Decoding the chemical language of RiPPs from the untapped Archaea domain

Zhi-Man Song[1,2], Cunlei Cai[1,2], Ying Gao[1], Xiaoqian Lin[1], Qian Yang[1], Dengwei Zhang[1], Gengfan Wu[1], Haoyu Liang[1], Qianlin Zhuo[1], Junliang Zhang[1], Peiyan Cai[1], Haibo Jiang[1], Wenhua Liu[2] and Yong-Xin Li[1*]

[1] Department of Chemistry and The Swire Institute of Marine Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China

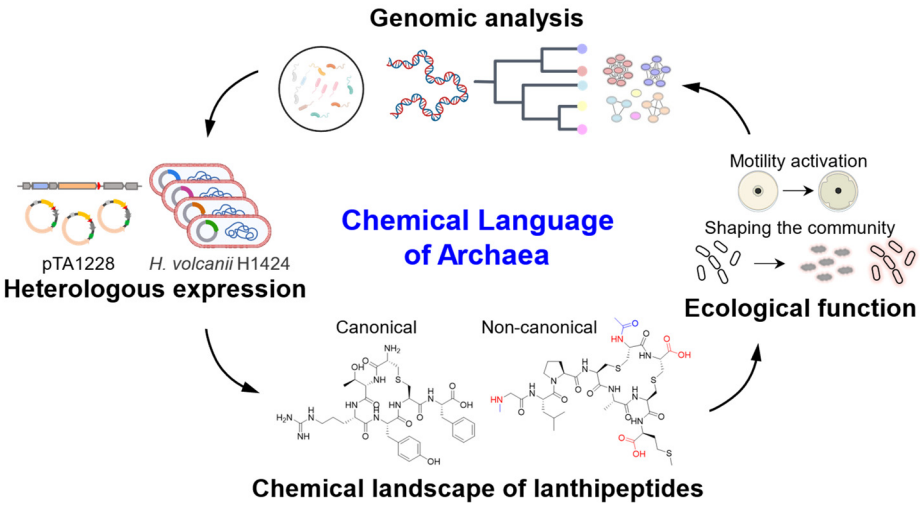
[2] Chemistry and Chemical Engineering Guangdong Laboratory, Shantou 515031, China

[*]To whom correspondence may be addressed.

Email: yxpli@hku.hk

Abstract

Chemical communication is crucial in ecosystems with complex microbial communities. However, the difficulties inherent to the cultivation of archaea have led to a limited understanding of their chemical language, especially regarding the structure diversity and function of secondary or specialized metabolites (SMs). Our comprehensive investigation into the biosynthetic potential of archaea, combined with metabolic analyses and the first report of heterologous expression in archaea, has unveiled the previously unexplored biosynthetic capabilities and chemical diversity of archaeal ribosomally synthesized and post-translationally modified peptides (RiPPs). We have identified twenty-four new lanthipeptides of RiPPs exhibiting unique chemical characteristics, including a novel subfamily featuring an unexplored type with diamino-dicarboxylic (DADC) termini, largely expanding the chemical landscape of archaeal SMs. This sheds light on the chemical novelty of archaeal metabolites and emphasizes their potential as an untapped resource for natural product discovery. Additionally, archaeal lanthipeptides demonstrate specific antagonistic activity against haloarchaea, mediating the unique biotic interaction in the halophilic niche. Furthermore, they showcase a new ecological role of RiPPs in enhancing the host's motility by inducing the rod-shaped cell morphology and upregulating the archaellin gene expression, facilitating the archaeal interaction with abiotic environments. These discoveries broaden our understanding of archaeal chemical language and provide promising prospects for future exploration of SM-mediated interaction.



Main

Archaea, the third domain of life, have evolved distinctly from bacteria and eukaryotes and are universally present in varied ecological niches[1]. They are particularly abundant in extreme environments, where their cultivation often poses challenges despite their significant roles in ecosystems[2]. Archaea have developed various mechanisms to thrive in different environments, involving direct and indirect interactions with biotic and abiotic factors[3-9]. These interactions extend to symbiosis, predation, antagonism, mutualism, and participation in essential processes like carbon, nitrogen, and sulfur cyclings. Bacteria and fungi are renowned for their production of a wide range of secondary or specialized metabolites (SMs), which include signaling molecules, antibiotics, and siderophores that aid in their interaction with their biotic and abiotic environments. However, our understanding of chemical-mediated interactions in archaea lags far behind that of bacteria, except for certain UV-protective carotenoids[10], putative siderophores[11] and quorum-sensing molecules[12,13].

