from zero (p-value < 0.007) in all 19 species.

Correlation of total peptide set between peptide length and instability indices is -0.077 (Figure 2B). Average correlation coefficient between peptide length and instability indices range from -0.138 to 0.011, with a grand mean and standard error of -0.071 and 0.0370 respectively; which is not significantly different (p-value = 0.86) from the correlation using total peptide set. This is consistent with a study suggesting possible relationship between peptide length and protein half-life [36]. At the individual species level, correlation is significantly different from zero (p-value < 0.03) in 15 of the 19 species with *Geoglobus ahangari* 234, *Picrophilus torridus* DSM 9790, *Methanococcus aeolicus* Nankai-3, and *Thermoplasma volcanium* GSS1 showing no significant correlations (p-value > 0.1).

Correlation of total peptide set between peptide length and hydropathy indices is -0.002 (Figure 2C) with is not significantly different from no correlation (p-value = 0.97). Average correlation coefficient between peptide length and instability indices range from -0.111 to 0.083, with a grand mean and standard error of 0.0003 and 0.0616 respectively; which is not significantly different (p-value = 0.97) from the correlation using total peptide set. At the individual species level, correlation is significantly different from zero (p-value < 0.03) in 11 of the 19 species with *Archaeoglobus fulgidus* DSM 4304, *Ferroglobus placidus* DSM 10642, *Geoglobus ahangari* 234, *Desulfurococcus mucosus* DSM 2162, *Aciduliprofundum boonei* T469, *Natronomonas moolapensis* 8.8.11, *Methanobrevibacter millerae* SM9, and *Palaeococcus pacificus* DY20341 showing no significant correlations (p-value > 0.3).

Correlation of total peptide set between peptide length and aromaticity indices is 0.089 (Figure 2D). Average correlation coefficient between peptide length and aromaticity indices range from 0.026 to 0.270, with a grand mean and standard error of 0.119 and 0.0698 respectively; which is not significantly different (p-value = 0.68) from the correlation using total peptide set. At the individual species level, correlation is significantly different from zero (p-value < 0.04) in 18 of the 19 species with *Nitrososphaera viennensis* EN76 showing no significant correlations (p-value = 0.17).

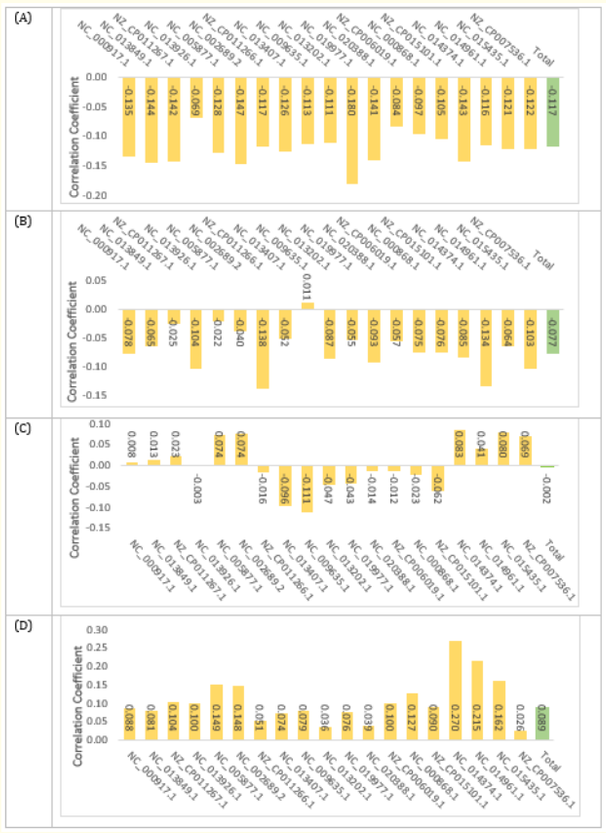

Figure 2: Correlations between Peptide Length and Peptide Physical Properties. "Total" represents the union of all 19 archaebacterial proteomes. Panel A shows correlation between peptide length and isoelectric point. Panel B shows correlation between peptide length and instability. Panel C shows correlation between peptide length and hydropathy. Panel D shows correlation between peptide length and aromaticity.

