

Distribution and evolutionary significance of the DnaK-DnaJ-GrpE (KJE) chaperone system in archaea

Mathieu E. Rebeaud^{*} 

Received: 10 November 2025 / Revised: 16 January 2026 / Accepted: 19 January 2026

© 2026 The Author(s). Published by Elsevier Inc. on behalf of Cell Stress Society International. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Abstract

A prevailing hypothesis for the emergence of life on Earth holds that it might have originated in hydrothermal vents, where the environmental conditions, although physically and chemically extreme (acidity, lack of oxygen, high pressure, very high temperature), vary very little. According to this view, single-celled organisms that appeared under these conditions subsequently began to colonize first all aquatic environments, then terrestrial ones. Here, I study the proteomes of more than 250 reference proteomes of archaea as well as those of a few non-reference Promethearchaeati (ASGARD), which have an optimal growth temperature of between 10 °C and 100 °C. I found a correlation between the chaperone present in these organisms, and in particular the presence/absence of the Hsp70 family (DnaK-DnaJ-GrpE, KJE for brevity), the proteome size, and the optimal growth temperature. These findings suggest that the presence of Hsp70s in mesophilic archaea is associated with larger proteomes and may have facilitated adaptation to more diverse environments.

Keywords DnaK · Evolution · Temperature · Archaea · Chaperones

Background

Since life first appeared on Earth at the bottom of the oceans near hydrothermal vents, some 3.5-4 billion years ago, constant evolution has driven organisms to colonize the entire Earth, taking forms as varied as single-celled bacteria and archaea, plants, fungi, and sponges, right up to metazoans like cats, blue whales and human.¹ Life appeared in conditions that, while extreme if compared with the environment in which most eukaryotes tend to evolve, due to high acidity, lack of oxygen, extreme pressure, and temperatures up to 120 °C, are remarkably stable, unchanging, and, for well-adapted organisms, rather unchallenging.² Indeed,

organisms living in these extreme environments have proteins that are optimally adapted,^{3,4} with high evolutionary pressure not to change anymore.

Singer and Hickey⁴ showed that thermophilic organisms living in elevated growth temperatures are imposed with selective constraints at all three molecular levels; the nucleotide content, codon usage, and amino acid composition. It has been established that the amino acid composition of thermophilic proteins confers enhanced stability of the folded state at high temperatures.^{4,5} Moreover, several studies have shown that hyperthermophiles such as the model organisms *Thermus thermophilus*⁶ and *Sulfolobus acidocaldarius*⁷ have lower mutation rates than their mesophilic counterparts. Indeed, neutral mutations can be compensated in mesophilic conditions, and as a consequence mesophilic organisms tend to evolve more rapidly and diversify more easily.⁸ Instead, the same mutations likely become deleterious in the extreme physical and chemical conditions of hyperthermophiles. Furthermore, in mesophilic bacteria⁹ and archaea,⁸ this selective pressure appears to be more pronounced for highly expressed genes and proteins.^{10,11}

The phylogenomics of 10,575 bacterial and archaeal genomes has recently demonstrated the proximity between

Abbreviations: ATP, Adenosine triphosphate; DNA, Deoxyribonucleic acid; GC, Guanine-Cytosine content; IDP, Intrinsically disordered proteins; TPS, Thermosomes (chaperonins subunits), Prefoldins and Small Heat Shock Proteins (sHSPs) system

^{*} Mathieu E. Rebeaud
mathieu.rebeaud@epfl.ch

Laboratory of Statistical Biophysics, Institute of Physics, School of Basic Sciences, École Polytechnique Fédérale de Lausanne – EPFL, 1015 Lausanne, Switzerland.

the two domains of prokaryotic life, as well as the numerous exchanges between them, mainly through horizontal gene transfer (HGT).^{12,13} The constant presence of HGTs also affects "housekeeping genes," emphasizing the importance of both inter- and intra-domain HGT.¹⁴ Because of the shuffling of the gene pool, tracing the appearance or disappearance of certain genes requires in-depth study of gene families, such as chaperones like Hsp60 (GroEL in bacteria, thermosomes in archaea) or the main component of the chaperone network, Hsp70 (DnaK for the most conserved Hsp70 member in prokaryotes). According to several studies, Hsp70s first appeared in mesophilic bacteria, then was transferred by HGT to archaea.^{1,15,16}

DnaK is considered to be the central hub of the chaperone network, and in *Escherichia coli* about 700, mostly cytosolic, proteins usually interact with DnaK under standard growth conditions.¹⁷ Furthermore, a recent study on the bacterium *Myxococcus xanthus* demonstrated that DnaK duplication and specialization in bacteria correlates with increased proteome complexity.¹⁸ DnaK has already been determined to increase mutational robustness, as well as to allow proteins that are its obligate clients to evolve faster than those that are not by supporting their folding.¹⁹ Therefore, the presence of DnaK, and more in general of the protein quality control (PQC) system, has profound long-term effects on genome evolution.²⁰⁻²² These studies provide evidence that even a single protein, or a limited set of proteins (the chaperones), can have a disproportionate effect on the evolution of the proteome.

This study examines the relationship between proteome size and optimal growth temperature in over 250 mesophilic or thermophilic archaea. Specifically, I analyze this relationship as a function of their chaperome composition, including the presence or absence of Hsp70 and its associated system (DnaJ, GrpE, as well as Hsp90-like HtpG and Hsp100-like ClpB).²³ The aim is to extend our understanding of the relationship between an organism's complexity, its environment, and its chaperone system, this time with a focus on archaea (and some bacteria), unlike our previous study that covered less specifically the whole Tree of Life.¹ In particular, using archaeal reference proteomes, I am testing whether the presence or absence of key components of the PQC, in particular the Hsp70/DnaK system, is associated with proteome size and growth temperature.

Results

Selection of archaea and comparison between optimal growth temperature and GC content of 16S RNA of prokaryotes

Phylogenetic analyses of 16S rRNA genes in thermophilic and hyperthermophilic organisms have consistently shown

a correlation between optimal growth temperature and GC composition, as first demonstrated by Kimura and colleagues²⁴ and later confirmed by Hu et al.²⁵

To investigate this relationship further, I analyzed the 16S rRNA sequences of selected prokaryotes. As expected, GC content in 16S rRNA increases with optimal growth temperature, likely reflecting the enhanced thermodynamic stability of GC-rich sequences²⁵ (Figure 1A). I next looked at the correlation between proteome size and optimal growth temperature, finding that organisms adapted to higher temperatures have on average smaller proteomes (Figure 1B), possibly due to constraints on protein stability or metabolic efficiency. These trends appear to be more prevalent in archaea, which is in keeping with the observation that most hyperthermophiles organisms are of the archaea domain²⁶ (archaea in general seem to be better adapted to more extreme conditions).

Although GC content has been linked to thermal adaptation because of its effect on DNA melting temperature,²⁵ it doesn't fully explain the variations of environment seen among archaeal lineages. Other molecular mechanisms may also contribute to adaptation to different thermal environments, in particular the PQC,²⁰ which is essential for protein folding and stress response in organisms. Thus, I next examined the correlation between the presence or absence of specific chaperone families and growth temperatures in archaea.

