

Research article

The methanogen core and pangenome: conservation and variability across biology's growth temperature extremes

Paula Prondzinsky^{1,2,*}, , Sakae Toyoda², and Shawn Erin McGlynn^{1,3,4,*}, 

¹Earth-Life Science Institute, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, 152-8550 Tokyo, Japan

²Department of Chemical Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, 226-8503 Yokohama, Japan

³Center for Sustainable Resource Science, RIKEN, 2-1 Hirosawa, Wako, 351-0198 Saitama, Japan

⁴Blue Marble Space Institute of Science, Seattle, WA 98154, USA

*To whom correspondence should be addressed. Tel: +81 3 5734 3154. Fax: +81 3 5734 3416. Email: prondzinsky@elsi.jp (P.P.); mcglynn@elsi.jp (S.E.M.)

Abstract

Temperature is a key variable in biological processes. However, a complete understanding of biological temperature adaptation is lacking, in part because of the unique constraints among different evolutionary lineages and physiological groups. Here we compared the genomes of cultivated psychrotolerant and thermotolerant methanogens, which are physiologically related and span growth temperatures from -2.5°C to 122°C . Despite being phylogenetically distributed amongst three phyla in the archaea, the genomic core of cultivated methanogens comprises about one-third of a given genome, while the genome fraction shared by any two organisms decreases with increasing phylogenetic distance between them. Increased methanogenic growth temperature is associated with reduced genome size, and thermotolerant organisms—which are distributed across the archaeal tree—have larger core genome fractions, suggesting that genome size is governed by temperature rather than phylogeny. Thermotolerant methanogens are enriched in metal and other transporters, and psychrotolerant methanogens are enriched in proteins related to structure and motility. Observed amino acid compositional differences between temperature groups include proteome charge, polarity and unfolding entropy. Our results suggest that in the methanogens, shared physiology maintains a large, conserved genomic core even across large phylogenetic distances and biology's temperature extremes.

Key words: methanogenesis, genome comparison, temperature adaptations, pangenome, physiology

Introduction

Life is a chemical process: it follows that temperature is a key determinant both of rate and reaction favourability.^{1,2} Temperature effects can be seen across biology,³ and it is of interest to consider these from an evolutionary perspective. Although recent phylogenetic trees do not necessarily indicate a basal placement of thermophiles, the history of life in relation to temperature has been debated ever since the first trees placed thermophilic lineages in basal positions^{4–7} suggesting that taxonomic evolution may be entangled with temperature.

An ongoing challenge has been to identify how organisms adapt to life at high temperatures.^{16–21} Able to grow from -2.5 to 122°C ,^{8,9} methanogens span perhaps the largest temperature range of any single physiological group. Methanogens generally produce methane from CO_2 and H_2 , acetate and/or methylated compounds,^{10,11} and are found in environments including animal rumens,¹² to wetlands,¹³ to hydrothermal vents,¹⁴ to subglacial sediments.¹⁵

To avoid confounding factors when considering different lifestyles and phylogenetic backgrounds, Lecocq et al.²² investigated the Methanococcales, finding that amino acid substitution patterns, rather than the acquisition of orthogroups best explain thermotolerance. Other previous comparative

genomic studies of methanogens focused on pairs of species or organisms within a specific clade.^{23–25}

We sought to investigate possible genome adaptations to temperature by comparatively analysing genomes of cultured organisms for which growth temperature and substrate utilization information is available. By analysing methanogens as a physiologically similar yet phylogenetically diverse group, we specifically aimed to constrain metabolic variables and to search for temperature effects on genome evolution which might exist as general adaptation strategies rather than as synapomorphies unique to specific phylogenetic groups. Our study complements and adds to our understandings of what makes a methanogen a methanogen, and how they have genetically evolved over time.

Methodology

Establishing a methanogen database

As a first step, the diversity of alkane-activating archaea was investigated based on the presence of alpha, beta and gamma subunits of the methyl-coenzyme M reductase (Mcr) enzyme using Annotree v1.2.0.²⁶ The KEGG UniRef100 database²⁷ was used to query the alpha (K00399), beta (K00401) and gamma (K00402) subunits in archaea with

default parameters (min. 30% identity, max. E value 0.00001, min. 70% subject and query alignment). This search resulted in a total of 290 genome hits across six archaeal phyla including Halobacteriota (156), Methanobacteriota (94), Thermoplasmatota (27), Thermoproteota (10), Asgardarchaeota (2) and Hadarchaeota (1). Their protein sequences were downloaded from the GTDB data repository (GTDB release 95).²⁸ A total of 255 genomes were considered for further analyses after a quality cut-off (checkM score ≥ 90). In previous taxonomic descriptions, methanogens were separated into Class one (Methanobacteriota) and Class two (Halobacteriota) methanogens within the Euryarchaeota phylum,²⁹ however here we describe their taxonomy using that of GTDB release 95.

Temperature data for cultured organisms were accessed through the Database of Growth TEMPeratures of Usual and Rare Prokaryotes (TEMPURA, Sato et al.³⁰) and substrate data was added from the PhyMet2 database.³¹ This data compilation is available in [Supplementary Table S1](#). Phylogenetic trees with annotations were visualized using the interactive tree of life online tool.³²

A total of 86 organisms had temperature data available and were used for the main analyses herein. There is no universal definition for growth temperatures to classify organisms into psychrotolerant and thermotolerant groups. Here we defined these temperature groups based on the overall distribution of growth temperatures for methanogenesis ([Supplementary Fig. S1](#)). Organisms were classified into psychrotolerant organisms with a minimum growth temperature $\leq 15^\circ\text{C}$ and with no growth observed above 45°C . Thermotolerant organisms have a maximum growth temperature $\geq 45^\circ\text{C}$ and no growth observed below 15°C . Organisms spanning $\leq 15^\circ\text{C}$ to $\geq 45^\circ\text{C}$ are considered as mesophilic. This classification resulted in 21 psychrotolerant, 41 thermotolerant and 24 mesophilic species.

Core and pangenomes

Core- and pangenomes were established with OrthoFinder³³ running with default parameters on both the 255 species with checkM score ≥ 90 and the 86 species with available temperature data. The resulting orthogroups were separated into the core (here we used the extended core of orthogroups present in more than 95% of organisms to account for incomplete genomes), shared (orthogroups present in at least two but less than 95% of organisms) and unique (genes present only in one organism) groups. These groupings were used for all subsequent analyses.

For comparison to other microbes, genera with temperature data for at least five of their member species were selected from the TEMPURA database and five species selected randomly for core genome analyses using Proteinortho,³⁴ run with the default parameters. The number of species was kept constant to account for the reduction of the core size with the addition of species³⁵ and to keep the analyses comparable. Fourteen archaeal and bacterial genera, including two methanogenic ones, are represented by at least five species and were used for this analysis, with eight thermotolerant (*Pyrococcus*, *Pyrobaculum*, *Thermococcus*, *Methanocaldococcus*, *Acidianus*, *Thermotoga*, *Caldicellulosiruptor*, and *Fervidobacterium*), four mesophilic (*Acholeplasma*, *Bacteroides*, *Eubacterium* and *Methanosarcina*) and two psychrotolerant (*Cryobacterium* and *Paraglaciecola*) genera.

Composition and function

Functional annotations of all genes in each orthogroup were made with eggNOGmapper version 5.³⁶ Initial overviews of functional differences between orthogroup and temperature groups were done based on cluster of orthologous genes (COG) categories. Subsequently KEGG orthology output from eggNOGmapper was used to map reactions unique to temperature groups as well as substrate groups with the KEGG pathway module.²⁷

Amino acid biases were assessed by counting amino acids coded by open reading frames in the whole genomes as well as core and shared fractions. Based on amino acid compositions, proteome properties were calculated based on Gromiha et al.³⁷ indicating the average property of a residue within a species' proteome. To test for statistically significant trends with temperature in amino acid composition and amino acid residue properties, Pearson correlation P-values were calculated for amino acid counts as well as residue properties for each species using the statistical functions module in python's scipy package³⁸ and tested for statistical significance with the Benjamini Hochberg correction using the fdr_correction function with method = 'indep' in the mne module in python.³⁹

Ancestral reconstruction for orthogroups

Count⁴⁰ was used for estimations of gene gain and loss events using Wagner parsimony as well as to establish rate model optimization using the gain–loss–duplication model⁴¹ with the Poisson family size distribution at the root. Lineage-specific variations were left unspecified, and all other parameters were kept at default. The input data for this analysis was the group of 86 methanogens associated with growth temperature data.