Archaea possess a variety of enzymes and metabolic pathways distinct from those in bacteria[14-18], marking them as a promising and largely untapped reservoir of novel enzymes and SMs. However, due to the challenges in archaeal cultivation and genetic manipulation, the chemical landscape, biosynthesis, and ecological function of archaeal metabolites remain largely unexplored[2,19,20]. Haloarchaea, easily lab-cultured and known for producing antagonistic proteins like halocins[21,22], are ideal models for studying archaeal interactions, cell biology, and metabolic pathways[6,23-25]. Despite the presence of hundreds of biosynthetic gene clusters (BGCs)[26-29] and numerous attempts of wild-type strain fermentation, previous research identified only two ribosomally synthesized and post-translationally modified peptides (RiPPs) from Haloarchaea[26]. Expressing archaeal biosynthetic pathways heterologously in bacterial hosts can be problematic due to distinct metabolic features, potentially leading to improper folding, low expression levels, or inactive archaeal enzymes[30]. The scarcity of SMs from even well-defined archaeal BGCs predicted bioinformatically *in silico*[26-29,31], alongside the challenges involved in their heterologous expression in model bacterial hosts, underscores the necessity for new strategies in discovering archaeal metabolites.

Uncovering the yet unexplored chemistry of archaeal SMs will provide crucial insights into their chemically mediated interactions, invigorating the relatively underexplored field of archaeal chemical biology. Here, our study reports the systematic investigation and the first heterologous expression-based discovery of SMs in archaea, aiming to uncover their biosynthetic potential, chemical landscape, and ecological functions. We successfully validated twenty-four classic and noncanonical lanthipeptides of RiPPs originating from eleven BGCs by combining the heterologous expression with biosynthetic rule-based metabolomic analysis[32]. These archaeal lanthipeptides exhibited a specific antagonistic activity against Haloarchaea. Intriguingly, our investigation unveiled a new ecological function of lanthipeptides in stimulating the heterologous host's motility, by promoting a rod-like morphology and upregulating archaellin gene expression. These observations indicate that lanthipeptides potentially enhance the motility of archaea, thereby improving nutrient uptake or space occupation for environmental adaptation. Our study represents a significant advancement in understanding the archaeal chemical landscape and paves the way for archaeal natural product discovery and chemical ecology study.

Results

Genome mining reveals the largely untapped biosynthetic potential of archaeal RiPPs

Our previous genomic analysis indicated that archaea contain a relatively small number of BGCs per genome compared to the extensively studied and well-documented bacterial counterparts[26]. They encode a broad range of known SM classes, such as RiPPs, terpenes, nonribosomal peptides (NRPs), and polyketides (PKs). Our initial fermentation-based discovery identified a new lanthipeptide from Haloarchaea, marking the first discovery of lantibiotic and anti-archaeal SM in the archaeal domain[26]. In the current study, we further applied the antiSMASH 7.0 tool[33], which is capable of analyzing both bacterial and archaeal BGCs, to examine 11,644 archaeal genomes from the NCBI database to delve deeper into exploring the biosynthetic landscape, diversity, and novelty of archaeal RiPPs. From 3,451 representative archaeal genomes, 7,731 BGCs were identified, mainly from phyla Halobacteriota (3,379, 43.7%), Thermoplasmatota (1,359, 17.6%), Methanobacteriota (1,271, 16.4%), Thermoproteota (962, 12.4%) and Nanoarchaeota (221, 2.9%) (Fig. 1a and Supplementary Data 1). Among them, RiPPs (3,599, 46.6%), Terpenes (1,990, 25.7 %), and NRPs (581, 7.5 %) emerged as dominant BGCs in the domain of Archaea (Fig.1a). These BGCs exhibited significant novelty compared to known BGCs from MIBiG[34], indicating a largely unexplored source for natural product discovery.