Because hyperthermophilic organisms are particularly informative for addressing these questions, the analysis focused primarily on archaeal lineages. The Tempura²⁷ database, which was already curated, was used as a reference to see which prokaryotes are the most interesting ones. Very few archaeal organisms were included in it (Figure 1A and B), so it was necessary to use another database (BacDive²⁸) and select organisms manually from UniProt references proteomes to refine the results for archaea. In total, 234 archaeal reference proteomes were collected (see Methods and Supplementary Information S2). Their average proteome size (2763 ± 960 proteins) was comparable to that of the initial selection (Supplementary Figure 1), which is 2920 ± 880 (see Supplementary Information S1).²⁸

A general, albeit weak, correlation between proteome size and optimal growth temperature was observed across Archaea (Supplementary Figure 2). Other statistical methods were tested against the dataset (Supplementary Figure 3, Supplementary Information S10). This correlation between optimal growth temperature and proteome size also correlates weakly with the taxonomic classification of different archaeal

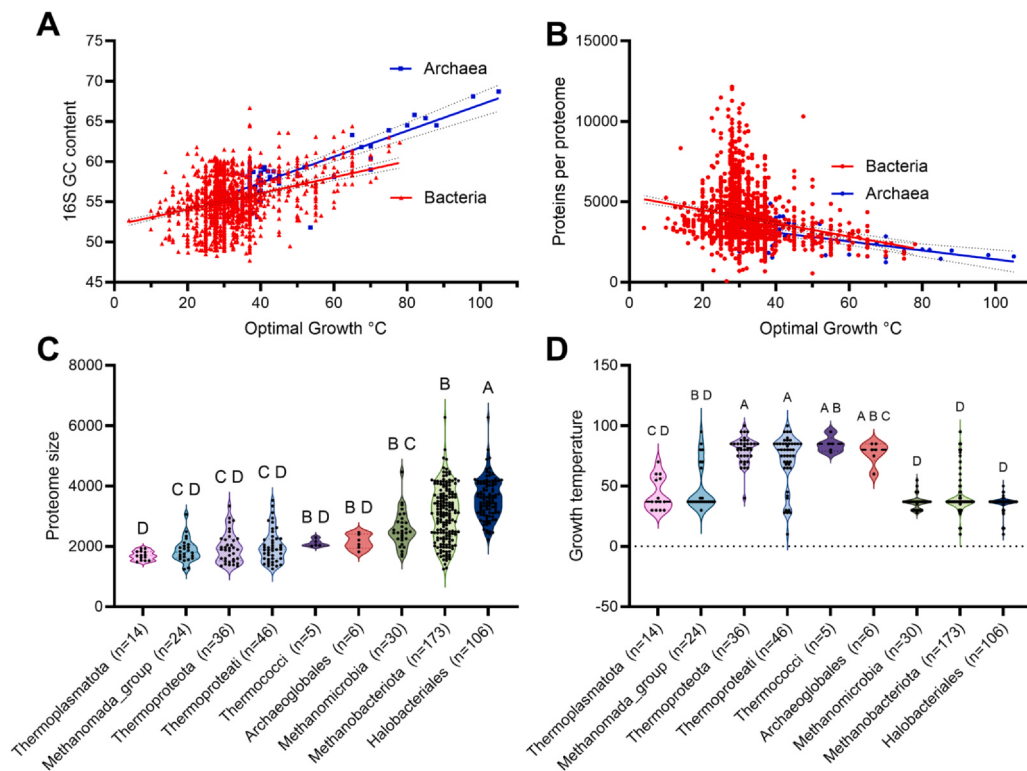


Fig. 1 Scatter plot of proteins per proteome and 16S GC content of selected organisms from the Tempura and Uniprot databases and comparison of different microbial groups based on their mean proteome size and mean growth temperature. **(A)** Bacteria and archaea show a positive correlation between GC content and optimal growth temperatures. Archaea (blue) have a steeper trend line than bacteria (red), suggesting a stronger relationship between temperature and GC content in archaeal species (equation for archaea and bacteria: $Y = 0.1617 \cdot X + 50.90$, $R^2 0.6901$, P value < 0.0001 and $Y = 0.1064 \cdot X + 51.77$, $R^2 0.184$, P value < 0.0001). **(B)** Mean values for proteins of archaeal proteomes (blue) are 2920 and for bacterial (red) 4026. Fitted trend lines show a negative correlation between optimal growth temperatures and proteomes size (Equation for Archaea and Bacteria: $Y = -27.87 \cdot X + 4209$, $R^2 0.07697$, P value < 0.0001 and $Y = -34.92 \cdot X + 5047$, $R^2 0.2946$, P value < 0.0001). Dotted lines: 95% CI. Trend lines are simple linear regression between the x and y axis calculated with PRISM. **(C)** Proteome sizes vary significantly across groups, with some, like Halobacteriales, having larger ranges. **(D)** Differences in growth temperatures, where thermophilic groups (eg. Thermoprotei, Thermococci) have higher mean growth temperatures than others. Black dots represent individual values. A Kruskal-Wallis test followed by a post-hoc (Dunn's multiple comparison test) was used to test for statistically differences between groups. Groups with different letters are statistically distinct, while groups with the same letter are not statistically different from each other.

groups, which show distinct trends in proteome size and growth temperature. Some groups, like Halobacteriales, display a broader range of proteome sizes while preferring lower temperatures. Others, such as Thermococci, have smaller proteomes and grow at higher temperatures (Figure 1C and D).

To go further than just looking at these overall correlations and check out these patterns in terms of chaperone system composition, I then analyzed how the presence or absence of the DnaK-DnaJ-GrpE (KJE) system changes based on proteome size and optimal growth temperature (Figure 2A and B).

As shown in Figure 2, archaea encoding for the KJE system tend to have larger proteomes and are mainly associated with mesophilic to moderately thermophilic growth temperatures, while organisms lacking KJE are more frequently found among hyperthermophiles with smaller proteomes.

Co-occurrence patterns between chaperones and selected genes in archaea

A detailed survey of the presence of the different chaperones and the proteasome subunits (psmA and psmB, since the proteasome is required for cell growth and stress responses³⁰), is shown in Figure 3. The vast majority of all the archaea analyzed here (~98%-100%) possess one or more thermosomes (chaperonins subunits), small Heat Shock Proteins (sHSPs), prefoldins (here called TPS system)¹ and 20S proteasomes,³⁰ attesting for their crucial relevance at all temperatures. As a matter of fact, prefoldins, chaperonins, and sHSPs are considered to be among the most conserved in archaea and constitute the oldest chaperone system that, according to several bioinformatics analyses, was already part of the Last Universal Common Ancestor (LUCA).² Using these protein families also enabled us to check