Results and discussion

Phylogenetic distribution of growth temperature and physiology among methanogens

Using Annotree, we found 290 archaeal genomes contain the Mcr alpha, beta and gamma subunits, out of which 255 are considered to have a high enough quality for analyses (checkM score ≥ 90). Those genomes range in size from 1.112 Mbp (WYZ-LMO11 sp004348015 of the *Methanomethylaceae* family) to 5.752 Mbp (*Methanosaerina acetivorans*) and span five phyla, 17 classes, 21 orders, 38 families and 99 genera, which highlights the continued taxonomic expansion of Mcr containing archaea.^{11,42,43} A list of these genomes with information on their size, quality and where available growth temperature ranges, isolation environments and growth substrates of species is available in [Supplementary Table S1](#).

Among the 255 high-quality archaeal genomes, substrate data from cultivation studies is available for 112³¹ ([Fig. 1](#)). The *Methanomicrobia*, *Methanococci*, *Methanocellia* and *Methanobacteria* genera are strictly hydrogenotrophic, *Methanotrichales* strictly aceticlastic and all genera in the *Methanosaerinae* family besides *Methanosaerina* are strictly methylotrophic, while some species within *Methanosaerina* are capable of methanogenesis from CO₂ and H₂, acetate and/or methylated compounds. Eighty-three species have capabilities for hydrogenotrophic methanogenesis, while 72 are strict hydrogenotrophs unable to use other methanogenesis pathways. Psychrotolerant hydrogenotrophic organisms include members of the class

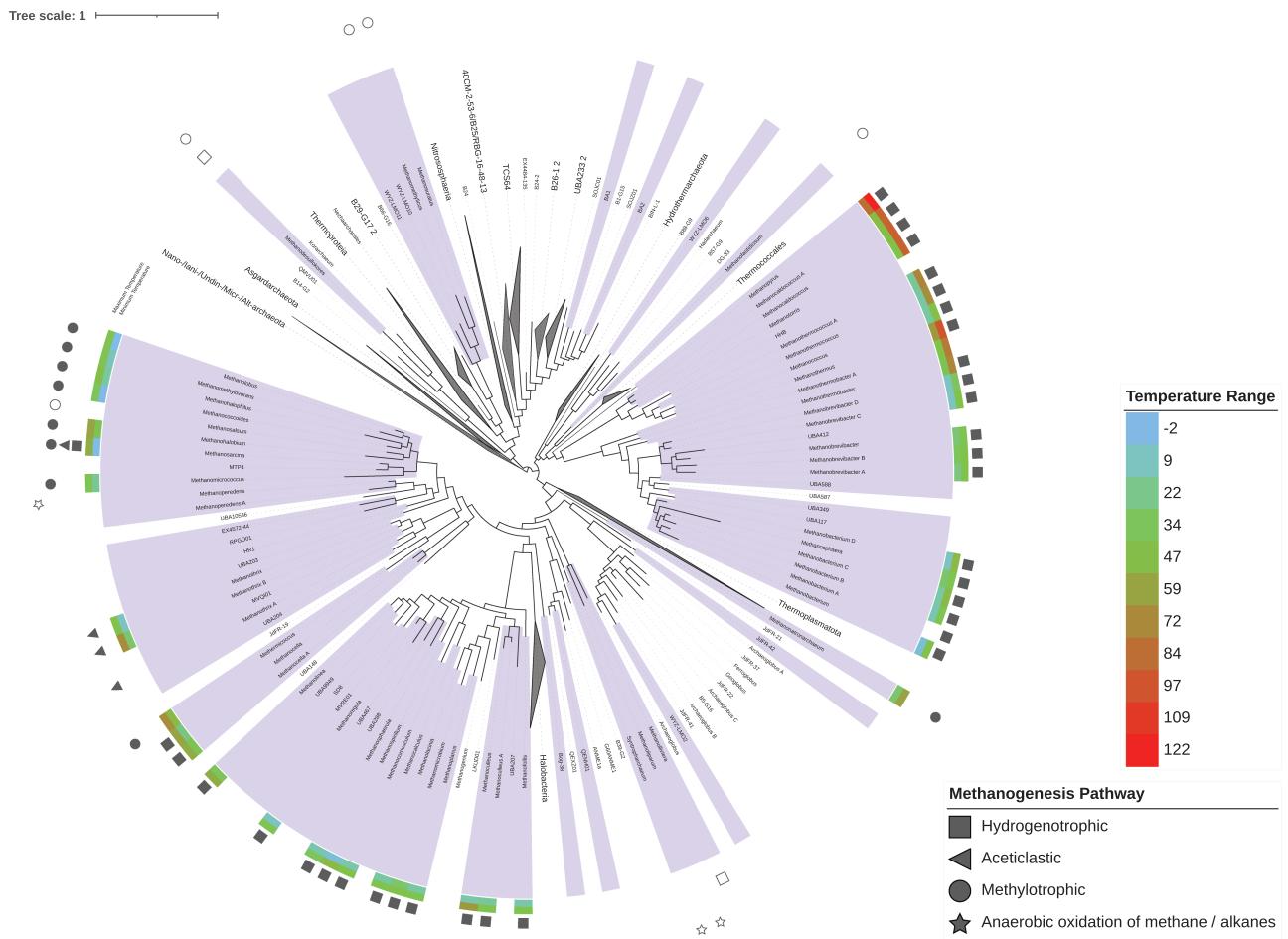


Figure 1. Genera containing mcrABG (marked in purple) throughout the archaeal tree of life. Coloured rings indicate minimum and maximum growth temperatures in °C observed within each genus, symbols indicate the substrate used for alkane metabolism. Unfilled symbols indicate putative pathways which were inferred for yet uncultured species.

Methanomicrobia, whereas thermotolerant ones include Methanopyri, Methanobacteria and Methanococci. With the exception of *Methermicoccus shengliensis* (methylotroph) and *Methanothrix B thermoacetophila* (aceticlastic), all methanogens with maximum growth temperatures $\geq 65^{\circ}\text{C}$ are hydrogenotrophic. Thirty species are methylotrophic with a majority of these being psychrotolerant (Methanosarcinales) and 20 being strict methylotrophs. Finally, 17 are aceticlastic, mostly psychrotolerant species (Methanotrichales), with seven strict aceticlastic organisms. Other growth substrates for methanogenesis include butanol, 2-butanol, isobutanol, propanol, 2-propanol, carbon monoxide, cyclopentanol, dimethyl sulfide, ethanol, and formate, though these were omitted from our analysis since these substrates have been less frequently tested for growth compared with hydrogen, acetate and methyl-compounds.³¹

Thermotolerant organisms are associated with the shortest branches to the root of the archaeal species tree (Fig. 2; this figure is also redrawn using minimum and maximum growth temperatures in Supplementary Fig. S2). This follows the initial observation of Woese,⁴⁴ who found that hyperthermophilic archaea were basal in the tree of life. In the case of the methanogens, the combined observation of short branch lengths and small genomes (see below, Fig. 4A) may run counter to a thermo-reduction hypothesis of genome evolution.⁴⁵

The methanogen pan- and core genomes

Eighty-six isolated methanogens have growth temperature data available in the TEMPURA database³⁰ (a species tree is shown in Supplementary Fig. S4). These span 3 phyla, 8 classes, 10 orders, 19 families and 45 genera. A total of 211,659 genes from these 86 organisms were grouped into 10,131 orthogroups, out of which 791 orthogroups containing 2,329 genes are unique to a single genome. The core includes 330 orthogroups, which are shared by all 86 species, while 552 orthogroups belong to the extended core, which is shared by more than 95% of species. We adopt this relaxed measure proposed by Charlebois and Doolittle³⁵ to account for partially incomplete genomes.