Phylogenetic distribution analysis unveiled a remarkable diversity within RiPP BGCs, widely distributed across the Asgard, TACK, DPANN superphyla, and predominantly within Euryarchaeota (Fig. 1a). Among these abundant RiPP classes are YcaO-related RiPPs (thioamitides and thiopeptides), radical S-adenosylmethionine (rSAM)-dependent enzyme-modified RiPPs (ranthipeptides and sactipeptides), lanthipeptides, and lassopeptides (Fig. 1b). While RiPPs are widely distributed throughout different phyla/classes, the scarcity of identified archaeal RiPPs hinders us from elucidating their unique chemical structure and putative ecological functions. Our attempts to activate YcaO- and rSAM-related BGCs in the wild-type strain or through heterologous expression have been unsuccessful, prompting a shift in focus towards lanthipeptides. Class II lanthipeptides, unique among other RiPPs, are solely present in Haloarchaea (refer to the class Halobacteria), implying halophilic niche-specific metabolites (Fig. 1b, Supplementary Data 1 and 2). Consequently, an updated library was constructed *in silico* to decipher the chemical landscape of archaeal lanthipeptides. This library contains 353 putative precursors (LanAs) and 110 lanthipeptide synthase LanMs from 102 archaeal lanthipeptide BGCs, through antiSMASH-based genome mining method[26,33] (Supplementary Data 2). Among these, 67 deduplicated archaeal LanMs, exclusively present in Haloarchaea, are mainly from the genera *Halorussus* (26, 38.8%) and *Haloferax* (14, 20.9%), followed by *Haladaptatus* (5, 7.5%), *Halobacteriales* (5, 7.5%), *Natrinema* (5, 7.5%) (Supplementary Fig. 1a and Supplementary Data 2). Moreover, 196 unique LanA sequences distinctive from bacterial LanA have been categorized into 157 families through sequence similarity network (SSN) analysis (Fig. 1c), highlighting untapped chemical space for lanthipeptide discovery.

Co-evolved archaeal LanMs and LanAs exhibit environmental adaptation

Our collective analysis of the phylogenetic tree, amino acid composition bias, SSN, and sequence logo indicates a unique chemical distinctiveness of archaeal lanthipeptides compared to their bacterial

counterparts (Fig. 1 and Supplementary Figs. 1-2). The phylogenetic tree of archaeal LanMs highlighted their evolutionary divergence from bacterial LanMs (Supplementary Fig. 1b and Supplementary Data 2). The tree of archaeal LanAs exhibited two primary clades: one encompassing our previously characterized LanAs, Alnα and Alnβ, and another unexplored clade that offers prospects for further investigation (Supplementary Fig. 2). Additionally, archaeal LanAs are predicted to be shorter and comprise a higher portion of C and S instead of T residues than bacterial LanAs (Fig. 1c, d), which are crucial conserved residues for thioether ring formation in lanthipeptide biosynthesis[35]. More importantly, the distinctive amino acid composition bias of archaeal LanAs and LanMs supports the hypothesis that they underwent eco-evolution in response to hypersaline environments (Fig. 1d, e). Specifically, archaeal LanMs and LanAs evolved to over-represent D, V, R, and P residues while under-representing K, I, N, Q, M, L, W, and F residues (Fig. 1d, e). The increased abundance of D/E residues and the reduced portion of K residue is linked to the "salt-in" strategy of Haloarchaea, enabling interactions with cations in their high-salinity environment[6,36]. Furthermore, the reduced occurrence of hydrophobic amino acids (I, L, M, W) may increase random-coil structure formation, thereby preventing protein aggregation or inactivation in low-water environments. The prevalence of basic R residue in archaeal LanMs and LanAs, as pronounced as acidic D residue, may be a unique evolutionary trait of archaeal lanthipeptide biosynthetic pathways that have not been observed in other protein analyses to the best of our knowledge[36]. Considering the evolutionary divergence of LanMs and their amino acid composition biases, the co-evolutionary relationship between archaeal LanMs and LanAs hints at a unique evolutionary path for archaeal lanthipeptides, which occupy distinct chemical realms (Fig. 1f). These results imply that archaeal LanAs, having co-evolved with LanMs as a part of environmental adaptation, could potentially enhance the chemical diversity of lanthipeptides exclusive to Haloarchaea, thereby playing significant ecological roles in halophilic environments.