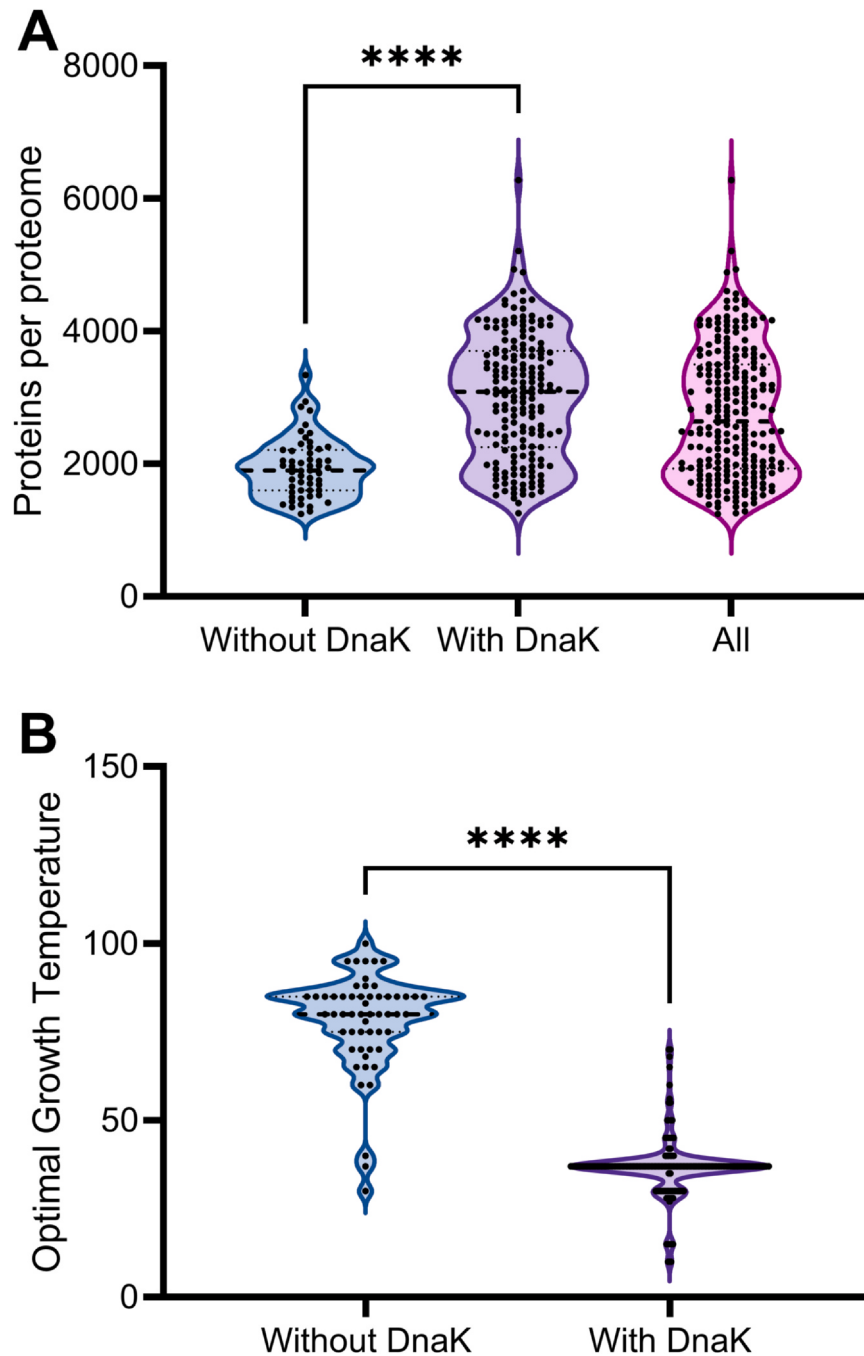


Fig. 2 Mean archaeal proteome size and optimal growth temperature depending on the presence or absence of DnaK. **(A)** Mean archaeal proteome size of all archaea in the subset is 2763 proteins. Without DnaK, 1936 and with DnaK 3023. The y-axis represents the number of proteins per proteome, ranging from 0 to 8000. **(B)** Mean for organisms with DnaK is 36.75 °C and for those without 78.18 °C. Black dots representing individual proteomes. Statistical comparisons (Student *t*-test) between datasets are denoted by asterisks, ****: *P* value < 0.0001.

the databases and the robustness of the analysis, given that thermosomes are thought to be ubiquitously present and highly expressed.³¹

One chaperone system showed instead a striking correlation with optimal growth temperature: the KJE system, which is mostly associated with mesophilic species (Figure 3; Supplementary Figure 4). In

particular, attesting for the stringent relation between DnaK and DnaJ, they are always present (or absent) together. Instead, GrpE was absent in a few organisms that have DnaJ and DnaK but was never present in their absence, a feature that has already been observed in the past in several parasitic organisms.³² Nonetheless, the physiological interest in losing GrpE, which enhances

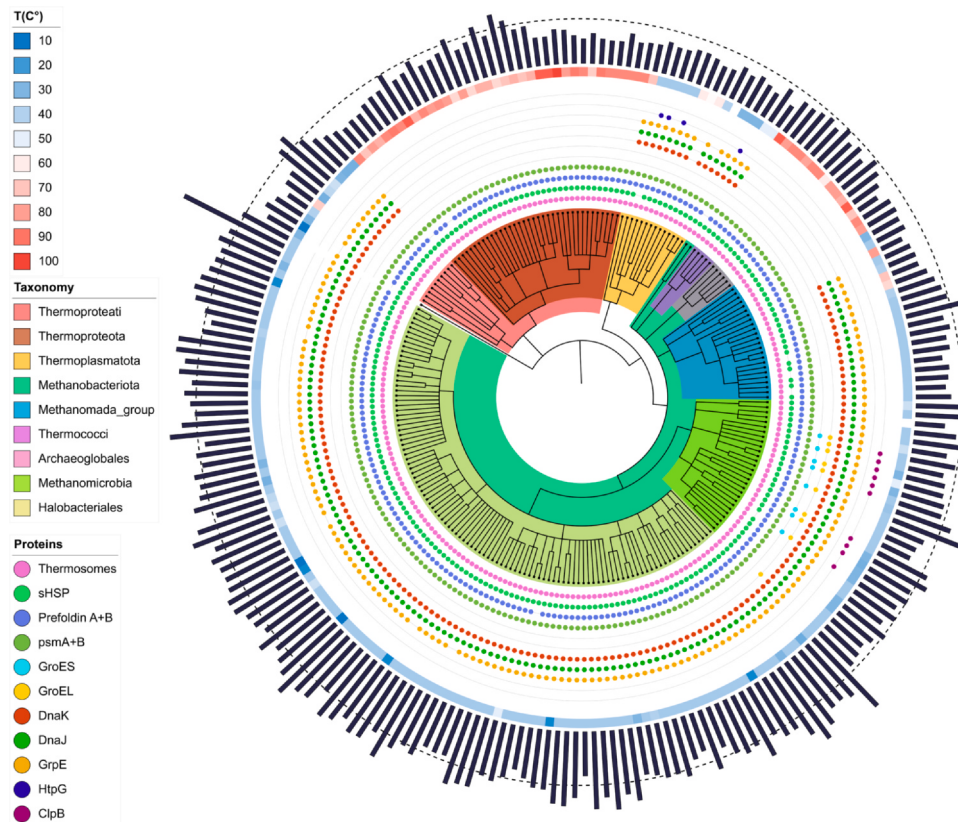


Fig. 3 Taxonomic tree of the selected 234 archaeal organisms. Different taxonomy has been represented on the inner tree; Thermoproteati (light red), Thermoproteota (maroon), Thermoplasmatota (orange), Methanobacteriota (sea green), Methanomada group (sea blue), Thermococci (purple), Archaeoglobales (pink, grey), Methanomicrobia (lime), and Halobacteriales (pear). On the outside of the tree are different classes of chaperones and proteases represented by the presence (colored circle) or absence of the protein; Thermosomes (pink), sHSP (sea green), Prefoldin A/B (blue), Proteasome subunit A/B (green), GroES (light blue), GroEL (yellow), DnaK (red), DnaJ (green), GrpE (yellow), HtpG (deep blue), and ClpB (purple). Then an heatmap of the optimal growth temperature of the organisms between 10 (blue) to 100 °C (red). On the outer side are proteome size of each organism (black bars) with the mean size (dotted line). This tree has been constructed using the NCBI taxonomy²⁹ of selected organisms and visualized in iTOL.²⁵

the efficiency of the KJ system, would be evolutionarily difficult to explain.^{1,32} It is nevertheless possible that the transfer of certain operons possessing only DnaK or DnaJ by HGT may have been possible, as a review of bacterial genomes³³ showed that tripartite JEK or bipartite KE/KJ/JE operons compositions exist. It is also notable that the proteomes of archaea and bacteria¹⁸ that possess DnaK, and the KJE system in general, are larger (Figure 2A and 3).