The extended core of these 86 methanogens consists largely of genes related to biosynthesis (57%), coenzyme transport and metabolism (13%). Eleven percent of genes have unknown functions and the remaining 19% are functionally related to lipid biosynthesis, ion transport, secondary metabolites, intracellular transports, motility, cell wall and membrane biogenesis, cell cycle control or carbohydrate transport (Supplementary Fig. S5). The remaining 8,788 orthogroups found in these species belong to the shared fraction of the pan-genome; those genes shared by at least two but less than 95% of organisms in the analysis. Distributions of the number of orthogroup sizes are shown in Fig. 3A. Extended

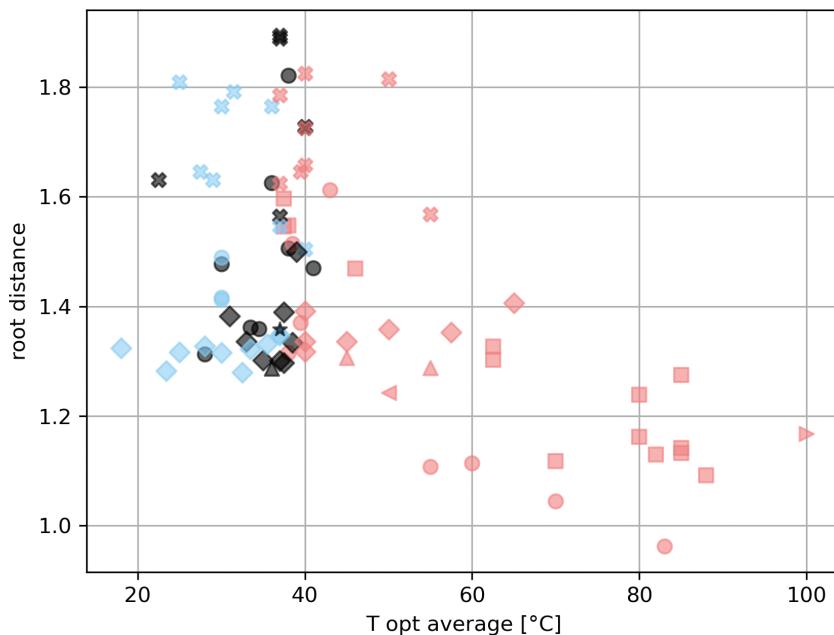


Figure 2. Branch lengths from the root of the archaeal tree to each species versus their optimal growth temperatures. Colours indicate temperature groups used in this study (blue—psychrotolerant, red—thermotolerant, black—no group/mesophilic). Shapes indicate class level taxonomy (circle—Methanobacteria, upward triangle—Methanocellia, square—Methanococci, x—Methanomicrobia, left point triangle—Methanomethanarchaeia, right point triangle—Methanopyri, diamond—Methanosaerica, star—Thermoplasmata).

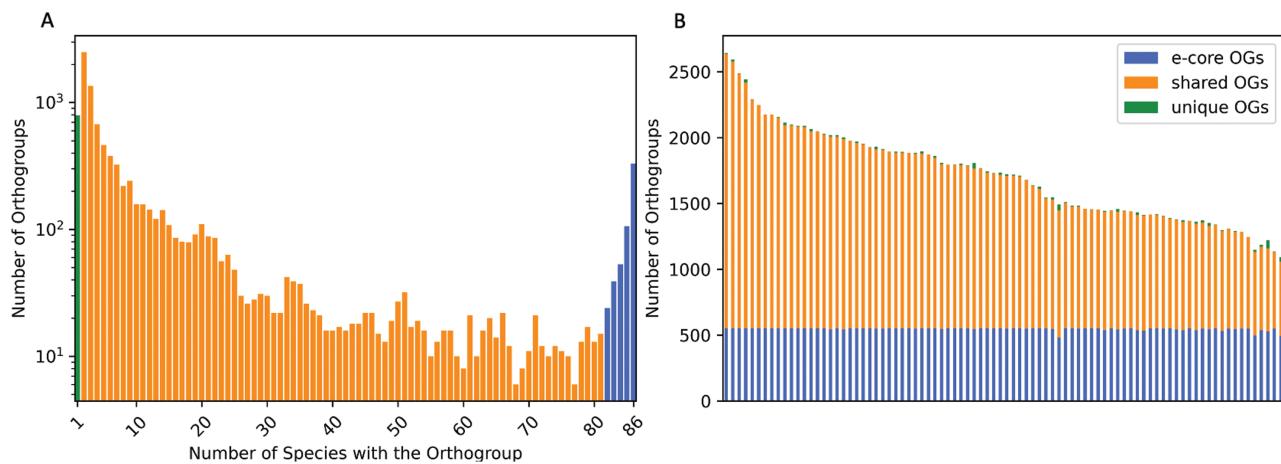


Figure 3. Core- and pangenome analysis of cultivated methanogens. (A) Distribution of the number of orthogroups across the methanogens. (B) Core- and pangenomes of the methanogens, where each bar is one species. Colours in both A and B indicate extended core (shared by at least 95% of the organisms), shared and unique genome fractions.

core- and pan-genome sizes for each genome are shown in Fig. 3B.

The mean size of the extended genomic core is 33.74% (30% in psychrotolerant and 35% in thermotolerant organisms). This fraction is large, considering that these organisms span three different phyla and comparable core sizes can be found at the genus level (e.g. *Comamonas*⁴⁶ or *Streptococcus*⁴⁷). Interestingly, a large core fraction between phylogenetically diverse thermoacidophilic archaea was also observed.⁴⁸ In that case, the constraints of life at low pH were thought to contribute to the large shared genome core. For the methanogens however, the environmental conditions are not constrained: in addition to temperature, growth pH of the

organisms studied here ranges from 4.3 to 10,³¹ suggesting that the methanogen core genome may be constrained by the biochemical requirements of methanogenesis.

The addition of more genomes into core comparison analyses will reduce the core size.³⁵ We find that when including all genomes that contain McrABG and have a checkM score of ≥ 90 , the extended core contains 403 orthogroups, which is a 27% reduction in comparison to the core size observed when including only the 86 cultivated members. This set includes 255 genomes from methanogens, methanotrophs and other alkane-activating archaea: it appears to be the common core for alkane metabolizing archaea (all orthogroups for this analysis in Supplementary Table S2).

Temperature and methanogen genome content

Thermal adaptation in the methanogens does not appear to be accomplished by the use of specific orthologous gene sets: There is neither a thermotolerant nor psychrotolerant specific core (i.e. orthogroups found in all organisms of one temperature group but not in the other). This being said, there are genes found exclusively in one temperature but not the other. Thermotolerant organisms are found with (metal) transporter genes and transcriptional regulators and contain the DNA repair gene *rad51* and reverse gyrase, while psychrotolerant organisms have genes related to membrane and salt transport, membrane fusion, signaling and motility, chaperones and the DNA repair gene *recN* (comprehensive list in [Table 1](#)). Since these genes are not conserved in the temperature groups, they might be related to environmental factors other than temperature, such as high metal concentrations at hydrothermal vent systems, or high salinity in cold environments requiring additional membrane proteins and salt transporters in psychrotolerant organisms.

It may be that reverse gyrase is the only conserved and specific hyperthermophile protein,^{49,50} while chaperones and genes involved in membrane and cell wall biogenesis have been implicated in psychrotolerant species.⁵¹ Our analysis

suggests that thermal adaptation in the methanogens may be largely accomplished by mechanisms other than the use of specific genes—perhaps as detailed below by amino acid substitution patterns.

In a previous analysis with hyperthermophilic bacteria and archaea,⁵² 107 COGs specific to hyperthermophilic archaea were identified out of which 58 are shared across the domains. From these 58 COGs, those with known functions belong to predicted DNA repair systems, metabolic functions ($n = 9$), replication and repair genes ($n = 4$), and one each to translation-related genes and the COG category transcription-related genes and cellular processes. In our analysis, the majority of temperature group specific orthogroups have no functional assignment, and these may be experimental targets for understanding the molecular basis of thermotolerance.

Temperature and methanogen genome size

The genomes of thermotolerant species are on average ~30% smaller than those of organisms growing at colder temperatures ([Fig. 4A](#); also see [Supplementary Fig. S3](#) for a comparison of genome size with minimum and maximum growth temperature), a trend that is not unique to methanogens and has been observed throughout the bacteria and archaea;³⁰

Table 1. Selected temperature group specific functional proteins

Psychrotolerant specific	COG category	Thermotolerant specific
VsrA	L	TopG, Rgy Reverse gyrase
DNA mismatch endonuclease		
RadN	L	Rad51 DNA repair
DNA repair		
YtrF	V	EtaA Exfoliative toxin A/B
Acetoin utilization transporter		
BlpB	VM/M	ZipB, TC.BAT1, OpuD, BetL, BetT, BetS, ChaA, ChrA, PhaA Electrochemical potential driven transporters
Membrane fusion protein		
WzxC	M/P	AfuA, Afub Iron transporters
Lipopolysaccharide transporter		
RscC	M/P	MntA, MntB, MntC Manganese transporter
Capsule synthesis		
RfpC	T/P	MtsA Iron/manganese/copper transporter
Cell-to-cell signaling		
K20974	T/P	PsaA, ScaA, SloC Zinc/manganese transporter
Swarming activity and biofilm formation		
HtpG	O/P	ABC.MN.A, ABC.MN.P, ABC.MN.S Manganese/iron transporter
Molecular chaperone		
WprA	O/P	SLC39A9 Metal ion transporter
Cell wall associated protease		
ChpA	W/P	SLC42A Rh ammonium transporter
Chemosensory pili system protein		
MshQ	W/P	FieF Ferrous iron efflux pump
MSHA pilus biogenesis protein		
SLC7A4	E/J	SelB Selenocysteine specific elongation factor
Cationic amino acid transporter/glycoprotein associated		
YfeH	-U	CpaA, TadV, CpaE, TadZ Tight adherence export apparatus
Sodium bile salt cotransporter		
	—	SLC24A3 Na ⁺ /Ca ²⁺ – K ⁺ exchanger

Selected proteins found in one, but not the other temperature group. COG categories: L: replication, recombination and repair, V: defense mechanism, M: cell Wall/membrane/envelope biogenesis, P: inorganic ion transport and metabolism, T: signal transduction mechanism, O: post-translational modification, protein turnover and chaperones, W: extracellular structures, E: amino acid transport and metabolism, J: translation, ribosomal structure and biogenesis, U: intracellular trafficking, secretion and vesicular transport. The slash mark in the COG column separates the COG category designation for the genes listed on the left or right.