Correlations between isoelectric point and instability, hydropathy, or aromaticity

Correlation of total peptide set between isoelectric point and instability indices is -0.011 (Figure 3A). Average correlation coefficient between peptide length and instability indices range from -0.073 to 0.097, with a grand mean and standard error of -0.002 and 0.0451 respectively; which is not significantly different (p-value = 0.85) from the correlation using total peptide set. At the individual species level, correlation is not significantly different from zero (p-value > 0.08) in 13 of the 19 species with *Desulfurococcus mucosus* DSM 2162, *Halomicrobium mukohataei* DSM 12286, *Natronomonas moolapensis* 8.8.1, *Methanocaldococcus vulcanius* M7, *Methanococcus aeolicus* Nankai-3, and *Pyrococcus abyssi* showing significant correlations (p-value < 0.03).

Correlation of total peptide set between isoelectric point and hydropathy indices is 0.165 (Figure 3B). Average correlation coefficient between peptide length and hydropathy indices range from 0.024 to 0.375, with a grand mean and standard error of 0.137 and 0.0973 respectively; which is not significantly different (p-value = 0.75) from the correlation using total peptide set. At the individual species level, correlation is significantly different from zero (p-value < 0.03) in 16 of the 19 species with *Desulfurococcus mucosus* DSM 2162, *Thermoplasma volcanium* GSS1, and *Methanococcus aeolicus* Nankai-3 showing no significant correlations (p-value > 0.1).

Correlation of total peptide set between isoelectric point and aromaticity indices is 0.212 (Figure 3C). Average correlation coefficient between peptide length and aromaticity indices range from 0.027 to 0.224, with a grand mean and standard error of 0.146 and 0.0605 respectively; which is not significantly different (p-value = 0.29) from the correlation using total peptide set. At the individual species level, correlation is significantly different from zero (p-value < 0.01) in 17 of the 19 species with *Desulfurococcus mucosus* DSM 2162, and *Nitrososphaera viennensis* EN76 showing no significant correlations (p-value > 0.06).

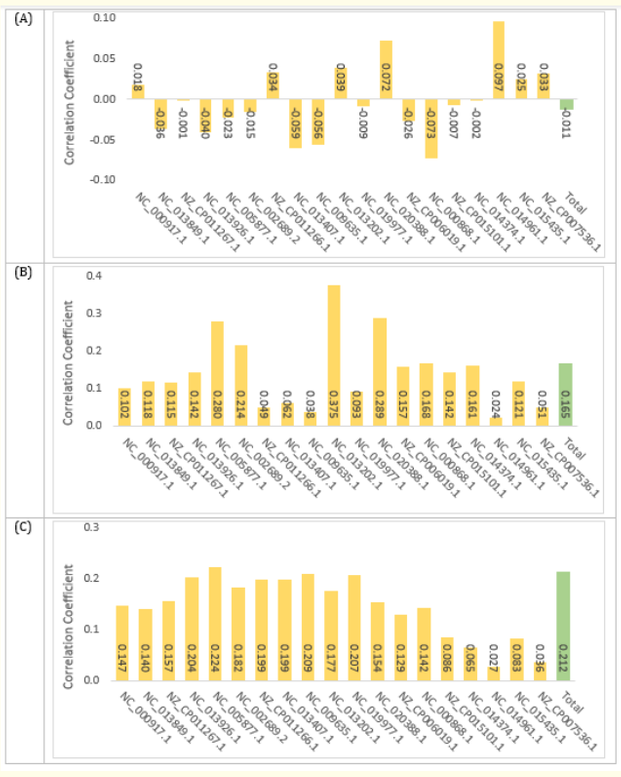

Figure 3: Correlations between Isoelectric Point and Instability, Hydropathy, or Aromaticity. "Total" represents the union of all 19 archaebacterial proteomes. Panel A shows correlation between isoelectric point and instability. Panel B shows correlation between isoelectric point and hydropathy. Panel A shows correlation between isoelectric point and aromaticity.

Correlations between instability and hydropathy or aromaticity

Correlation of total peptide set between instability indices and hydropathy indices is -0.338 (Figure 4A). Average correlation co-

efficient between instability indices and hydropathy indices range from -0.444 to -0.295, with a grand mean and standard error of -0.340 and 0.0444 respectively; which is not significantly different (p-value = 0.96) from the correlation using total peptide set. At the individual species level, correlation is significantly different from zero (p-value < 1 x 10^{-31}) for all 19 species.

Correlation of total peptide set between instability indices and aromaticity indices is -0.109 (Figure 4B). Average correlation coefficient between instability indices and aromaticity indices range from -0.196 to 0.004, with a grand mean and standard error of -0.104 and 0.0443 respectively; which is not significantly different (p-value = 0.91) from the correlation using total peptide set. At the individual species level, correlation is significantly different from zero (p-value < 0.03) in 18 of the 19 species with *Nitrososphaera viennensis* EN76 showing no significant correlations (p-value = 0.08).