Differential distribution of chaperone systems in archaeal lineages

The Methanomicrobia class, and more specifically the Methanosarcinales order, comprises mesophilic to moderately thermophilic archaea, including *Methanosarcina acetivorans*³⁴ (lime clade in Figure 3 and Supplementary Information S5-S6 and S9). Several members of this group, such as *Methanosarcina mazei*,³⁵

possess not only thermosomes but also group I chaperonins (GroELs),¹ which have different substrate specificity and that have been likely acquired from bacteria by HGT.

In addition to GroELs, several members of this group of archaea seems to have also acquired the AAA+ chaperone of the Hsp100 family (ClpB) (purple dots on Figure 3, supplementary information S6, S9).³⁶ This suggests that members of this mesophilic group have undergone a fairly large number of HGTs for the chaperone network. But why is ClpB, together with GroELs, only present in Methanomicrobia? Is it the specificity of the environmental niche³⁷ in which these archaea live that has enabled successive HGTs with bacteria sharing it? Studying the prokaryotic communities that cohabit in this type of niche, as well as finding which other genes might have been transferred, could provide an answer. Another observation that deserves to be explored in more detail is the presence in

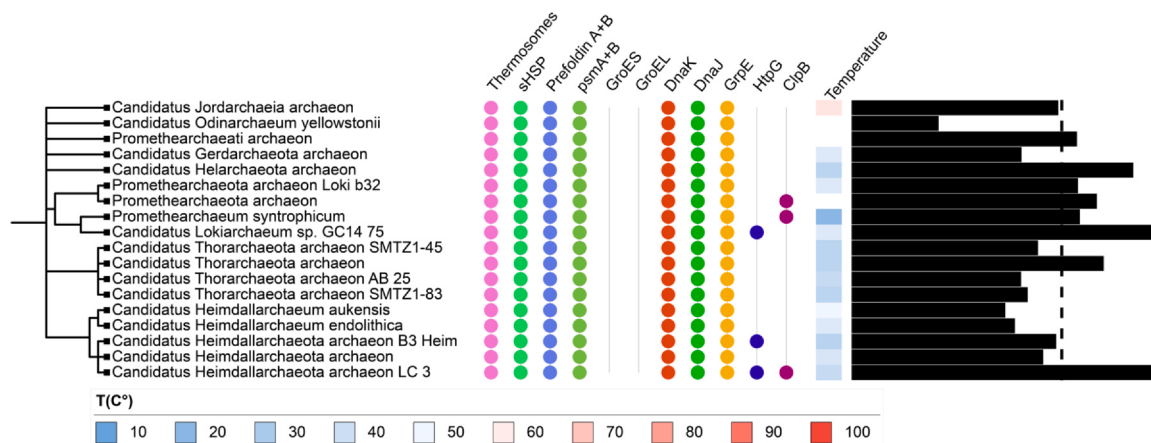


Fig. 4 Taxonomic tree of the selected 18 *Promethearchaeati* (ASGARD) organisms. On the outside of the tree are different classes of chaperones and proteases represented by the presence (colored circle) or absence of the protein; Thermosomes (pink), sHSP (sea green), Prefoldin A/B (blue), Proteasome subunit A/B (green), GroES (light blue), GroEL (yellow), DnaK (red), DnaJ (green), GrpE (yellow), HtpG (deep blue), and ClpB (purple). Then an heatmap of the optimal growth temperature of the organisms between 10 (blue) to 100 °C (red). On the outer side are the proteome size of each organism (black bars) with the mean size (dotted line, 3626). This tree has been constructed using the NCBI taxonomy²⁹ of selected organisms and visualized in iTOL.²⁵

some *Thermoplasmatota* organisms of the chaperone Hsp90 (HtpG, deep blue dots on Figure 3), which is generally accepted to be absent from most archaea.³⁸

Analysis of the chaperone network in the *Promethearchaeati* (ASGARD) family

In recent years, several new members of the archaea domain have been discovered, among which the ASGARD family³⁹ which was recently renamed and classified in the *Promethearchaeati* kingdom.⁴⁰ This group is currently thought to be the closest to the archaea that fused with bacteria during endosymbiosis to become the Last Eukaryotic Common Ancestor, LECA.⁴¹ Although most of the available proteomes of these new archaea are not considered reference proteomes and thus were not considered in the analysis above, I selected some of the most complete and annotated ones in order to characterize their chaperome, optimal growth temperatures, and the relative size of their proteomes. Most of these taxa have not been validly described based on their nomenclatural status,⁴² *Prometheoarchaeum syntrophicum* MK-D1 being the only reference organism. I reprise the same analysis here on this more specific dataset, showing that compared to other archaea, *Promethearchaeati* tend to have larger proteomes (Supplementary Figure 4, Supplementary Information S3+S8), in keeping with their mostly mesophilic or moderately thermophilic nature.^{43,44}

I then compared these data with the chaperone composition, finding that they all possess the KJE system (Figure 4), and a few have Hsp90 (HtpG) or Hsp100 (ClpB). The presence of a more robust chaperone repertoire in these mesophilic archaea, which

are considered to be the closest relatives of eukaryotes,⁴³ is not surprising, because the proteome and eukaryotic-related complexity found in *Promethearchaeati*⁴⁵ often go hand in hand with the expansion of the chaperone system.¹ Interestingly, in these archaea it appears that HGT from bacteria did not occur for the GroELS system (Figure 4), but due to the small set of genomes analyzed here, it is yet unwarranted to draw general statements.