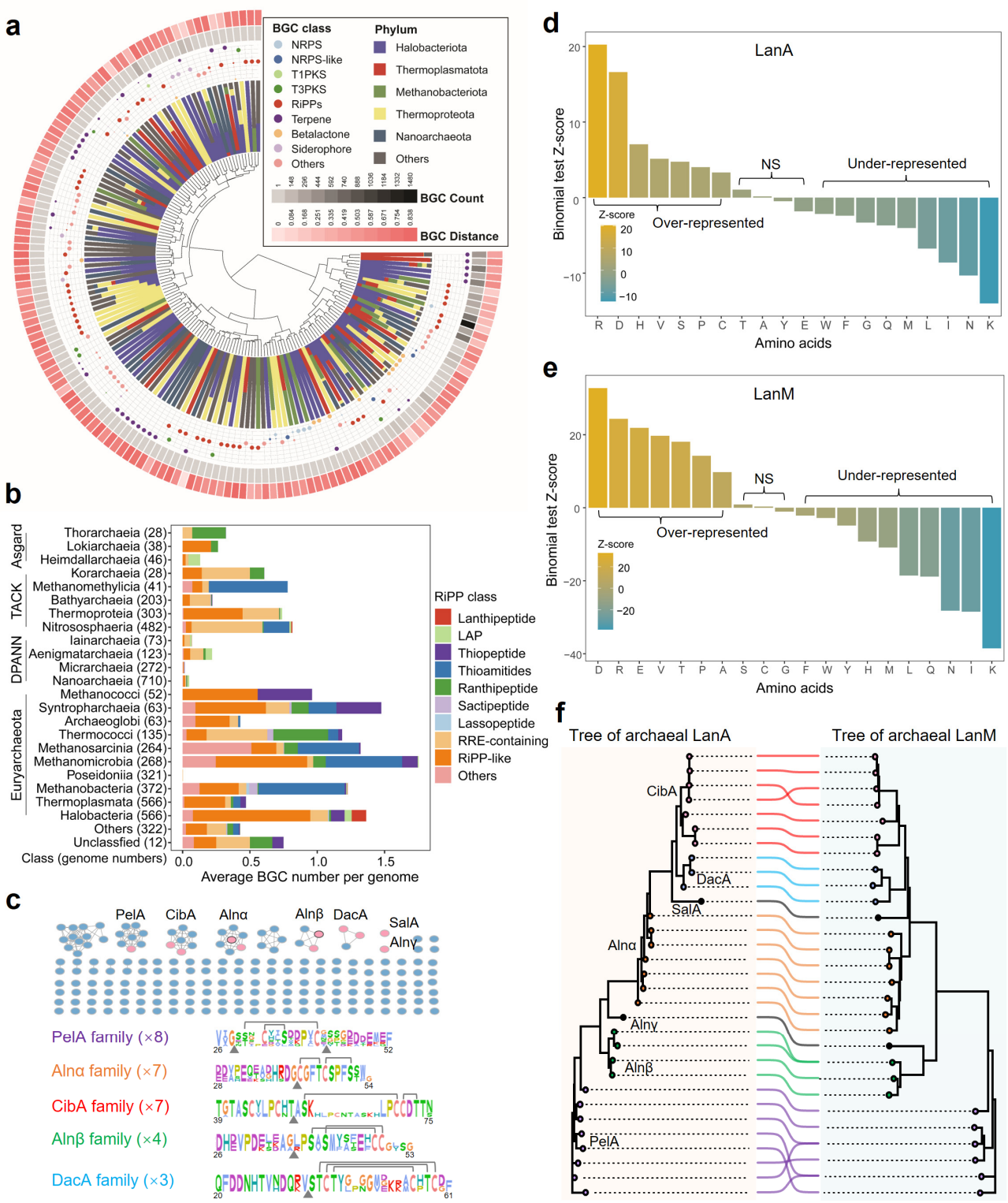


Figure 1. Uncovering the biosynthetic potential of archaeal RiPPs. a, The novelty, diversity, and distribution of archaeal BGCs. Layers from innermost to outermost show hierarchical clustering of the 148 gene cluster clans (GCCs) based on the BGC distances, phylum distribution for each clan, the proportion of different BGC classes for each clan, BGC count, and average cosine distance to known BGCs. **b**, The distribution of RiPP BGCs in the domain of Archaea at the class level. **c**, The SSN analysis of putative archaeal LanAs and sequence logos of experimentally verified LanA families. Sequences with 100% identity are grouped into a single node. Pink nodes represent the characterization

