Comparative Genomic Insights into Chemoreceptor Diversity and Habitat Adaptation of Archaea

Guihong Cha, Yugeng Liu, Qing Yang, Liping Bai, Lei Cheng, Wei Fan

Abstract

Diverse archaea, including many unknown species and phylogenetically deeply rooted taxa, survive in extreme environments. They play crucial roles in the global carbon cycle and element fluxes in many terrestrial, marine, saline, host-associated, hot-spring, and oilfield environments. There is little knowledge of the diversity of chemoreceptors that are presumably involved in their habitat adaptation. Thus, we have explored this diversity through phylogenetic and comparative genomic analyses of complete archaeal genomes. The results show that chemoreceptors are significantly richer in archaea of mild environments than in those of extreme environments, that specific ligand-binding domains of the chemoreceptors are strongly associated with specific habitats, and that the number of chemoreceptors correlates with genome size. The results indicate that the successful adaptation of archaea to specific habitats has been associated with the acquisition and maintenance of chemoreceptors, which may be crucial for their survival in these environments.

Importance

Archaea are capable of sensing and responding to environmental changes by several signal transduction systems with different mechanisms. Much attention is paid to model organisms with complex signaling networks to understand their composition and function, but general principles regarding how an archaeal species organizes its chemoreceptor diversity and habitat adaptation are poorly understood. Here, we have explored this diversity through phylogenetic and comparative genomic analyses of complete archaeal genomes. Signaling sensing and adaptation processes are tightly related to the ligand-binding domain, and it is clear that evolution and natural selection in specialized niches under constant conditions have selected for smaller genome sizes. Taken together, our results extend the understanding of archaeal adaptations to different environments and emphasize the importance of ecological constraints in shaping their evolution.

Keywords

Archaea, habitat adaptation, chemoreceptors, ligand-binding domains

Archaea are considered one of the earliest life forms on earth (1). They occupy a key position in the tree of life and contribute a major fraction of microbial diversity (2). Currently, four major supergroups are recognized, designated Euryarchaeota, TACK (Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota), Asgard, and DPANN (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanohaloarchaeota, and Nanoarchaeota) (3). Here, the phylogeny tree distinguishes five archaeal superphyla: Euryarchaeota, Asgard, TACK, DPANN, and Candidatus Thermoplasmata (Fig. 1; see also Fig. S1 in the supplemental material). Each of the supergroups includes several potentially phylum-rank clades (3–5), and its members are present in diverse environments, including oceanic, terrestrial, hot-spring, and oilfield habitats (6, 7).
Many archaeal and bacterial species have chemotactic abilities to sense signals regarding environmental gradients and use them to direct their movement toward more favorable locations (8–10). This process involves various signal transduction pathways, which are also called chemoreceptor-based signaling cascades or chemosensory systems (11–15). Archaeal chemotaxis has received less attention than the well-explored bacterial chemotaxis systems, but chemoreceptors called methyl-accepting chemotaxis proteins (MCPs) sense environmental signals and then drive chemotactic responses in conjunction with other components in both bacteria and archaea (16, 17). It is also known that despite variants of the chemotaxis signaling system among bacterial and archaeal species (18), there is substantial conservation...
of both the overall mechanism and proteins involved (10). Previous studies on the model organism *Halobacterium salinarum* have shown that MCPs might contribute to the adaptation of archaea to their habitats. This model halophilic organism displays tactic behavior and has chemoreceptors that sense a multitude of environmental stimuli, such as various chemicals (e.g., acetate and some amino acids), oxygen, osmolarity, and light (8, 18, 19). They regulate the autophosphorylation activity of the histidine kinase CheA, through the coupling protein CheW (20, 21). After autophosphorylation, a phosphoryl group is transferred from CheA to the response regulator CheY (17, 22). Phosphorylated CheY interacts with the archaea-specific chemotaxis protein CheF (23), and the CheF/CheY complex can then change the direction of rotation of the archaellum filament through interaction with the ArlCDE complex of the euryarchaeal archaellum motor (21, 24–31). Important structural components of the filament (both "major" and "minor") include ArlA and ArlB archaellins (29, 32–36).