Phylogenetic analysis of Hsp70 in selected organisms

I have then selected one Hsp70 sequence from each organism displaying one or more Hsp70 genes (ie, 178 archaea from the first dataset and the 18 *Promethearchaeati*) and they were aligned to verify their sequence homology (see Supplementary Information S7). The presence of one or more copies of DnaK or Hsp70 in mesophilic or light thermophilic organisms, as well as in all *Promethearchaeati*, is an interesting marker for the joint hypothesis of the maintenance of proteins that possess otherwise deleterious mutations, as well as proteome expansion.¹⁸ The selected Hsp70 sequences were then used to generate a phylogenetic tree (Figure 5). To compare the sequences present in these archaea with bacteria, a selection of Hsp70 from *Pseudomonadota* (gram-negative) and *Bacillota* (mostly gram-positive) were added (see method and Supplementary Information S11), as well as a recently discovered archaeon (*Candidatus Sukunaarchaeum mirabile*) which has undergone extreme genome reduction but retained the KJE system.⁴⁶ The results indicate that DnaK from archaea tend to cluster with those from gram-positive bacteria, having the same indels as previously showed,⁴⁷ and that archaea with more than one Hsp70 all possess a bona fide DnaK besides

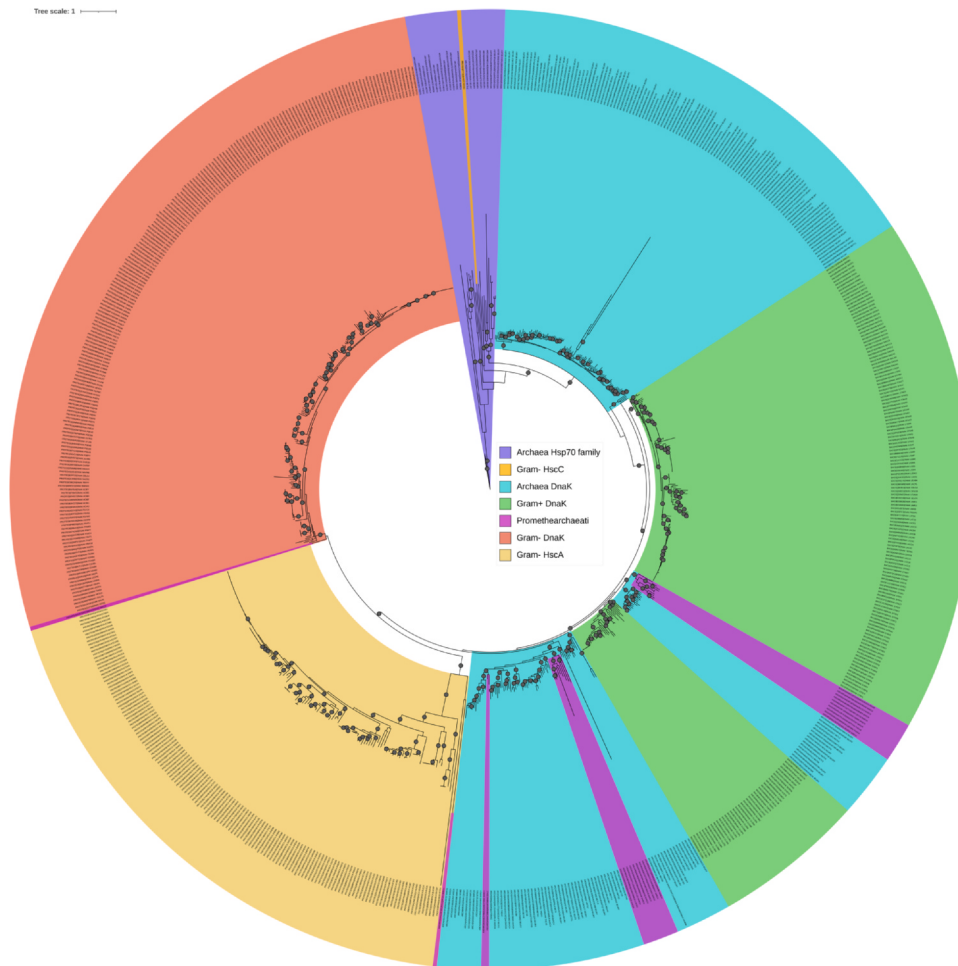


Fig. 5 Maximum likelihood phylogeny of archaeal and bacterial Hsp70 (DnaK/HscA/other) proteins. The tree was reconstructed with IQ-TREE2 using the LG+F+I+G4 substitution model, as selected by ModelFinder. Branch support was assessed with 1000 ultrafast bootstrap replicates and -aLRT tests. Archaeal sequences are shown in light blue for the non Promethearchaeati (ASGARD) and purple for the Promethearchaeati, while bacterial sequences are shown in green for the Bacilliota (Gram+) and in red for the Pseudomonadota (Gram-) DnaK, yellow for HscA and orange for the *E. coli* HscC. Bootstrap support values of at least 95% are represented by black dots. The scale bar indicates the number of amino acid substitutions per site.

other Hsp70-like (Figure 5). Interestingly, all archaea that have more than one copy of Hsp70 are Halobacterium, Thermoplasmatota or Promethearchaeati. Halobacterium⁴⁸ live in high-salinity environment and Thermoplasmatota⁴⁹ are acidophiles and light thermophilic, growing in pH < 2. This could explain the diversification of their Hsp70s to adapt to these environments,⁵⁰ as described in acidophiles bacteria⁵¹ or archaea.⁵² Promethearchaeati encode KJE and several Hsp70 genes and have relatively large proteomes, although these data remain preliminary given current genome completeness and curation.

Discussion

The present analysis reveals that archaeal lineages encoding the KJE chaperone system tend to have larger

proteomes and occupy mesophilic⁵³ niches. This consistent trend across all independent clades suggests that maintenance of the KJE system is linked to proteome expansion and folding requirements, rather than temperature alone. Therefore, the KJE complex may act as a molecular capacitor, buffering folding stress. Conversely, the recurrent absence of KJE in hyperthermophilic lineages likely reflects an adaptation to intrinsically stable proteomes and a reduced need for ATP-dependent folding assistance. Moreover in prokaryotes, ATP production is limited by the available cell surface⁵⁴ and the cost of repairing or degrading proteins can also influence what will be the choice in the cell.⁵⁵ The energy demands of a cell are primarily dictated by their volume regardless of complexity,⁵⁶ but complexity drives complexity, and protein synthesis accounts for 70%-80% of the ATP budget of prokaryotes.⁵⁷ Finally it

has been hypothesized that costly in energy PQC pathways can systematically promote the evolution of protein networks.⁵⁸ These results indicate that the interaction between proteome complexity, environmental temperature, and the distribution of folding mechanisms such as KJE played a role in shaping the evolutionary trajectories of archaea.⁵³

Possible interpretations

One explanation is that chaperones, here the KJE system, contribute directly to mutational buffering of their clients,^{59,60} thereby permitting the retention of more proteins and supporting proteome expansion in moderate environments.^{15,53,61} Experimental work in bacteria has shown that Hsp70 clients evolve faster and accumulate destabilizing mutations that are tolerated thanks to chaperone assistance.^{19,21,62}

Another possibility is that KJE distribution reflects HGT dynamics. Archaeal Hsp70 operons are thought to have been acquired from bacteria,^{15,61} and because most donor bacteria are mesophiles, archaeal mesophiles would be more likely to retain these genes, while hyperthermophiles would not.

A third explanation is that hyperthermophilic environments inherently constrain proteome expansion, as proteins in such conditions must remain stable and highly optimized, and organisms utilize multiple advantages of IDPs, leaving little tolerance for the accumulation of destabilizing variants^{8,4,63} and making some chaperones, like the KJE system, unnecessary. In this view, the correlation between KJE and proteome size would be indirect.