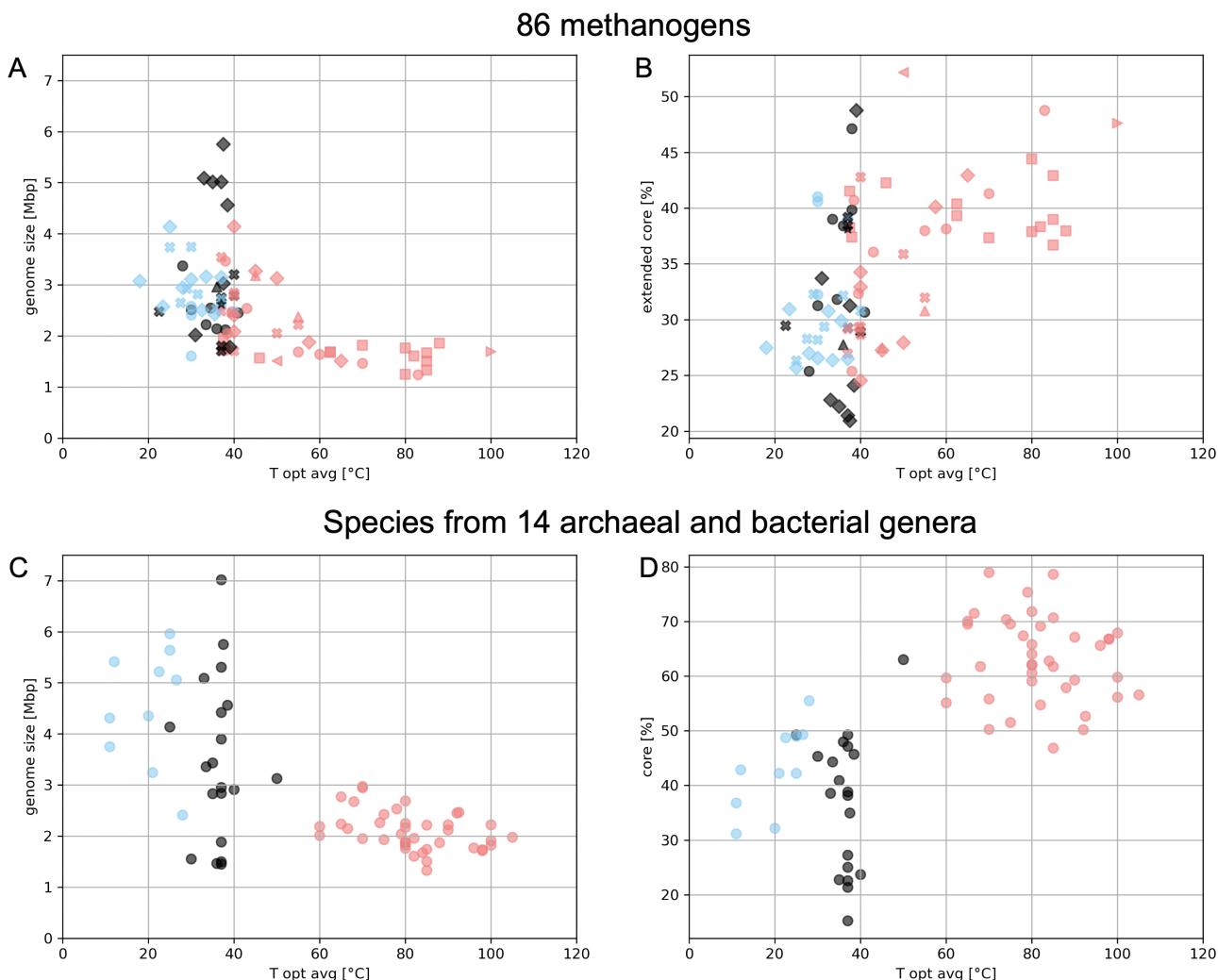


Figure 4. (A) Distribution of genome sizes with T_{opt} for 86 methanogen species. (B) Core sizes with T_{opt} for 86 methanogen species. (C) Distribution of genome sizes with T_{opt} for each species across the genera Pyrococcus, Pyrobaculum, Thermococcus, Methanocaldococcus, Acidianus, Thermotoga, Caldicellulosiruptor, Fervidobacterium, Acholeplasma, Bacteroides, Eubacterium, Methanoscincina, Cryobacterium and Paraglacielcola. Five genomes from each genus were randomly chosen so as to have equal genomic representation between the different groups. (D) Core sizes with T_{opt} for each species across the aforementioned genera. Colours in (A)–(D) indicate temperature groups used in this study (blue—psychrotolerant, red—thermotolerant, black—no group/mesophilic). Shapes in (A) and (B) indicate class level phylogenies (circle—Methanobacteria, upward triangle—Methanocellia, square—Methanococci, x—Methanomicrobia, left point triangle—Methanonatronarchaeia, right point triangle—Methanopyri, diamond—Methanosarcina, star—Thermoplasmata).

Fig. 4C). Among all methanogens, genome size ranges from 1.24 Mbp (*Methanothermus fervidus*) to 5.75 Mbp (*Methanosarcina acetivorans*). For psychrotolerant species the mean genome size is 2.85 Mbp with 2748 ORFs (min. 1.61 Mbp in *Methanobrevibacter D filiformis*, max. 4.14 Mbp in *Methanosarcina lacustris*), while for their thermotolerant counterparts the mean genome size is 2.12 Mbp with 2096 ORFs (min. 1.24 Mbp in *Methanothermus fervidus*, max. 4.14 Mbp in *Methanosarcina mazei*).

To investigate if core size is correlated with growth temperature, the 86 methanogens and 5 species each from a total of 14 bacterial and archaeal genera were plotted. As temperature increases, genome size decreases while the core fraction increases both for the methanogens and archaea and bacteria (Fig. 4B and D). The two methanogenic genera included in these analyses (*Methanocaldococcus* and *Methanosarcina*, both containing more than five species)

group with other genera of similar growth temperature ranges (Fig. 4D).

Substrate-specific orthogroups

Between the substrate groups of aceticlastic, methylotrophic and hydrogenotrophic methanogens, the methylotrophs contain specific methyltransferases as has previously been reported,^{53–55} while hydrogenotrophs contain hydrogenases and dehydrogenases as described by Costa and Leigh.⁵⁶ Interestingly, we find that aceticlastic methanogens have no unique genes related to methane metabolism (comprehensive list in [Supplementary Table S3](#)). This could indicate a general ability of methanogens to metabolize acetate, which has been observed in the hydrogenotroph *Methanococcus maripaludis*.⁵⁷ In addition to considering variations with respect to carbon chemistry, Lyu and Lu⁵⁸ showed genome

content differences in Class I and Class II methanogens (those formerly known as Methanobacteriota and Halobacteriota), with the latter having greater oxidative adaptation abilities.

Phylogenetic distance negatively correlates with shared orthogroups

Across the phylogenetic diversity of methanogens, we compared the pairwise number of orthogroups shared between two genomes at the genus, family and phylum level. Excluding those clades with only one genome member, our comparisons include 180 pairwise comparisons in 23 genera, 884 in 12 families and 3640 in 2 phyla (Fig. 5). We find that within a genus, two species can share as little as 40% of their orthogroups (within the genus *Methanoscincina*) and a decrease in shared fractions with increasing phylogenetic distance. On average 67.95% of orthogroups are shared between two species within a genus, 54.66% of orthogroups are shared between two species in a family and 43.27% of orthogroups are shared between two species in a phylum. Our observations are related to those presented by Touchon et al.,⁵⁹ who found that gene repertoire relatedness decreases with increasing phylogenetic distance, but more work is needed to understand how genome content varies at different taxonomic depths.