Chemoreceptors have two principal modules: input and output (37). The input module is usually composed of a single domain, although it has two or more domains in some chemoreceptors (38). The output module has a conserved structure, consisting of a long dimeric four-helix bundle composed of two symmetric antiparallel coils, which comprises the cytoplasmic signaling methyl-accepting (MA) chemotaxis-like domain (20, 39). In transmembrane chemoreceptors (the most common membrane topology class) input-output signaling is mediated by a HAMP linker (Histidine kinases, Adenylate cyclases, Methyl-accepting chemotaxis proteins, or Phosphatases) domain (20, 40). Chemoreceptors can recognize signals via their input module in several ways (8). A common mode of sensing is through the direct binding of effectors to the input ligand-binding domain (LBD) (20, 39, 41). Some input domains have cofactors, such as heme, which enables chemoreceptors to recognize oxygen (14). For example, Htr8 and Htr10 (HemAT) are required for attraction and phobic responses to oxygen of the archaeon *H. salinarum* (14). Although the structural composition of chemoreceptors is variable, the secondary structure of the MA domain, which interacts with the CheW adaptor protein and CheA histidine kinase, is highly conserved (41, 42), at least in taxa investigated to date. However, significant studies of the archaeal chemoreceptors remain confined to a few archaea: *H. salinarum* (19), *Methanocaldococcus jannaschii* (8), *Methanoseta harundinacea* (13), *Thermococcus kodakarensis* (11), and recently *Haloferax volcanii* (33). Thus, further information about chemoreceptor diversity, evolution, and roles in habitat adaptation is required.

In efforts to obtain such information, we applied comparative genomic analysis using well-annotated archaeal phylogeny to explore chemoreceptor diversity in strains isolated from the six types of habitats in which archaea are most prevalent: terrestrial, marine, saline, host associated, hot spring, and oilfield (Fig. 1). We found that genomic size variations among archaea are associated with habitat adaptation. We observed a strong correlation between the GC contents of the genomes and their phylogenetic positions, with strains located within clusters having similar GC contents. In addition, archaea of similar habitats and metabolic characteristics are generally clustered in the tree and separated by small numbers of branches, but they are relatively scattered in terms of overall evolutionary relationships. We also noticed no indication of any pattern in the genome size in the tree or correlation with evolutionary relationships. Thus, many characteristics of archaea do not appear to be strongly linked to these evolutionary relationships, raising questions about their diversity, which we sought to elucidate by studying the diversity of their chemoreceptors and dynamic changes in them. We focused on MCP diversity in the analysis to identify the major classes of archaeal MCP domains and to discover correlations between them and adaptation to various environments. That the archaea MCP repertoire evolved following patterns similar to that observed in bacteria (e.g., Campylobacterota) indicates that the more heterogenous conditions are in niches in which archaea swim, the more diverse the MCP repertoire (43, 44). Our findings indicate that the tight coupling of signal identification by the LBDs and adaptative functions of the chemotaxis systems contributed to substantial diversity in chemotaxis mechanisms among archaean groups.
RESULTS AND DISCUSSION

Diversity of chemoreceptors in archaea. Archaea are widespread and occur in all environments on earth, where they comprise substantial portions of the microbial biomass, and they exhibit great diversity in metabolism, morphology, and physiology (3). MCPs are the first components of chemotaxis system to sense outside signals and initiate chemotactic responses in both bacteria and archaea (8). Therefore, we focused on chemoreceptors to elucidate their genomic evolution and the mechanisms of their adaptation to different environments.