Implications for archaeal evolution

The case of *Promethearchaeati* is especially noteworthy. These lineages, which have been proposed as close relatives of the host that gave rise to eukaryotes,^{39,64} possess expanded proteomes and encode the full KJE machinery, together with the occasional acquisition of Hsp90 and Hsp100. This combination supports the idea that expanded chaperone networks may have facilitated increased proteomic complexity during early archaeal diversification and possibly prior to eukaryogenesis.¹ At the same time, other lineages such as Halobacteriales or Methanomicrobia illustrate how possible repeated HGTs can reshape the chaperone repertoire,^{35,65,66} suggesting that ecological context and microbial community composition also played an important role in the formation of these systems. Thermoplasmata constitutes another interesting group, with light thermophilic and, above all, extreme acidophilic species.

Despite their reduced genomes (Figure 1C), their acidophilic nature and the ecological niche in which they evolve seem to have favored the acquisition of the KJE system,^{49,51} despite the light thermophilic environment.

Alternative evolutionary scenarios

These results can be discussed in the broader framework of archaeal and early cellular evolution. Several studies have argued that LUCA was a thermophilic or even hyperthermophilic organism^{2,41}, which would imply that KJE was later acquired as archaea adapted to cooler environments. Other analyses, including phylogenetic reconstructions of reverse gyrase, suggest instead that LUCA was mesophilic.⁶⁷ Polygenic models of adaptation propose that temperature shifts result from the cumulative effect of horizontally acquired genes rather than single determinants.^{61,68,69} Another type of adaptation to a mesophilic environment involves genes playing a role in the regulation of membrane fluidity and membrane composition⁷⁰ and transport across membranes, as a large proportion of HGT from bacteria belong to two of these metabolic categories.^{61,66} Differentiating between these scenarios requires gene-level phylogenies of the extended PQC network in the archaeal and bacterial domains, coupled with ecological and environmental reconstructions.

Aside from whether LUCA was mesophilic or thermophilic, other evolutionary scenarios can be proposed specifically for the KJE system itself. One could think of a progressive recruitment, in which ancestral archaeal lineages initially relied solely on group II chaperonins (thermosomes) for protein folding, then incorporated bacterial-type Hsp70s through HGT. In this view, the KJE system would have been selectively conserved in lineages experiencing increased proteomic diversity or more variable environments. A contrasting co-evolutionary model could suggest that KJE did not simply complement existing systems but co-evolved with the expansion of proteomes, forming an integrated network that increased folding capacity as complexity increased. Previous work on bacterial systems showed correlations between DnaK duplication and proteome complexity.¹⁸

Limitations

This work is subject to several limitations. The dataset is unevenly sampled, with many reference proteomes reported at 37 °C, which may bias correlations despite robustness checks. Presence/absence hits may be affected by annotation or assembly errors. While these results are consistent with a relation between the

expansion of chaperone systems and the complexity of the proteome, it seems nonetheless important to point out that the KJE system does not act alone as an evolutionary capacitor and that other members of the PQC (ie, GroELS, Hsp90) could be implicated in these adaptations.^{3,22,59,71,72} Finally, targeted experiments (eg, reconstruction of ancestral proteins and associated growth tests, knock-in/knock-out experiments in model archaea) and more detailed phylogenetic analyses to test different scenarios more directly would be necessary.

Future directions

Further work will be needed to evaluate the mechanisms underlying these associations. Experimental approaches to determine whether archaeal proteins with and without KJE differ in terms of folding efficiency and mutational tolerance would allow direct testing of the capacitor hypothesis in similar experimental system as in Aguilar-Rodríguez et al.¹⁹ Comparative metagenomic studies of archaeal habitats could help identify the ecological conditions under which Hsp70 transfers occur.

Conclusion

In summary, I have analyzed over 250 archaeal proteomes to make the broadest possible comparison using quality proteomes that are as complete as possible so as to limit the possibility that the absence of proteins is due to poor sequencing or lack of completion. The KJE system in archaea is associated with mesophilic growth temperatures and larger proteomes. However, it is unknown whether KJE specifically played a role in expanding archaeal proteomes in mesophilic archaea or if it simply emerged through repeated horizontal transfers from cohabiting organisms. Future phylogenetic and experimental studies will be needed to clarify the evolutionary role of KJE.

Material and methods

Prokaryotic database

The Tempura²⁷ and BacDive²⁸ databases were used to retrieve information on optimal growth temperatures for the various organisms, as well as information on the GC content of 16S ribosomal RNA. For missing information and values, publications on the various organisms were retrieved to complete the database. In the

case of Promethearchaeati, several papers have been used to infer temperatures of most of them.^{43,44}

Selection of organisms and local datasets creation

The archaea selected to create the local database were retrieved on UNIPROT⁷³ from the reference proteomes and filtered against the BUSCO⁷⁴ Score with less than 10% of the proteome missing. For the first dataset, around 2000 archaeal and bacterial organisms were selected based on the Tempura values. Then, 234 archaea were selected to create the second dataset, to which several organisms from the Promethearchaeati (AS-GARD)⁴⁰ clade were added in a separate third dataset. The Promethearchaeati proteomes are not reference proteomes but are those of the archaea closest to the common ancestor of all eukaryotes, LUCA.⁴¹ The FASTA files from each of these organisms have been fetched and assembled with BLAST+⁷⁵ to be used on a local database against selected sequences.

Establishing presence or absence of proteins in the selected datasets

The two datasets containing exclusively archaea (the 234 archaea and the 18 Promethearchaeati) were then blasted against a selection of proteins from bacteria (*Escherichia coli* K12, UP000000625) and archaea (*Thermoplasma acidophilum*, UP000001024; *Methanothermobacter thermautotrophicus*, UP000005223; *Thermoplasma acidophilum*, UP000001024 and *Methanocaldococcus jannaschii* UP000000805). Each selected protein from each organism was blasted against the two local databases with an E_value of 1e-5 for most of the proteins (1e-2 for sHSP) and then manually curated. Based on the result of the BLAST searches and if the presence of hits in each organism were not present, another BLAST was performed with another closely related protein and/or with a lowered E_value up to 1. Subsequent results were analyzed by hand, and the presence or absence of each protein was reported. Finally, the results were compared with the InterPro⁷⁶ database to check that the proteins obtained and the domains present were correctly annotated. All proteins used can be found in supplementary information S4. To ensure that some of the proteins were correct, an alignment was carried out using MUSCLE⁷⁷ to check the MSA.

Phylogenetic analysis of Hsp70 homologs

Phylogenetic analyses of Hsp70 (DnaK, HscA, and others) homologs were performed using IQ-TREE 2.4.0.⁷⁸ Protein sequences were aligned with the MUSCLE3⁷⁷ algorithm, and poorly aligned regions were

trimmed with trimAl (v1.4) using the -gt (gapthreshold) and -cons (minimum percentage of positions in the original alignment to conserve) methods.⁷⁹ The best-fitting substitution model was determined by ModelFinder⁸⁰ as LG+F+I+G4, which was subsequently used for maximum likelihood (ML) tree reconstruction. Branch support was estimated using 1000 ultrafast bootstrap replicates²⁹ and approximate likelihood-ratio test-aLRT tests.⁸¹ The resulting tree was visualized and annotated in iTOL.⁸²

Figure creation

The figures were all made using GraphPad PRISM version 10 and iTOL⁸² for the phylogenetic and taxonomic trees.