In a previous analysis including two hydrogenotrophic species from the *Methanothermobacter* genus,²³ orthogroups shared between the two genomes were largely related to CO₂ reduction and energy conservation, as well as the translocation of metals and ions other than sodium. Between those two species from the same genus, 85% and 91% of genomic content was shared respectively. In another study, Borton et al.²⁴ showed that across eleven genomes within the genus *Methanohalophilus*, most core genes accounted for house-keeping and methylotrophic methanogenesis-related functions. The core accounted for only 42.5–49.4% of the total genome size, which may be attributed to the inclusion of MAGs that tend to have lower completeness and quality scores.

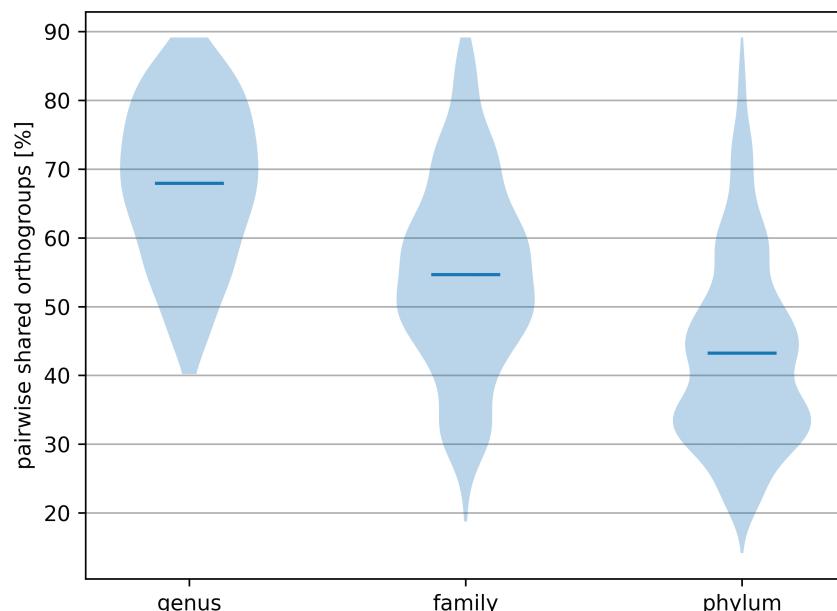


Figure 5: Pairwise shared orthogroups across clades. Species–species comparisons within each genus ($n = 180$ in 23 genera), family ($n = 884$ in 12 families) and phylum ($n = 3640$ in 2 phyla). The mean is marked with a solid line.

Amino acid composition of the proteome

It has been suggested that compositional biases in amino acid content are not observable at the whole genome level,⁶⁰ however for the methanogens, we observe compositional shifts when comparing methanogens at both the whole proteome and conserved extended core levels. Out of 20 amino acids, 14 show significant correlations with the optimum growth temperature of organisms, with slightly more pronounced differences in the shared proteome compared to the extended core (Fig. 6). Polar uncharged amino acids are increased in the psychrotolerant species, while lysine, glutamic acid, leucine and isoleucine are increased in thermotolerant species. In psychrotolerant organisms, amino acid substitutions increase protein flexibility and specific activity and decrease thermostability,^{61,62} while in thermotolerant organisms stability is increased.⁶³

Our results are in line with Saunders et al.,⁶⁴ who compared then available methanogen species ($n = 9$) and found an almost linear trend in the content of glutamine, threonine and leucine over the range of studied growth temperatures (15–98°C) with glutamine and threonine being more abundant and leucine being less abundant in cold-tolerant archaea. Our results are also in agreement with a previous study of the methanogenic genus *Methanococcus*, which found that amino acid biases are affected mostly by optimal growth temperature and not by phylogenetic distance,⁶⁵ and most recently with the result of Lecocq et al.²² who found that amino acid frequency variation is the main driver in temperature adaptation in the order *Methanococcales*.

Interestingly, amino acid compositional trends were generally found to be uncorrelated between psychrotolerant and thermotolerant species analysed at the family level.⁶⁶ Our results suggest that for the methanogens, a shared biochemical repertoire underlies coordinated amino acid changes across the growth temperature range.

The amino acid composition of a proteome has a direct impact on its structural and energetic features: We find that

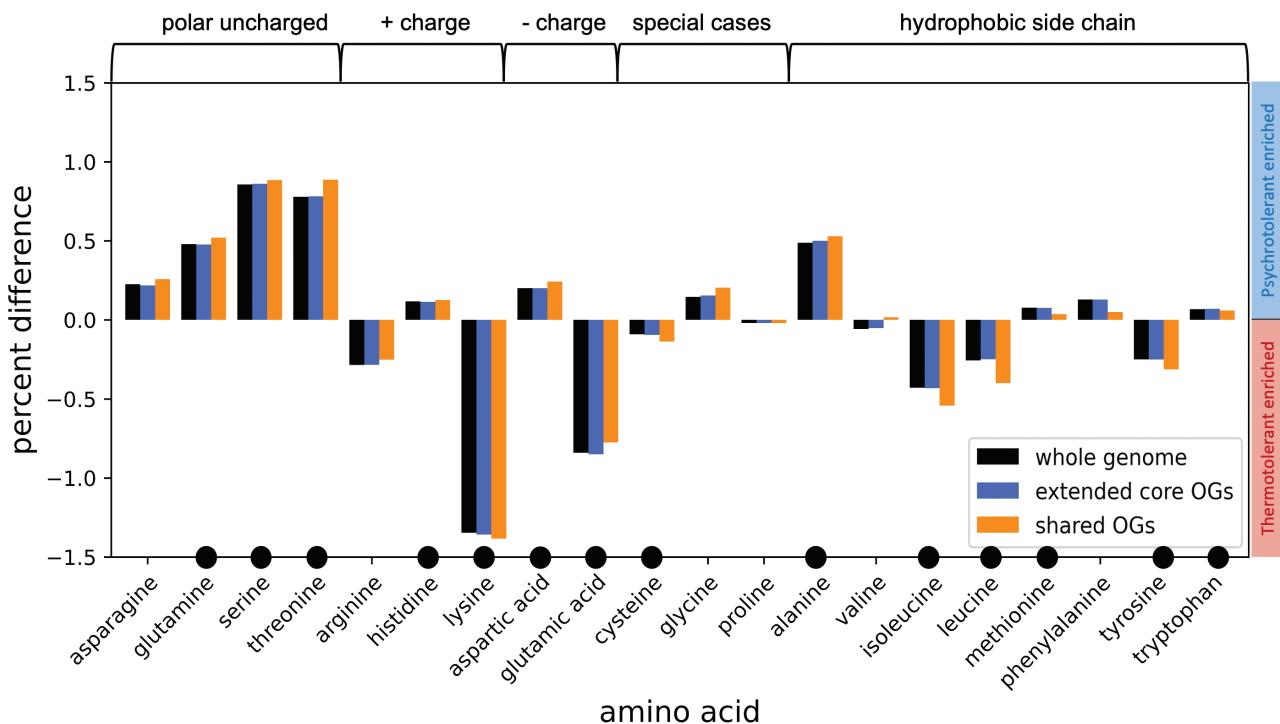


Figure 6: Amino acid compositional differences between temperature groups. Positive percentages indicate amino acid enrichment in psychrotolerant species, and negative percentages indicate amino acid enrichment in thermotolerant species. Percentages were obtained by averaging the relative amino acid abundance over the proteome/extended core orthogroups/shared orthogroups of the two temperature groups and subtracting the thermotolerant mean from the psychrotolerant mean. Circles denote amino acids that have significant correlation with temperature across the whole range of temperatures and for all 86 species, when optimal growth temperature was plotted against each amino acid's abundance (Benjamini Hochberg correction on Pearson P-values). Amino acid abundances for each OG grouping can be found in [Supplementary Table S4](#).

the proportion of charged amino acids and hydrophobicity have positive correlations with optimum growth temperature, while amino acid properties associated with structural complexity such as turn and coil tendencies, have negative correlations with the optimum growth temperature (Fig. 7; Table 2). Comparing of unfolding entropy and enthalpy (Fig. 7B and C) reinforces our understanding that protein folding is entropically driven. Some of these properties have been described for thermophilic enzymes in Ando et al.,⁶⁷ where the authors propose mechanisms that increase the stability of those enzymes, including entropic stabilization.⁶⁸ Interestingly, although the average isoelectric point of amino acids in the proteome increases with increasing temperature, the charge state of the amino acids across the whole proteomes of hyperthermophilic methanogens may be closer to zero compared to their mesophilic relatives since the average amino acid isoelectric point is closer to neutral water at higher temperatures (Supplementary Fig. S6; recall that neutral pH changes from 7 at 25°C to 6.1 at 100°C¹). It is also possible that the increase in lysine observed in thermotolerant organisms drives the increase in proteome isoelectric point (Fig. 7).