We searched all the chemoreceptors in the selected archaeal genomes and found that 150 of the 352 archaeal genomes have genes that encode MCPs and that 202 lack such genes (Fig. 1; see also Table S3 in the supplemental material). MCPs can be classified as either integral inner membrane proteins or soluble cytoplasmic proteins. Both cytoplasmic and membrane-bound MCPs reportedly have more heterogenous subcellular distributions than their bacterial counterparts, since archaeal arrays are frequently denser close to midcell positions (except by cell poles), whereas bacterial MCPs are more evenly distributed at the cell poles (11, 45). We found that 44% MCPs were cytoplasmic, and 56% were transmembrane (see Fig. S2A). We also observed that about 78% of chemoreceptors contain a membrane-proximal HAMP domain and that about 22% of the cytoplasmic chemoreceptors do not contain HAMP in all their MCPs (Fig. 2A; see also Fig. S2B). HAMP domains are predicted to transmit signals detected by extracytosolic LBDs to the MA domains in cytoplasmic chemoreceptors, and we found that many of these receptors contain a HAMP domain in archaea (see Table S6). In previous studies, MCPs have been divided into four classes based on membrane topology and LBD location (20). Here, chemoreceptors were divided into five classes, each further divided into several subgroups based on LBD locations (Fig. 2B; see also Fig. S2C). Specifically, numerous cytoplasmic class V chemoreceptors contain a putative N-terminal LBD, and some also contain a C-terminal LBD (Fig. 2B; see also Table S5). The broad structural diversity of chemoreceptors that we detected, including in LBD locations and types, strongly suggests that these chemoreceptors could monitor wide arrays of signals from the cytoplasm and the environment.

Our phylogenetic tree also indicates that neither the most recently evolved group of archaea ("Candidatus Thermoplasmatota") nor the ancient DPANN group has a chemotaxis system (Fig. 2C). Components of a chemotaxis system were mainly found in the Euryarchaeota, TACK, and Asgard superphyla (see Table S6). In addition, most of the archaeal phyla that have MCPs include some strains that lack MCPs. Moreover, the number of MCPs in members of the Stenosarchaea group varies from 1 to 27 (Fig. 2C; see also Table S3). The MCP repertoire is extremely limited in some species, but much broader in other species. In addition, the changes number or diversity of MCPs are not well conserved along branches of the evolutionary tree (Fig. 2C). We also show that all MCP-containing archaeal species have chemoreceptor type 44H as the dominant signal input (see Table S4 and Fig. S4). Unlike in Bacteria (11, 22, 43, 44), the chemotaxis system then evolved largely vertically in Archaea (11, 20, 46); these results also indicate that the chemotaxis system did not evolve solely in this manner in archaea and that low horizontal-gene-transfer (HGT) frequencies contributed (11).

Enriched chemoreceptors in a specific habitat. To further study the relationship between the number of chemoreceptors and the ecological habitat of the archaea, we collected information on the niche of each strain and divided them into ecological groups. The results show that MCPs are present in most saline and terrestrial archaea, but fewer are present in host-associated, marine, thermal-spring, and oilfield archaea (Fig. 3A). These results suggest that the presence of MCPs in saline and terrestrial archaea enables them to move toward more optimal substrates and thereby cope more readily with changes in habitat conditions. In terms of metabolic types, we also observed that the numbers of MCPs were significantly richer in acetotrophic archaea than in hydrogenotrophic, methylotrophic, H2-dependent methylotrophic, heterotrophic, and alkylotrophic archaea (Fig. 3B). As in bacteria, the data for Archaea identified here are similar to what has been shown for Bacteria, namely, that genome size and niches with changing conditions and multiple gradients correlate with larger genomes, more diverse
metabolisms, and more MCPs (43, 44). Collectively, the results mentioned above indicate that the number of MCPs encoded by archaeal genomes is closely related to their habitats. Variations in habitats account for the diversity of LBDs. Chemoreceptors have multiple protein domains, most of which are structurally conservative, but not the LBDs. More than 80% of LBDs in chemoreceptors with known LBDs are members of the protoglobin, cache-like, and PAS/PAC families, while the PilJ, HBM, SBP_bac, and 4HB families account for about 9% (Fig. 4A and B; see also Table S5). Cache-like families, which include several types, are the most prevalent LBDs in chemoreceptors of the archaea-like stenosarchae group (Fig. 4A and B; see also Table S5). PAS/PAC domains are the second most abundant family (Fig. 4B). These domains are found almost exclusively in cytoplasmic chemoreceptors (Fig. 4C), and many family members contain flavin adenine dinucleotide (FAD) or heme involved in oxygen and redox sensing processes (37, 47). Protoglobin domain is the third most abundant type of LBD (Fig. 4A and B). In the archaeon *H. salinarum* this domain contains heme and mediates oxygen sensing (14). In addition, some chemoreceptors contain an unknown N-terminal or C-terminal potential LBD that is least 80 amino acids in length (Fig. 4A and B).
We also observed that LBDs were significantly richer in heterotrophic strains than in other metabolic types (Fig. 4D) and were present in the archaea of most terrestrial and marine environments. Moreover, the LBD types of host-associated, saline, and oil-field archaea are also present in those of terrestrial and marine environments, but some LBD types occur only in MCPs associated with specific ecological habitats (Fig. 4E). These results indicate that their LBD diversity is closely related to variation in ecological niches. At the same time, these results also reveal that archaea inhabiting very specialized niches have less need for
chemotaxis since there may not be much competition for nutrients, etc. Although the chemotaxis system evolved largely vertically in archaea, these findings suggest that gene duplication and mutation are not enough to create such diversity and that HGT may also be involved in the MCP enrichment and diversification of LBD types in specific habitats. Protein fusion between different classes of MCPs may also contribute to the evolution of new types during gene duplication, as well as HGT processes (46, 48). Overall, these results indicate that successful archaean adaptation to specific habitats is associated with the acquisition and maintenance of MCPs, which might enhance fitness.