Statistical analysis

All statistical analyses were performed in R (version ≥ 4.5) and GraphPad Prism version 10. Continuous relationships (eg, proteome size versus optimal growth temperature) were tested using linear regression (GLM) or generalized least squares; for each regression slope, R^2 , 95% confidence intervals, and P value were reported. Corrected analyses using generalized least squares with the ML model and other statistical methods implemented via the nlme/glmmTMB/DHARMA/lmtest packages in R were performed to compare the results. Comparison between models have been plotted with ggplot2 and modelsummary packages in R. For group comparisons (eg, mean proteome size between taxonomic groups), Kruskal-Wallis tests followed by Dunn's post-hoc tests, where used. Presence/absence analyses (eg, presence of KJE relative to temperature class) were tested using Student's t-test or one-way ANOVA in PRISM10.

Funding and support M.E.R. acknowledges SNSF funding under grant IC0010-227688 (project number 10000663).

Data availability All data are available in the main text or in the supplementary data.

Declaration of Competing Interest The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments The author would like to acknowledge Pierre Goloubinoff for many stimulating discussions that enabled me to work on several very interesting topics, such as this one. Paolo De Los Rios for reading the manuscript, amending it, thoughtful comments and suggestions on how to improve it and letting me work on this. Lisa Gennai for reading the manuscript and giving

me some feedback, and all my colleagues in the LBS lab. The author would like to acknowledge the 2 anonymous reviewers and the editor who helped improve the manuscript with their suggestions and comments.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.cstres.2026.100145](https://doi.org/10.1016/j.cstres.2026.100145).

References

1. Rebeaud ME, Mallik S, Goloubinoff P, Tawfik DS. On the evolution of chaperones and cochaperones and the expansion of proteomes across the Tree of Life. *Proc Natl Acad Sci U S A*. 2021;118.
2. Weiss MC, Sousa FL, Mrnjavac N, et al. The physiology and habitat of the last universal common ancestor. *Nat Microbiol*. 2016;1:16116.
3. Drake JW. Avoiding dangerous missense: thermophiles display especially low mutation rates. *PLoS Genetics*. 2009;5:e1000520.
4. Singer GA, Hickey DA. Thermophilic prokaryotes have characteristic patterns of codon usage, amino acid composition and nucleotide content. *Gene*. 2003;317:39–47.
5. Cherry JL. Highly expressed and slowly evolving proteins share compositional properties with thermophilic proteins. *Mol Biol Evol*. 2009;27:735–741.
6. Mackwan RR, Carver GT, Kissling GE, Drake JW, Grogan DW. The rate and character of spontaneous mutation in *Thermus thermophilus*. *Genetics*. 2008;180:17–25.
7. Grogan DW, Carver GT, Drake JW. Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. *Proc Natl Acad Sci U S A*. 2001;98:7928–7933.
8. Groussin M, Gouy M. Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in archaea. *Mol Biol Evol*. 2011;28:2661–2674.
9. Rocha EP, Danchin A. An analysis of determinants of amino acids substitution rates in bacterial proteins. *Mol Biol Evol*. 2004;21:108–116.
10. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. "Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A*. 2005;102:14338–14343.
11. Leuenberger P, Ganscha S, Kahraman A, et al. Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science*. 2017;355:eaai7825.
12. Fuchsman CA, Collins RE, Rocap G, Brazelton WJ. Effect of the environment on horizontal gene transfer between bacteria and archaea. *PeerJ*. 2017;5:e3865.
13. Zhu Q, Mai U, Pfeiffer W, et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and archaea. *Nat Commun*. 2019;10:5477.
14. Creevey CJ, Doerks T, Fitzpatrick DA, Raes J, Bork P. Universally distributed single-copy genes indicate a constant rate of horizontal transfer. *PLoS One*. 2011;6:e22099.
15. Petitjean C, Moreira D, López-García P, Brochier-Armanet C. Horizontal gene transfer of a chloroplast DnaJ-Fer protein to Thaumarchaeota and the evolutionary history of the DnaK chaperone system in archaea. *BMC Evol Biol*. 2012;12:226.
16. Philippe H, Budin K, Moreira D. Horizontal transfers confuse the prokaryotic phylogeny based on the HSP70 protein family. *Mol Microbiol*. 1999;31:1007–1010.