An important and interesting outlier in our study is *Methanopyrus kandleri*, which grows at the highest temperature of the methanogens at 122°C. Surprisingly, this organism exhibits many proteome properties, such as bulkiness, flexibility, helical contact area, solvent accessible surface area and volume, more similar to mesophilic than other thermotolerant organisms (Supplementary Fig. S6; *Methanopyrus kandleri* is found near the 100°C mark of its optimal temperature). This observation raises questions about the mechanisms of

thermal adaptation and the use of amino acid profiles to predict growth temperature. For example, some studies propose the use of certain amino acid ratios (e.g. charged vs. polar amino acids) to predict optimal growth temperatures of un-cultured organisms,⁶⁹ but *Methanopyrus kandleri* seems to challenge our ability to make generalizations. Although it is unknown how *Methanopyrus kandleri* can live at high temperatures without altering its proteome amino acid composition as other methanogens do, it is tempting to speculate that the organism may have an altered cytosol content in comparison to other methanogens, but work is needed to clarify this. We note that protein compositional differences are associated with other environmental variables other than temperature; for example, pH, oxidation state⁷⁰ and salinity.⁷¹ Future work is needed to build an integrative picture of how multiple variables affect the amino acid composition of the proteome.

Gene gain and loss in methanogens

We used a species tree of the 86 species to investigate a possible phylogenetic relationship between gene gain-loss-duplication with temperature. Although the thermotolerant organisms have smaller genomes, correlations between temperature and gain-loss-duplication were not observed (Supplementary Fig. S7). This observation leaves open questions as to how the temperature:genome size relationship (Fig. 4A) came to be.

Thomas et al.⁷² showed that in comparison with 21 other Methanosaecinales, the genome of the rumen dwelling and mesophilic methanogen *Methanosaecina blatticola* shows the largest genome reduction, with the loss of almost all genes

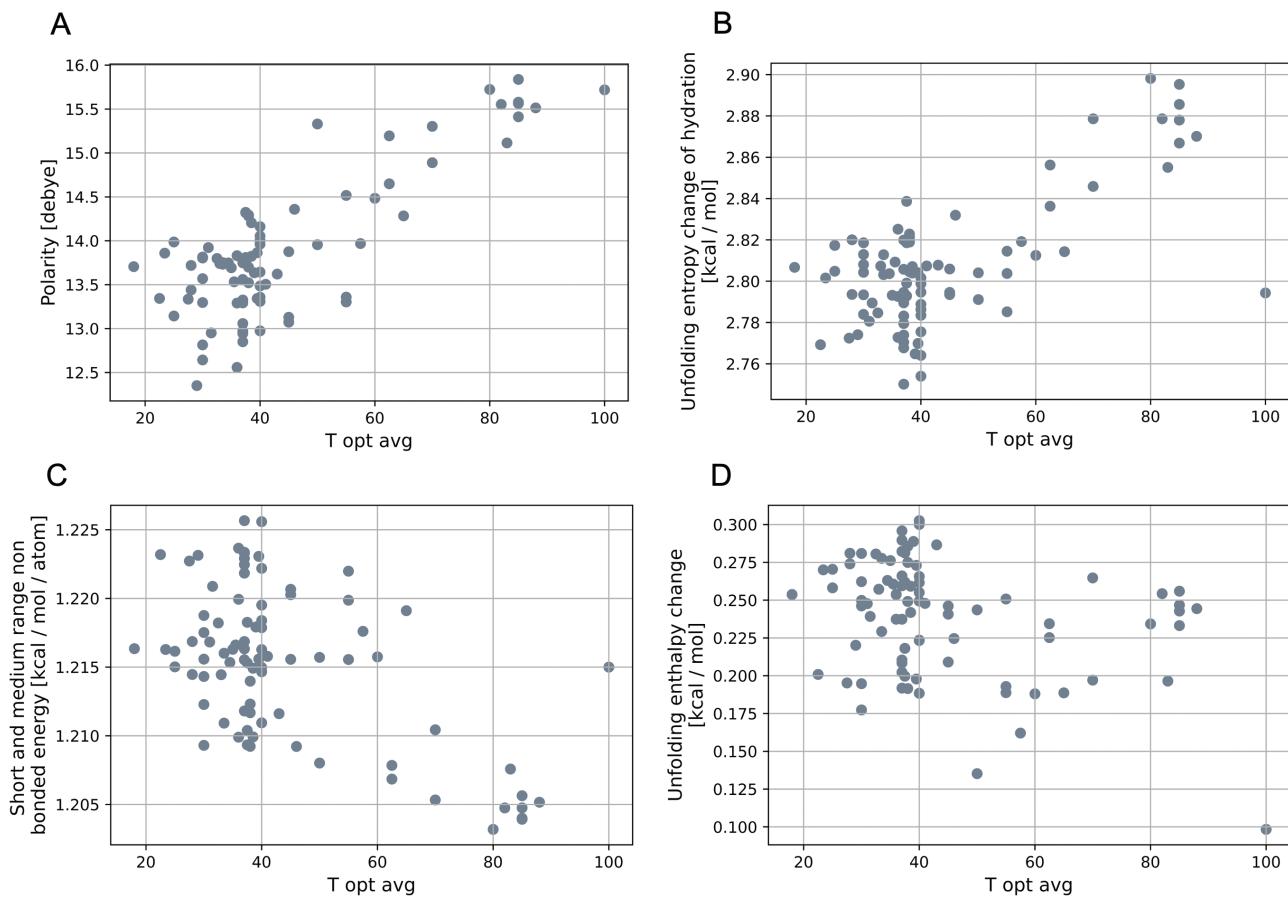


Figure 7. Selected proteome properties in relation to optimal growth temperature ($^{\circ}\text{C}$): polarity (A), unfolding entropy change of hydration (B), short and medium range non bonded energy (C) and unfolding enthalpy change (D). Plots for other amino acid properties are available in [Supplementary Fig. S6](#).

Table 2. Proteome amino acid properties with temperature correlations

Increase with temperature	Decrease with temperature
Isoelectric point (0.812)	Coil tendencies (-0.677)
Polarity (0.808)	Turn tendencies (-0.617)
Refractive index (0.737)	Short and medium range non bonded energy (-0.567)
Thermodynamic transfer hydrophobicity (0.715)	Sombined surrounding hydrophobicity globular and membrane (-0.356)
Unfolding entropy change of hydration (0.690)	Total non-bonded energy (-0.350)
Unfolding hydration heat capacity change (0.686)	Unfolding enthalpy change (-0.318)
Helical contact area (0.675)	Unfolding enthalpy change of hydration (-0.311)
Average medium contacts (0.637)	Buriedness (-0.264)
Solvent accessible surface area for denatured protein (0.622)	Compressibility (-0.251)
Flexibility (0.579)	Chromatographic index (-0.225)
Solvent accessible reduction ratio (0.561)	Unfolding Gibbs free energy change (-0.218)
Bulkiness (0.560)	Long range non-bonded energy (-0.058)
Volume (0.535)	
Molecular weight (0.528)	
Solvent accessible surface area for native protein (0.523)	
Unfolding entropy change (0.339)	
Shape (0.241)	

Protein amino acid properties with Significant correlation to temperature. The Pearson correlation coefficient for each property is in parentheses. Scatterplots showing the data are available in [Supplementary Fig. S6](#).

involved in the H₄MPT (tetrahydromethanopterin) methyl branch of the Wood–Ljungdahl pathway. In line with this, our analysis shows that *Methanosarcina blatticola* has the highest reduction with a net gain rate of −0.36, while other Methanosarcinales show both net gains and losses. This clade-specific variance within the Methanosarcinales suggests that gene gain–loss–duplication rates are not solely governed by the phylogenetic placement of organisms, and it may be that this type of variability is part of what clouds our ability to observe a temperature gain–loss–duplication rate relationship.