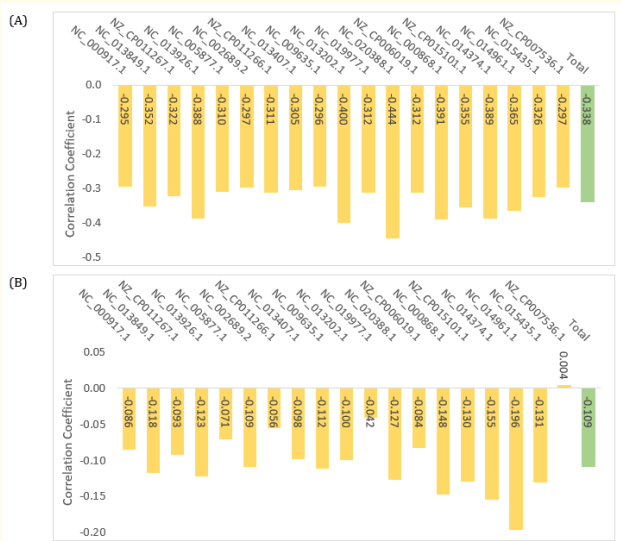

Figure 4: Correlations between Instability and Hydropathy or Aromaticity. "Total" represents the union of all 19 archaebacterial proteomes. Panel A shows correlations between instability and hydropathy. Panel B shows correlations between instability and aromaticity.

Correlations between hydropathy and aromaticity

Correlation of total peptide set between hydropathy indices and aromaticity indices is 0.283 (Figure 5). Average correlation coefficient between hydropathy indices and aromaticity indices range from 0.118 to 0.437, with a grand mean and standard error of 0.289 and 0.0777 respectively; which is not significantly different (p-value = 0.94) from the correlation using total peptide set. At the individual species level, correlation is significantly different from zero (p-value < 1 x 10^{-10}) in all 19 species. This is consistent with a study on rabies virus glycoproteins finding slight correlation between hydropathy and aromaticity [37].

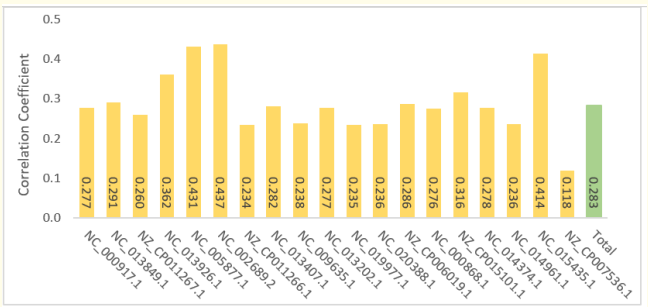

Figure 5: Correlations between Hydropathy and Aromaticity. "Total" represents the union of all 19 archaebacterial proteomes.

Conclusion and Future Work

There has been substantially fewer studies on archaebacteria and their diversity compared to eubacteria despite its potential in many industrial applications [7] and potential insights into the limits of life [4-6]. In this study, we examine the proteome of 19 archaeal species to provide a cursory view on diversity; in terms of peptide lengths, aromaticity, pI, instability. This re-iterating the call for further studies into these organisms and future work can expand on both the number of species and the properties examined.

Conflict of Interest

The authors declare no conflict of interest.

Bibliography

1. Woese CR and Fox GE. "Phylogenetic structure of the prokaryotic domain: the primary kingdoms". *Proceedings of the National Academy of Sciences of the United States of America* 74.11 (1977): 5088-5090.

2. Dombrowski N., *et al*. "Genomic diversity, lifestyles and evolutionary origins of DPANN archaea". *FEMS Microbiology Letters* 366.2 (2019): fnz008.

3. DeLong EF and Pace NR. "Environmental diversity of bacteria and archaea". *System Biology* 50.4 (2001): 470-478.

4. Pikuta EV., *et al*. "Microbial extremophiles at the limits of life". *Critical Reviews in Microbiology* 33 (2007): 183-209.

5. Schulze-Makuch D., *et al*. "The Physical, Chemical and Physiological Limits of Life". *Life* (Basel, Switzerland). 5 (2015): 1472-1486.

6. de Vera J-P., *et al*. "Limits of Life and the Habitability of Mars: The ESA Space Experiment BIOMEX on the ISS". *Astrobiology* 19 (2019): 145-157.