17. Calloni G, Chen T, Schermann SM, et al. DnaK functions as a central hub in the *E. coli* chaperone network. *Cell Reports*. 2012;1:251–264.
18. Pan Z, Zhuo L, Wan T-Y, Chen R-Y, Li Y-Z. DnaK duplication and specialization in bacteria correlates with increased proteome complexity. *mSystems*. 2024;9:e01154-01123.
19. Aguilar-Rodríguez J, Sabater-Muñoz B, Montagud-Martínez R, et al. The molecular chaperone DnaK is a source of mutational robustness. *Genome Biol Evol*. 2016;8:2979–2991.
20. Diaz Arenas C, Alvarez M, Wilson RH, Shakhnovich EI, Ogbunugafor CB. Protein quality control is a master modulator of molecular evolution in bacteria. *Genome Biol Evol*. 2025;17:1–15.
21. Kadibalban AS, Bogumil D, Landan G, Dagan T. DnaK-dependent accelerated evolutionary rate in prokaryotes. *Genome Biol Evol*. 2016;8:1590–1599.
22. Rutherford SL. Between genotype and phenotype: protein chaperones and evolvability. *Nat Rev Genet*. 2003;4:263–274.
23. Mayer MP. The Hsp70-chaperone machines in bacteria. *Front Mole Biosci*. 2021;8:694012.
24. Kimura H, Sugihara M, Kato K, Hanada S. Selective phylogenetic analysis targeted at 16S rRNA genes of thermophiles and hyperthermophiles in deep-subsurface geothermal environments. *Appl Environ Microbiol*. 2006;72:21–27.
25. Hu E-Z, Lan X-R, Liu Z-L, Gao J, Niu D-K. A positive correlation between GC content and growth temperature in prokaryotes. *BMC Genom*. 2022;23:110.
26. Vieille C, Zeikus GJ. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev*. 2001;65:1–43.
27. Sato Y, Okano K, Kimura H, Honda K. TEMPURA: database of growth TEMPeratures of usual and RAre prokaryotes. *Microbes Environ*. 2020;35.
28. Schober I, Koblit J, Sardà Carbasse J, et al. BacDive in 2025: the core database for prokaryotic strain data. *Nucleic Acids Res*. 2024;53:D748–D756.
29. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. "UFBoot2: improving the ultrafast Bootstrap approximation. *Mole Biol Evol*. 2017;35:518–522.
30. Zhou G, Kowalczyk D, Humbard MA, Rohatgi S, Maupin-Furlow JA. Proteasomal components required for cell growth and stress responses in the haloarchaeon *Haloferax volcanii*. *J Bacteriol*. 2008;190:8096–8105.
31. Karlin S, Mrázek J, Ma J, Brocchieri L. Predicted highly expressed genes in archaeal genomes. *Proc Natl Acad Sci U S A*. 2005;102:7303–7308.
32. Warnecke T. Loss of the DnaK-DnaJ-GrpE chaperone system among the aquificales. *Mole Biol Evol*. 2012;29:3485–3495.
33. Barriot R, Latour J, Castanié-Cornet MP, Fichant G, Genevaux P. J-domain proteins in bacteria and their viruses. *J Mol Biol*. 2020;432:3771–3789.
34. Maeder DL, Macario AJL, de Macario EC. Novel chaperonins in a prokaryote. *J Mole Evol*. 2005;60:409–416.
35. Hirtreiter AM, Calloni G, Forner F, et al. Differential substrate specificity of group I and group II chaperonins in the archaeon *Methanosarcina mazei*. *Mol Microbiol*. 2009;74:1152–1168.
36. Shih C-J, Lai M-C. Analysis of the AAA+ chaperone clpB gene and stress-response expression in the halophilic methanogenic archaeon *Methanohalophilus portucalensis*. *Microbiology*. 2007;153:2572–2583.
37. Baquero F, Coque TM, Galán JC, Martínez JL. The origin of niches and species in the bacterial world. *Front Microbiol*. 2021;12:657986.
38. Chen B, Zhong D, Monteiro A. Comparative genomics and evolution of the HSP90 family of genes across all kingdoms of organisms. *BMC Genom*. 2006;7:156.
39. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*. 2017;541:353–358.
40. Imachi H, Nobu MK, Kato S, et al. *Promethearchaeum syntrophicum* gen. nov., sp. nov., an anaerobic, obligately syntrophic archaeon, the first isolate of the lineage "Asgard" archaea, and proposal of the new archaeal phylum Promethearchaeota phyl. nov. and kingdom Promethearchaeati regn. nov. *Int J Systematic Evol Microbiol*. 2024;74.
41. Moody ERR, Álvarez-Carretero S, Mahendrarajah TA, et al. The nature of the last universal common ancestor and its impact on the early Earth system. *Nat Ecol Evol*. 2024;8:1654–1666.
42. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List of Prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int J Systematic Evol Microbiol*. 2020;70:5607–5612.
43. Eme L, Tamarit D, Caceres EF, et al. Inference and reconstruction of the heimdallarchaeal ancestry of eukaryotes. *Nature*. 2023;618:992–999.
44. Lu Z, Xia R, Zhang S, et al. Evolution of optimal growth temperature in Asgard archaea inferred from the temperature dependence of GDP binding to EF-1A. *Nat Commun*. 2024;15:515.
45. Köstlbacher S, van Hooff JJE, Panagiotou K, et al. Structure-based inference of eukaryotic complexity in Asgard archaea; 2024. bioRxiv: 2024.2007.2003.601958.
46. Harada R, Nishimura Y, Nomura M, et al. A cellular entity retaining only its replicative core: hidden archaeal lineage with an ultra-reduced genome; 2025. bioRxiv: 2025.2005.2002.651781.
47. Gupta RS. Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among Archaeobacteria, Eubacteria, and Eukaryotes. *Microbiol Mole Biol Rev*. 1998;62:1435–1491.
48. Gupta RS, Golding GB. Evolution of HSP70 gene and its implications regarding relationships between Archaeobacteria, Eubacteria, and Eukaryotes. *J Mole Evol*. 1993;37:573–582.
49. Gribaldo S, Lumia V, Creti R, Conway de Macario E, Sanangelantoni A, Cammarano P. Discontinuous occurrence of the hsp70 (dnaK) gene among archaea and sequence features of HSP70 suggest a novel outlook on phylogenies inferred from this protein. *J Bacteriol*. 1999;181:434–443.
50. Sugimoto S, Nakayama J, Fukuda D, et al. Effect of heterologous expression of molecular chaperone DnaK from *Tetragenococcus halophilus* on salinity adaptation of *Escherichia coli*. *J Biosci Bioeng*. 2003;96:129–133.
51. Izquierdo-Fiallo K, Muñoz-Villagrán C, Orellana O, Sjöberg R, Levicán G. Comparative genomics of the proteostasis network in extreme acidophiles. *PLoS One*. 2023;18:e0291164.
52. Matarredona L, Camacho M, Zafrilla B, Bonete M-J, Esclapez J. The role of stress proteins in haloarchaea and their adaptive response to environmental shifts. *Biomolecules*. 2020;10:1390.
53. Dunn CD. Some liked it hot: a hypothesis regarding establishment of the proto-mitochondrial endosymbiont during eukaryogenesis. *J Mole Evol*. 2017;85:99–106.
54. Raval PK, Garg SG, Gould SB. Endosymbiotic selective pressure at the origin of eukaryotic cell biology. *eLife*. 2022;11:e81033.
55. Fauvet B, Rebeaud ME, Tiwari S, De Los Rios P, Goloubinoff P. Repair or degrade: the thermodynamic dilemma of cellular protein quality-control. *Front Mol Biosci*. 2021;8:768888.
56. Muñoz-Gómez SA. The energetic costs of cellular complexity in evolution. *Trends Microbiol*. 2024;32:746–755.
57. Lane N. How energy flow shapes cell evolution. *Curr Biol*. 2020;30:R471–R476.

58. Pechmann S, Frydman J. Interplay between Chaperones and Protein disorder promotes the evolution of protein networks. *PLoS Comput Biol*. 2014;10:e1003674.
59. Rutherford SL, Lindquist S. Hsp90 as a capacitor for morphological evolution. *Nature*. 1998;396:336–342.
60. Tokuriki N, Tawfik DS. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature*. 2009;459:668–673.
61. López-García P, Zivanovic Y, Deschamps P, Moreira D. Bacterial gene import and mesophilic adaptation in archaea. *Nat Rev Microbiol*. 2015;13:447–456.
62. Pechmann S, Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol*. 2013;20:237–243.
63. Uversky VN. Protein intrinsic disorder and adaptation to extreme environments: resilience of chaos. *J Mole Biol*. 2025;169547.
64. Spang A, Saw JH, Jørgensen SL, et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*. 2015;521:173–179.
65. Nelson-Sathi S, Dagan T, Landan G, et al. Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc Natl Acad Sci U S A*. 2012;109:20537–20542.
66. Nelson-Sathi S, Sousa FL, Roettger M, et al. Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature*. 2015;517:77–80.
67. Catchpole RJ, Forterre P. The evolution of reverse gyrase suggests a nonhyperthermophilic last universal common ancestor. *Mol Biol Evol*. 2019;36:2737–2747.
68. Farrell AA, Nesbø CL, Zhaxybayeva O. Bacterial growth temperature as a horizontally acquired polygenic trait. *Genome Biol Evol*. 2024;17:1–20.
69. Pollo SM, Zhaxybayeva O, Nesbø CL. Insights into thermoadaptation and the evolution of mesophily from the bacterial phylum Thermotogae. *Canadian J Microbiol*. 2015;61:655–670.
70. Tourte M, Schaeffer P, Grossi V, Oger PM. Functionalized membrane domains: an ancestral feature of archaea? *Front Microbiol*. 2020;11:526.
71. Sabater-Muñoz B, Prats-Escriche M, Montagud-Martínez R, et al. Fitness trade-offs determine the role of the molecular chaperonin GroEL in buffering mutations. *Mol Biol Evol*. 2015;32:2681–2693.
72. Wyganowski KT, Kaltenbach M, Tokuriki N. GroEL/ES buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. *J Mol Biol*. 2013;425:3403–3414.
73. Consortium TU. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res*. 2024;53:D609–D617.
74. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–3212.
75. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinf*. 2009;10:421.
76. Blum M, Andreeva A, Florentino LC, et al. InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res*. 2025;53:D444–D456.
77. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–1797.
78. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mole Biol Evol*. 2020;37:1530–1534.
79. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–1973.
80. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14:587–589.
81. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–321.
82. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucl Acids Res*. 2024;52:W78–W82.