Conclusions

We observed differences in genome size and functional composition across the genomes of methanogens spanning a wide growth temperature range. We show that differences in amino acid composition can be found at the whole genome level as well as in the functionally constant extended core. Thermotolerant methanogens have smaller, reduced genomes with an increase in charged amino acids and functional genes related to ion transport which are not found in their psychrotolerant counterparts. The latter have larger genomes with an increase in polar uncharged amino acids and additional functional genes related to cell structure and mobility. The organisms in our core genome analysis span three phyla but share a large extended core, suggesting that physiological similarity has a stronger effect on temperature adaptation than phylogenetic distance in the methanogens. We show that core size decreases with the increased phylogenetic distance between the methanogenic archaea, and it will be interesting to explore this trend across the tree of life.

The expansion of our knowledge of alkane-utilizing archaea, their phylogenies, metabolisms and genomic features has widened our understanding of archaeal evolutionary history in the past decade.^{7,69,73–80} Further studies focusing on their genome evolution can help us further narrow down the first appearances of methanogens and methanotrophs in Earth’s history. The type of study conducted here could be expanded across a greater diversity of organisms, further helping to clarify how environment, physiology and evolution each contribute to genomic content.

Acknowledgements

Computational work was partly carried out at the Supercomputer System, Institute for Chemical Research, Kyoto University. We thank two anonymous reviewers for comments which improved the manuscript.

Funding

P.P. was supported by JST SPRING, Grant Number JPMJSP2106.

Conflict of interest

The authors declare no conflict of interest.

Supplementary Data

Supplementary data are available at *DNARES* online.

References

- Amend, J.P., and Shock, E.L. 2001, Energetics of overall metabolic reactions of thermophilic and hyperthermophilic archaea and bacteria, *FEMS Microbiol. Rev.*, **25**, 175–243.
- Gillooly, J.F., Brown, J.H., West, G.B., Savage, V.M., and Charnov, E.L. 2001, Effects of size and temperature on metabolic rate, *Science* (80-), **293**, 2248–51.
- Arroyo, J.I., Díez, B., Kempes, C.P., West, G.B., and Marquet, P.A. 2022, A general theory for temperature dependence in biology, *Proc. Natl. Acad. Sci.*, **119**, e2119872119.
- Woese, C.R., Kandler, O., and Wheelis, M.L. 1990, Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya, *Proc. Natl. Acad. Sci.*, **87**, 4576–9.
- Brochier, C., and Philippe, H. 2002, A non-hyperthermophilic ancestor for bacteria, *Nature*, **417**, 244.
- Forterre, P. 2015, The universal tree of life: an update, *Front. Microbiol.*, **6**, 717.
- Martinez-Gutierrez, C.A., and Aylward, F.O. 2021, Phylogenetic signal, congruence, and uncertainty across bacteria and archaea, *Mol. Biol. Evol.*, **38**, 5514–27.
- Franzmann, P.D., Springer, N., Ludwig, W., De Macario, E.C., and Rohde, M. 1992, A methanogenic archaeon from Ace Lake, Antarctica: *Methanococcoides burtonii* sp. nov, *Syst. Appl. Microbiol.*, **15**, 573–81.
- Takai, K., Nakamura, K., Toki, T., et al. 2008, Cell proliferation at 122 °C and isotopically heavy CH₄ production by a hyperthermophilic methanogen under high-pressure cultivation, *Proc. Natl. Acad. Sci.*, **105**, 10949–54.
- Thauer, R.K. 1998, Biochemistry of methanogenesis: a tribute to Marjory Stephenson: 1998 Marjory Stephenson prize lecture, *Microbiology*, **144**, 2377–406.
- Liu, Y., and Whitman, W.B. 2008, Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea, *Ann. N. Y. Acad. Sci.*, **1125**, 171–89.
- Lovley, D.R., Greening, R.C., and Ferry, J.G. 1984, Rapidly growing rumen methanogenic organism that synthesizes coenzyme M and has a high affinity for formate, *Appl. Environ. Microbiol.*, **48**, 81–7.
- Torres-Alvarado, R., Ramírez-Vives, F., Fernández, F.J., and Barriga-Sosa, I. 2005, Methanogenesis and methane oxidation in wetlands: implications in the global carbon cycle. *Hidrobiología*, **15**, 327–49.
- Jones, W.J., Stugard, C.E., and Jannasch, H.W. 1989, Comparison of thermophilic methanogens from submarine hydrothermal vents, *Arch. Microbiol.*, **151**, 314–8.
- Boyd, E.S., Skidmore, M., Mitchell, A.C., Bakermans, C., and Peters, J.W. 2010, Methanogenesis in subglacial sediments, *Environ. Microbiol. Rep.*, **2**, 685–92.
- Kreil, D.P., and Ouzounis, C.A. 2001, Identification of thermophilic species by the amino acid compositions deduced from their genomes, *Nucleic Acids Res.*, **29**, 1608–15.
- Vieille, C., and Zeikus, G.J. 2001, Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability, *Microbiol. Mol. Biol. Rev.*, **65**, 1–43.
- Tekaia, F., Yeramian, E., and Dujon, B. 2002, Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis, *Gene*, **297**, 51–60.
- Farias, S.T., and Bonato, M.C. 2003, Preferred amino acids and thermostability, *Genet. Mol. Res.*, **2**, 383–93.
- Zeldovich, K.B., Berezovsky, I.N., and Shakhnovich, E.I. 2007, Protein and DNA sequence determinants of thermophilic adaptation, *PLoS Comput. Biol.*, **3**, e5.
- Smole, Z., Nikolic, N., Supek, F., et al. 2011, Proteome sequence features carry signatures of the environmental niche of prokaryotes, *BMC Evol. Biol.*, **11**, 1–10.
- Lecocq, M., Groussin, M., Gouy, M., and Brochier-Armanet, C. 2021, The molecular determinants of thermoadaptation: methanococcales as a case study, *Mol. Biol. Evol.*, **38**, 1761–76.