Correlation between MCP number and genome size. Chemoreceptors are crucial components of a chemotaxis system (8), but other proteins (such as CheA, CheW, and CheY) are also needed for signal transduction (24). Since CheY is a response regulator with a single REC domain, and proteins with a single REC domain are present in diverse signal transduction systems, study of the CheY protein of the archean *Halobacterium salinarum* showed that, as in bacteria, CheY is phosphorylated and the protein and its phosphorylation state are essential for the chemotactic behavior of archaea (22). Thus, we examined CheW, CheY, and CheA proteins in further analyses of archaean chemotaxis systems. We also examined the ArlA and ArlB components of their archaellin systems, which accept signals from the chemotaxis system and induce changes in organism behavior (23, 29).

We found that nearly half (43%) of the analyzed strains contain MCPs, that most of
these strains encode CheA, CheY, and CheW, and that only 4% of them do not encode CheA or CheW (Fig. 5A; see also Tables S6 and S7). In contrast, very few of the MCP-free strains encode CheA, and none encode CheW, which has no role when CheA is missing (Fig. 5A; see also Fig. S3). We also observed that few archaeal strains contain ArlA and ArlB (see Fig. S3). We postulated two hypotheses based on these results: (i) gene families of the chemotaxis and archaellin systems may have been lost or extended, and (ii) the chemotaxis system may play a key role in the regulation of not only motility patterns but also other systems, e.g., pili, in archaea.

Since the signal networks are complex and involve numerous signal transduction-related genes, and complex of signal networks maybe correlate with genome size. Thus, we also explored the potential relationship between the numbers of MCPs and genome size, which varies from 0.49 to 5.75 Mb (see Table S6). We found that MCP-free strains tend to have smaller genomes than do MCP-containing strains (Fig. 5B). MCP-containing strains have CheA, CheW, and other components of the chemotaxis system, as well as other systems, such as the archaellin system required for the archaellum filament. Next, we tested the possibility that the sizes of archaeal genomes have been affected by the extremity or heterogeneity of their habitats, which may clearly affect the diversity of many pathways involved in microbes’ metabolic capabilities, stresses, growth, and development. We found that strains growing in relatively extreme ecosystems had smaller genome sizes and fewer MCPs than in mild environments (Fig. 5B). For example, the number of archaeal MCPs appears to have been affected by habitat type and correlates with genome size (Fig. 1; see also Table S1). These results indicate that genome size correlates with the diversity and numbers of these systems and networks. In other words, the evolution of Archaea in specialized niches with more stable and constant conditions (e.g., host associated) has selected for Archaea with smaller genome sizes, streamlined and likely specialized metabolisms, and signal transduction systems.