7. Krzmarzick MJ., *et al*. "Diversity and Niche of Archaea in Bioremediation". *Archaea* (2018): 3194108.

8. Al-Mailem DM., *et al*. "Biostimulation of indigenous microorganisms for bioremediation of oily hypersaline microcosms from the Arabian Gulf Kuwaiti coasts". *Journal of Environmental Economics and Management* 193 (2017): 576-583.

9. Sarkar J., *et al*. "Biostimulation of Indigenous Microbial Community for Bioremediation of Petroleum Refinery Sludge". *Front Microbiology* 7 (2016): 1407.

10. Zhao Y., *et al*. "Bioremediation of contaminated urban river sediment with methanol stimulation: Metabolic processes accompanied with microbial community changes". *Science of the Total Environment* 653 (2019): 649-657.

11. Klein R., *et al*. "Bioremediation of mine water". *Advances in Biochemical Engineering* 141 (2014): 109-172.

12. Holmes DE., *et al*. "Potential for Methanosarcina to Contribute to Uranium Reduction during Acetate-Promoted Groundwater Bioremediation". *Microbial Ecology* 76 (2018): 660-667.

13. Marques CR. "Extremophilic Microfactories: Applications in Metal and Radionuclide Bioremediation". *Front Microbiology* 9 (2018): 1191.

14. Tully BJ. "Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns". *Nature Communication* 10 (2019): 271.

15. Spang A., *et al*. "Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life". *Science* 357 (2017): eaaf3883.

16. Adam PS., *et al*. "The growing tree of Archaea: new perspectives on their diversity, evolution and ecology". *ISME Journal* 11 (2017): 2407-2425.

17. Cordero OX and Polz MF. "Explaining microbial genomic diversity in light of evolutionary ecology". *Nature Reviews Microbiology* 12 (2014): 263-273.

18. Reza MS., *et al*. "Taxonomic profiles in metagenomic analyses of free-living microbial communities in the Ofunato Bay". *Gene* 665 (2018): 192-200.

19. Ausiannikava D and Allers T. "Diversity of DNA Replication in the Archaea". *Genes* 8 (2017): 56.

20. Wu Z., *et al*. "DNA replication origins in archaea". *Frontiers in Microbiology* 5 (2014): 179-179.

21. Roy Chowdhury A., *et al*. "A pursuit of lineage-specific and niche-specific proteome features in the world of archaea". *BMC Genomics* 13 (2012): 236.

22. Sun C-T., *et al*. "A proteome view of structural, functional, and taxonomic characteristics of major protein domain clusters". *Scientific Reports* 7 (2017): 14210-14210.

23. Koonin EV. "Archaeal ancestors of eukaryotes: not so elusive any more". *BMC Biology* 13 (2015): 84.

24. O'Malley MA., *et al*. "Concepts of the last eukaryotic common ancestor". *Nature Ecology and Evolution* 3 (2019): 338-344.

25. Cock PJA., *et al*. "Biopython: freely available Python tools for computational molecular biology and bioinformatics". *Bioinformatics* 25 (2009): 1422-1423.

26. Lobry JR and Gautier C. "Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 Escherichia coli chromosome-encoded genes". *Nucleic Acids Research* 22 (1994): 3174-3180.

27. Kyte J and Doolittle RF. "A simple method for displaying the hydropathic character of a protein". *Journal of Molecular Biology* 157 (1982): 105-132.

28. Bjellqvist B., *et al*. "The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences". *Electrophoresis* 14 (1993): 1023-1031.

29. Guruprasad K., *et al*. "Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence". *Protein Engineering, Design and Selection* 4 (1990): 155-161.

30. Koonin EV and Wolf YI. "Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world". *Nucleic Acids Research* 36 (2008): 6688-6719.

31. Lynch M and Walsh B. "The origins of genome architecture". Sinauer Associates Sunderland, MA 2007.

32. Zhang J. Protein-length distributions for the three domains of life". *Trends in Genetics* 16 (2000): 107-109.

33. Macario AJ., *et al*. "Stress genes and proteins in the archaea". *Microbiology and Molecular Biology Reviews : MMBR* 63 (1999): 923-967.

34. Reed CJ., *et al*. "Protein Adaptations in Archaeal Extremophiles". *Archaea* 2013 (2013): 373275.

35. Moura A., *et al*. "Relative Amino Acid Composition Signatures of Organisms and Environments". *PLOS ONE* 8 (2013): e77319.

36. Rahman M., *et al*. "Predicting the protein half-life in tissue from its cellular properties". *PLOS ONE* 12 (2017): e0180428.

37. Morla S., *et al*. "Synonymous codon usage pattern in glycoprotein gene of rabies virus". *Gene* 584 (2016): 1-6.

Volume 2 Issue 5 May 2019