23. Kaster, A.-K., Goenrich, M., Seedorf, H., et al. 2011, More than 200 genes required for methane formation from H_2 and CO_2 and energy conservation are present in *Methanothermobacter marburgensis* and *Methanothermobacter thermautotrophicus*, *Archaea*, 2011, 1–23.
24. Borton, M.A., Daly, R.A., O'Banion, B., et al. 2018, Comparative genomics and physiology of the genus *Methanohalophilus*, a prevalent methanogen in hydraulically fractured shale, *Environ. Microbiol.*, 20, 4596–611.
25. Poehlein, A., Schneider, D., Soh, M., Daniel, R., and Seedorf, H. 2018, Comparative genomic analysis of members of the genera Methanospaera and Methanobrevibacter reveals distinct clades with specific potential metabolic functions, *Archaea*, 2018, 1–9.
26. Mendler, K., Chen, H., Parks, D.H., et al. 2019, AnnoTree: visualization and exploration of a functionally annotated microbial tree of life, *Nucleic Acids Res.*, 47, 4442–8.
27. Kanehisa, M., Araki, M., Goto, S., et al. 2007, KEGG for linking genomes to life and the environment, *Nucleic Acids Res.*, 36, D480–4.
28. Parks, D.H., Chuvochina, M., Rinke, C., et al. 2022, GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy, *Nucleic Acids Res.*, 50, D785–94.
29. Brochier-Armanet, C., Forterre, P., and Gribaldo, S. 2011, Phylogeny and evolution of the Archaea: one hundred genomes later, *Curr. Opin. Microbiol.*, 14, 274–81.
30. Sato, Y., Okano, K., Kimura, H., and Honda, K. 2020, TEMPURA: database of growth TEMPeratures of usual and RAre prokaryotes, *Microbes Environ.*, 35, ME20074.
31. Michał, B., Gagat, P., Jabłoński, S., et al. 2018, PhyMet2: a database and toolkit for phylogenetic and metabolic analyses of methanogens, *Environ. Microbiol. Rep.*, 10, 378–82.
32. Letunic, I., and Bork, P. 2021, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, *Nucleic Acids Res.*, 49, W293–6.
33. Emms, D.M., and Kelly, S. 2019, OrthoFinder: phylogenetic orthology inference for comparative genomics, *Genome Biol.*, 20, 1–14.
34. Lechner, M., Findeiß, S., Steiner, L., et al. 2011, Proteinortho: detection of (co-) orthologs in large-scale analysis, *BMC Bioinf.*, 12, 124.
35. Charlebois, R.L., and Doolittle, W.F. 2004, Computing prokaryotic gene ubiquity: rescuing the core from extinction, *Genome Res.*, 14, 2469–77.
36. Huerta-Cepas, J., Forslund, K., Coelho, L.P., et al. 2017, Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper, *Mol. Biol. Evol.*, 34, 2115–22.
37. Gromiha, M.M., Oobatake, M., and Sarai, A. 1999, Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins, *Biophys. Chem.*, 82, 51–67.
38. Seabold, S. and Perktold, J. 2010, Statsmodels: econometric and statistical modeling with python. In Proceedings of the 9th Python in Science Conference. Austin, TX, pp. 10–25080.
39. Genovese, C.R., Lazar, N.A., and Nichols, T. 2002, Thresholding of statistical maps in functional neuroimaging using the false discovery rate, *Neuroimage*, 15, 870–8.
40. Csűös, M. 2010, Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood, *Bioinformatics*, 26, 1910–2.
41. Csűös, M., and Miklós, I. 2009, Streamlining and large ancestral genomes in Archaea inferred with a phylogenetic birth-and-death model, *Mol. Biol. Evol.*, 26, 2087–95.
42. Borrel, G., O'Toole, P.W., Harris, H.M.B., et al. 2013, Phylogenomic data support a seventh order of methylotrophic methanogens and provide insights into the evolution of methanogenesis, *Genome Biol. Evol.*, 5, 1769–80.
43. Shao, N., Fan, Y., Chou, C.-W., et al. 2022, Expression of divergent methyl/alkyl coenzyme M reductases from uncultured archaea, *Commun. Biol.*, 5, 1–11.
44. Woese, C.R. 1987, Bacterial evolution, *Microbiol. Rev.*, 51, 221–71.
45. Forterre, P. 1995, Thermoreduction, a hypothesis for the origin of prokaryotes, *C. R. Acad. Sci. III*, 318, 415–22.
46. Wu, Y., Zaiden, N., and Cao, B. 2018, The core- and pan-genomic analyses of the genus *Comamonas*: from environmental adaptation to potential virulence, *Front. Microbiol.*, 9, 3096.
47. Barajas, H.R., Romero, M.F., Martínez-Sánchez, S., and Alcaraz, L.D. 2019, Global genomic similarity and core genome sequence diversity of the *Streptococcus* genus as a toolkit to identify closely related bacterial species in complex environments, *PeerJ*, 2019, e6233.
48. Fütterer, O., Angelov, A., Liesegang, H., et al. 2004, Genome sequence of *Picrophilus torridus* and its implications for life around pH 0, *Proc. Natl. Acad. Sci.*, 101, 9091–6.
49. Forterre, P. 2002, A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein, *Trends Genet.*, 18, 236–7.
50. Catchpole, R.J., and Forterre, P. 2019, The evolution of reverse gyrase suggests a nonhyperthermophilic last universal common ancestor, *Mol. Biol. Evol.*, 36, 2737–47.
51. Wang, F., Wang, J., Jian, H., et al. 2008, Environmental adaptation: genomic analysis of the piezotolerant and psychrotolerant deep-sea iron reducing bacterium *Shewanella piezotolerans* WP3, *PLoS One*, 3, e1937.
52. Makarova, K.S., Wolf, Y.I., and Koonin, E.V. 2003, Potential genomic determinants of hyperthermophily, *Trends Genet.*, 19, 172–6.
53. Burke, S.A., Lo, S.L., and Krzycki, J.A. 1998, Clustered genes encoding the methyltransferases of methanogenesis from monomethylamine, *J. Bacteriol.*, 180, 3432–40.
54. Borrel, G., Adam, P.S., McKay, L.J., et al. 2019, Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea, *Nat. Microbiol.*, 4, 603–13.
55. Söllinger, A., and Ulrich, T. 2019, Methylotrophic methanogens everywhere—physiology and ecology of novel players in global methane cycling, *Biochem. Soc. Trans.*, 47, 1895–907.
56. Costa, K.C., and Leigh, J.A. 2014, Metabolic versatility in methanogens, *Curr. Opin Biotechnol.*, 29, 70–5.
57. Vo, C.H., Goyal, N., Karimi, I.A., and Kraft, M. 2020, First observation of an acetate switch in a methanogenic autotroph (*Methanococcus maripaludis* S2), *Microbiol. Insights*, 13, 1178636120945300.
58. Lyu, Z., and Lu, Y. 2018, Metabolic shift at the class level sheds light on adaptation of methanogens to oxidative environments, *ISME J.*, 12, 411–23.
59. Touchon, M., Perrin, A., De Sousa, J.A.M., et al. 2020, Phylogenetic background and habitat drive the genetic diversification of *Escherichia coli*, *PLoS Genet.*, 16, e1008866.
60. Feller, G. 2010, Protein stability and enzyme activity at extreme biological temperatures, *J. Phys.: Condens. Matter.*, 22, 323101.
61. Metpally, R.P.R., and Reddy, B.V.B. 2009, Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: insights into the molecular basis of cold adaptation of proteins, *BMC Genomics*, 10, 111–10.
62. De Maayer, P., Anderson, D., Cary, C., and Cowan, D.A. 2014, Some like it cold: understanding the survival strategies of psychrophiles, *EMBO Rep.*, 15, 508–17.
63. Ward, O.P., and Moo-Young, M. 1988, Thermostable enzymes, *Biotechnol. Adv.*, 6, 39–69.
64. Saunders, N.F.W., Thomas, T., Curmi, P.M.G., et al. 2003, Mechanisms of thermal adaptation revealed from the genomes of the Antarctic Archaea *Methanogenium frigidum* and *Methanococcoides burtonii*, *Genome Res.*, 13, 1580–8.
65. Klipcan, L., Safro, I., Temkin, B., and Safro, M. 2006, Optimal growth temperature of prokaryotes correlates with class II amino acid composition, *FEBS Lett.*, 580, 1672–6.
66. Yang, L.-L., Tang, S.-K., Huang, Y., and Zhi, X.-Y. 2015, Low temperature adaptation is not the opposite process of high temperature

- adaptation in terms of changes in amino acid composition, *Genome Biol. Evol.*, 7, 3426–33.
67. Ando, N., Barquera, B., Bartlett, D.H., et al. 2021, The molecular basis for life in extreme environments, *Annu. Rev. Biophys.*, 50, 343–72.
68. Sawle, L., and Ghosh, K. 2011, How do thermophilic proteins and proteomes withstand high temperature? *Biophys. J.*, 101, 217–27.
69. Hua, Z.-S., Wang, Y.-L., Evans, P.N., et al. 2019, Insights into the ecological roles and evolution of methyl-coenzyme M reductase-containing hot spring Archaea, *Nat. Commun.*, 10, 1–11.
70. Dick, J.M., and Shock, E.L. 2011, Calculation of the relative chemical stabilities of proteins as a function of temperature and redox chemistry in a hot spring, *PLoS One*, 6, e22782.
71. Dick, J.M., Yu, M., and Tan, J. 2020, Uncovering chemical signatures of salinity gradients through compositional analysis of protein sequences, *Biogeosciences*, 17, 6145–62.
72. Thomas, C.M., Taib, N., Gribaldo, S., and Borrel, G. 2021, Comparative genomic analysis of *Methanimicrococcus blatticola* provides insights into host adaptation in archaea and the evolution of methanogenesis, *ISME Commun.*, 1, 1–11.
73. Petitjean, C., Deschamps, P., López-García, P., Moreira, D., and Brochier-Armanet, C. 2015, Extending the conserved phylogenetic core of archaea disentangles the evolution of the third domain of life, *Mol. Biol. Evol.*, 32, 1242–54.
74. Berghuis, B.A., Yu, B., Schulz, F., et al. 2019, Hydrogenotrophic methanogenesis in archaeal phylum Verstraetarchaeota reveals the shared ancestry of all methanogens, *PNAS*, 116, 5037–44.
75. Wolfe, J.M., and Fournier, G.P. 2018, Horizontal gene transfer constrains the timing of methanogen evolution, *Nat. Ecol. Evol.*, 2, 897–903.
76. Dombrowski, N., Williams, T.A., Sun, J., et al. 2020, Urdinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution, *Nat. Commun.*, 11, 1–15.
77. Wang, Y., Wegener, G., Williams, T.A., et al. 2021, A methylotrophic origin of methanogenesis and early divergence of anaerobic multicarbon alkane metabolism, *Sci. Adv.*, 7, 1–12.
78. Garcia, P.S., Gribaldo, S., and Borrel, G. 2022, Diversity and evolution of methane-related pathways in archaea, *Annu. Rev. Microbiol.*, 76, 727–55.
79. Laso-Pérez, R., Wegener, G., Knittel, K., et al. 2016, Thermophilic archaea activate butane via alkyl-coenzyme M formation, *Nature*, 539, 396–401.
80. Williams, T.A., Szöllősi, G.J., Spang, A., et al. 2017, Integrative modeling of gene and genome evolution roots the archaeal tree of life, *Proc. Natl. Acad. Sci.*, 114, E4602–11.