**Conclusions.** We found differences in the numbers of MCPs encoded by the genomes of terrestrial, marine, saline, host-associated hot-spring, and oilfield archaea. More importantly, MCPs, such as the sensing substrate, were found to be involved in key processes, suggesting that they play major roles in the survival and adaptation of archaea in specific habitats. These findings show that signaling and adaptation processes are tightly coupled with the LBD and that show that evolution sensing and natural selection in specialized niches under constant conditions has selected for smaller genome sizes. Taken together, these results extend our understanding of archaeal adaptations to different environments and emphasize the importance of ecological constraints in shaping their evolution.
MATERIALS AND METHODS

Data sources. Representative species of various archaea phyla with completely sequenced genomes were obtained from the NCBI RefSeq or GenBank database, and their taxonomic ranks were based on the NCBI taxonomy database (43). One strain was selected to represent each species with multiple completely sequenced strains. To cover some species that reportedly play an important role in extreme niches and to ensure genome quality, several complete but nearly complete genomes were included for analyses (see Table S1). Information on all the selected genomes (352 in total) was downloaded from the NCBI genome database (https://www.ncbi.nlm.nih.gov/genome/microbes/).

Bioinformatics and data analysis software. Phylogenetic trees were constructed by FastTree 2.0 (49). All homologous protein searches were conducted using BLAST 2.12.0 against local genomes (50), and MAFFT was used, with default parameters, for multiple sequence alignments (MSAs) (51). The topologies and domain architectures of all proteins were predicted by SMART, CDvist, and HHpred (52–54). Phylogenetic trees were modified and processed using EvolView (https://www.evolgenius.info/evolview/#login). Data were analyzed and figures were generated by using the Rstudio tool with the ggplot2, ggrepel, tidyverse, hrbrthemes, and Viridis packages. The significance of differences among groups was assessed using a Student t test, implemented in Rstudio, and Venn networks were generated by software accessed through the Evven website (http://www.ehbio.com/test/venn/#!/).

Phylogenetic analysis. To construct a phylogenetic tree from the 352 archaeal genomes downloaded from the NCBI database (see Table S1), MSAs were created through concatenation of 122 phylogenetically informative protein- or protein domain-encoding sequences compiled in the Pfam v27 or TIGRFAMs v15.0 databases (55, 56). The 122 archaean marker proteins (see Table S2) were selected using previously reported criteria (57, 58). A maximum-likelihood tree was constructed by FastTree v2.0 under the JC model, based on 16S rRNA gene sequences (~1,500 bp) obtained from annotated files of the 352 genomes in the NCBI RefSeq Archaean genome database (https://ftp.ncbi.nlm.nih.gov/genomes/refseq/archaea/) or Silva database (https://www.arb-silva.de/). The 16S rRNA sequences were aligned by MAFFT 7.0 (59) and alignment positions with gap characters from >80% of the sequences were excluded by Gblocks 0.91b (60). A maximum-likelihood phylogenetic tree based on the alignments was constructed by FastTree v2.0 under the JC model, and bootstrap analysis was applied with 1,000 replications (49).

Identification and analysis of protein components. The genes encoding all chemoreceptors in the selected genomes were identified by BLASTP, and hits with an E value of <e^-30 were assigned as candidates. The protein sequence of the MA domains of chemoreceptors from Methanothermobacter thermautotrophicus holandaica DSM_15978 were selected as query sequences to perform the BLASTP analysis. All the candidates were reexamined for their domain organization by searching the SMART, HHpred, and CDvist databases manually (52, 54, 61). All chemoreceptor proteins are listed in Table S6. The genes encoding chemotaxis proteins (CheA, CheW, and CheY) and archaeal proteins (Ada and Arsβ) in the selected genomes were identified by BLASTP, and the hits with E value of <e^-20 were assigned as candidates, but a threshold E value of 10 was applied for CheW and CheY. All of the candidates were reexamined for their domain organization by searching the SMART, HHPred, and CDvist databases manually (52, 54, 61). All identified proteins are listed in Table S6. For CheY, we chose sequences that were >200 amino acids in length and only had a receiver (REC) domain; all of the CheY candidate proteins are shown in Table S7.

Classification of MCPs based on LBD. The chemoreceptors were classified into different five classes based on the LBD locations, and each type was further divided into several subgroups based on their membrane topology. MA domains were divided into different H types (such as 4H, 40H, 28H, etc.) by Alexander and Zhulin (41) according to the sequence length and conservation of the MA domain in MCPs. All H types for MCPs in each genome were obtained from the MiST 3.0 database (61). The classifications of chemoreceptors are summarized in Table S4.

Sample collection and analysis. Archaea of diverse habitats and metabolic types were included in the analysis. Strain niches were characterized based on their isolation sites, which were determined manually by searching the NCBI Biosample (https://www.ncbi.nlm.nih.gov/biosample/?term=), DSMZ (https://www.dsmz.de/databases/dsmz-archaea-profile-and-figure/), GTDB (https://gtdb.gov/genomes), and ATCC (https://www.atcc.org/) databases (55). A Student t test was used to determine the significance of differences between groups of niches, and a Venn network was constructed by software accessed through Evven (62; http://www.ehbio.com/test/venn/#!/).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.7 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.2 MB.

ACKNOWLEDGMENTS

This study was supported by the China Postdoctoral Science Foundation (grant 2021M693468), the National Nature Science Foundation of China (grant 31970066), Agricultural Science and Technology Innovation Project of the Chinese Academy of Agriculture Science (Grant No. CAAS-ASTIP-2016-BIOMA) and the Central Public-interest Scientific Institution Basal Research Fund (Grant Y2021PT02, Grant Y2021XK06), and the Sichuan Provincial International Science and Technology Innovation Cooperation Project (grant 2021YFH0049).

G.C. and Y.L. carried out comparative analyses of the archaeal genomes to identify the MCPs reported here. Y.L. and Q.Y. constructed the phylogenetic trees. G.C., Y.L., Q.Y., L.B., L.C., Sichuan Provincial International Science and Technology Innovation Cooperation Project.
REFERENCES

1. Wang Y, Wegener G, Williams TA, Xie R, Hou J, Zhang Y, Xiao X. 2021. A methylotropic origin of methanogenesis and early divergence of anaerobic multicarbon alkaline metabolism. Sci Adv 7:eabd7180. https://doi.org/10.1126/sciadv.7180.

2. Adam PS, Borrel G, Brochier-Armanet C, Gribaldo S. 2017. The growing tree of Archaea: new perspectives on their diversity, evolution, and ecology. ISME J 11:2407–2425. https://doi.org/10.1038/ismej.2017.122.

3. Spang A, Caceres EF, Ettema TJG. 2017. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science 357: eaaf3883–10. https://doi.org/10.1126/science.aaf3883.

4. Evans PN, Boyd JA, Leu AO, Woodcroft BJ, Parks DH, Hugenholtz P, Tyson GW. 2019. An evolving view of methane metabolism in the archaea. Nat Rev Microbiol 17:219–232. https://doi.org/10.1038/s41579-018-0136-7.

5. Shu WS, Huang LN. 2022. Microbial diversity in extreme environments. Nat Rev Microbiol 2021–235. https://doi.org/10.1038/s41579-021-00648-y.

6. Borrel G, Brugere JF, Gribaldo S, Schmitt RA, Moissl-Eichinger C. 2020. The host-associated archaeaome. Nat Rev Microbiol 18:622–636. https://doi.org/10.1038/s41579-020-0407-y.

7. Nayfach S, Roux S, Seshardi R, Udway D, Varghese N, Schulz F, Wu D, Paez-Espino D, Chen I-M, Huntemann M, Palaniappan K, Ladau J, Mukherjee S, Reddy TBK, Nielsen T, Kirtom E, Faria JP, EDRISINHGE JS, Henry CS, JUNGBLUTH W. 2021. A genomic catalog of Earth microbiomes. Nat Rev Microbiol 18:622–636. https://doi.org/10.1038/s41579-020-0407-y.

8. Robinson NP, Quax TEF, Albers SV, Pfeiffer F. 2018. Taxis in archaea. Emerg Top Life Sci 2:535–546. https://doi.org/10.1042/ETLS20180089.

9. Raina JL, SBLS, Parks DH, Rinke C, Snipen L, Bramucci A, Ostrowski M, Signal B, Lutz A, Mendis H, Rubino F, Fernandez VL, Stocker R, Hugenholtz P, Tyson GW, Seymour JR. 2022. Chemotaxis shapes the microscale organization of the ocean’s microbiome. Nature 605:132–138. https://doi.org/10.1038/s41586-022-04614-3.

10. York A. 2022. Chemotaxis in the ocean. Nat Rev Microbiol 20:381–381. https://doi.org/10.1038/s41579-022-00743-8.

11. Briegel A, Ortega DR, Huang AN, Oikonomou CM, Gunsalus RP, Jensen GJ. 2019. An evolving view of methane metabolism in the Archaea. Nat Rev Microbiol 17:219–232. https://doi.org/10.1038/s41579-018-0136-7.

12. Szurmant H, Ordal GW. 2004. Diversity in chemotaxis mechanisms among Bacteria and archaea. Microbiol Mol Biol Rev 68:301–319. https://doi.org/10.1128/MMBR.68.2.301-319.2004.

13. Li J, Zheng X, Guo X, Qi L, Dong X. 2014. Characterization of an archaeal two-component system that regulates methanogenesis in Methanosaeta harundinacea. PLoS One 9:e95502. https://doi.org/10.1371/journal.pone.0095502.

14. Schlesner M, Miller A, Streif S, Staudinger WF, Muller J, Scheffer B, Siedler F, Oesterhelt D. 2009. Identification of Archaea-specific chemotaxis proteins which interact with the flagellar apparatus. BMC Microbiol 9:56. https://doi.org/10.1186/1471-2180-9-56.

15. Quax TEF, Altegoer F, Rossi F, Li Z, Rodriguez-Franco M, Kraus F, Bange G, Albers SV. 2018. Structure and function of the archaeal response regulator CheY. Proc Natl Acad Sci U S A 115:E1259–E1268. https://doi.org/10.1073/pnas.1716661115.

16. Chen SM, Staudinger WF, Oesterhelt D. 2012. The protein interaction network of a taxis signal transduction system in a halophilic Archaeon. BMC Microbiol 12:272. https://doi.org/10.1186/1471-2180-12-272.

17. Zhang W, Bouroon A, Mueller MM, Alam M. 1996. The primary structures of the archaeon Halobacterium salinarium blue light receptor sensory rhodopsin II and its transducin, a methyl-accepting protein. Proc Natl Acad Sci U S A 93:8230–8235. https://doi.org/10.1073/pnas.93.16.8230.

18. Koch MK, Staudinger WF, Siedler F, Oesterhelt D. 2008. Physiological sites of deamination and methyl esterification in sensory transducers of Bacterium salinarium. J Mol Biol 380:285–302. https://doi.org/10.1016/j.jmb.2008.04.063.
transmembrane signal transfer. FEBS Lett. 564:219–224. https://doi.org/10.1016/S0014-5793(04)00193-0.
41. Alexander RP, Zhulin IB. 2007. Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. Proc Natl Acad Sci U S A 104:2885–2890. https://doi.org/10.1073/pnas.0609359104.
42. Gordel’yi VI, Labahn J, Mokhanmetjanov R, Efremov R, Granzin J, Schlesinger B, Bulté G, Savopol T, Scheidig AJ, Klare JP, Engelhardt M. 2002. Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. Nature 419:484–487. https://doi.org/10.1038/nature01109.
43. Mo R, Liu YG, Chen YY, Mao YJ, Gao BL. 2022. The evolutionary path of chemosensory and flagellar macromolecular machines in Campylobacterota. PLoS Genet 18:e1010316-24. https://doi.org/10.1371/journal.pgen.1010316.
44. Gestwicki JE, Lamanna AC, Harshey RM, McCarter LL, Kiesling LL, Adler J. 2000. Evolutionary conservation of methyl-accepting chemotaxis protein location in Bacteria and Archaea. J Bacteriol 182:6499–6502. https://doi.org/10.1128/JB.182.22.6499–6502.2000.
45. Sheridan PO, Raguideau S, Quince C, Holden J, Zhang L, Thames C, Williams TA, Gubry-Rangin C, Thames Consortium. 2020. Gene duplication drives genome expansion in a major lineage of Thaumarchaeota. Nat Commun 11: 5494. https://doi.org/10.1038/s41467-020-19132-x.
46. Ortega DR, Fleetwood AD, Krell T, Harwood CS, Jensen GJ, Zhulin IB. 2017. Assigning chemoreceptors to chemosensory pathways in Pseudomonas aeruginosa. Proc Natl Acad Sci U S A 114:12809–12814. https://doi.org/10.1073/pnas.1708842114.
47. Spang AJHS, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJG. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521:173–179. https://doi.org/10.1038/nature14447.
48. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2: approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
49. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
50. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. 2019. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res 47:e5–W10. https://doi.org/10.1093/nar/gkz342.
51. Letunic I, Khedkar S, Bork P. 2021. SMART: recent updates, new developments and status in 2021. Nucleic Acids Res 49:D458–D460. https://doi.org/10.1093/nar/gkaa937.
52. Letunic I, Khedkar S, Bork P. 2021. SMART: recent updates, new developments and status in 2021. Nucleic Acids Res 49:D458–D460. https://doi.org/10.1093/nar/gkaa937.
53. Adebali O, Ortega DR, Zhulin IB. 2015. CDvist: a webserver for identification and visualization of conserved domains in protein sequences. Bioinformatics 31:1475–1477. https://doi.org/10.1093/bioinformatics/btu836.
54. Gabler F, Nam S-Z, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V. 2020. Protein sequence alignment using the MPI bioinformatics tool kit. Curr Protoc Bioinformatics 72:e108. https://doi.org/10.1002/cpbl.108.
55. Rinke C, Chuvcholina M, Mussig AJ, Chaumeil P-A, Davin AA, Waite DW, Whitman WB, Parks DH, Hugenholtz P. 2021. A standardized archaeal taxonomy for the Genome Taxonomy Database. Nat Microbiol 6:946–959. https://doi.org/10.1038/s41564-021-00918-8.
56. Baker BJ, De Anda V, Seitz KW, Dombrowski N, Santoro AE, Lloyd KG. 2020. Diversity, ecology, and evolution of Archaea. Nat Microbiol 5:887–900. https://doi.org/10.1038/s41564-020-0715-z.
57. Parks DH, Rinke C, Chuvcholina M, Chaumeil PA, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol 2:1533–1542. https://doi.org/10.1038/s41564-017-0012-7.
58. Jie LXZ, Li L, Zhang S, Ren M, Huang L, Dong X. 2021. The archaeal transcription termination factor aCPSF1 is a robust phylogenetic marker for archaeal taxonomy. Microbiol Spectr 9:e01539-21. https://doi.org/10.1128/spectrum.01539-21.
59. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:727–740. https://doi.org/10.1093/molbev/mst010.
60. Talavera G, Castresana J. 2007. Improvement of phylogeny after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56:564–577. https://doi.org/10.1080/10635150701472164.
61. Gumerov VM, Ortega DR, Adebali O, Ulrich LE, Zhulin IB. 2020. MiST 3.0: an updated microbial signal transduction database with an emphasis on chemosensory systems. Nucleic Acids Res 48:D459–D464. https://doi.org/10.1093/nar/gkaa988.
62. Chen T, Zhang H, Liu Y, Liu YX, Huang L. 2021. eVenn: easy to create repeatable and editable Venn diagrams and Venn networks online. J Genet Genomics 48: 863–866. https://doi.org/10.1016/j.jgg.2021.07.007.