of heterogeneous expression products or wild-type natural products in this study. Lines show the validated crosslinks between C and S/T residues. The main cleavage sites observed between the leader and core peptides are displayed as triangles below the LanAs. **d-e**, Binomial tests were conducted to compare the proportions of all 20 amino acids in LanAs (**d**) and LanMs (**e**) between archaeal and bacterial counterparts. *Z*-scores were calculated relative to archaeal LanAs and LanMs, where $Z > 2.0$ indicated significant enrichment of a given amino acid in Archaea ('Over-represented'), $Z < -2.0$ indicated significant depletion ('Under-represented'), and some amino acids showed no significant biases ('NS'). **f**, Co-phylogenomic analysis of archaeal LanAs (left) and LanMs (right) from verified SSN families. The nodes and lines are colored according to their LanA families in **c**. Two black notes represent verified orphan notes in the SSN analysis, SalA and Alnβ.

Heterologous expression-based discovery of archaeal lanthipeptides

Moving beyond the traditional method of natural product discovery through wild-type cultivation, we successfully heterologously expressed eleven BGCs from various SSN families in *Haloferax volcanii* H1424 using the pTA1228 vector[37,38] (Fig. 2a, Supplementary Table 2 and Supplementary Data 3). The procedure was initiated by cloning our previously verified lanthipeptide BGCs[26], *alnα* and *alnβ*, which resulted in the production of corresponding analogs of the natural lanthipeptides, archalans α2 (**7**), β1 (**9**), and β2 (**10)** (Fig. 2b, c and Supplementary Figs. 3-4). It marks the first endeavor to authenticate archaeal RiPP BGCs through heterologous expression (Fig. 2), laying the groundwork for the subsequent expression of other archaeal lanthipeptide BGCs. Based on this approach, we heterologously expressed fifteen lanthipeptide BGCs, covering all seven major lanthipeptide families in SSN (Fig.1). This endeavor linked nine BGCs from five SSN families and two orphan BGCs with their lanthipeptide products, representing the first comprehensive exploration of archaeal SMs so far (Figs.1, 2 and Supplementary Table 1).

Except for the unusual peptides from DcaA family, nine other classical lanthipeptides were pinpointed by biosynthetic rule-based metabolic analysis[32](Fig. 2 and Supplementary Figs. 5-18), including sallans (**1** and **2**, from BGC *sal*), ciblans (**3** and **4**, from BGCs *cib1* and *cib2*), pellan (**5**, from BGC *pel*), archalan γ (**6**, from BGC *alnγ*), medlan B (**8**, from BGC *medb*) and litlans (**11** and **12**, from BGC *lit*). It's important to note that sallans (**1** and **2**), pellan (**5**), and litlans (**11** and **12**) were not detected in their corresponding wild-type strains. Others were detected in relatively low titers compared to the results from heterologous expression. Most of these lanthipeptides are detected in low quantities, either in the wild type or the heterologous host. Over the past decades, this has been a major obstacle in discovering natural products from Archaea.

To pinpoint the low-yield BGC-encoded peptides within the metabolic profile, we mapped the calculated mass data, which was predicted based on lanthipeptide post-translational modification rules[32], with the HRMS data to detect corresponding hits. For instance, a UPLC-HRMS analysis of the culture extract from the heterologous expression strain of the BGC *sal* (*Halorussus salinus* YJ-37-H) exposed numerous low peptidic signals (Supplementary Fig. 5). The mass matching process revealed a hit of $[M + 2H]^{2+} = 379.6688$ (**2**) that matched the calculated mass data well of "CTRYSF" with one dehydration (calculated $[M + 2H]^{2+} = 379.6682$), linking this low-yield peptide to BGC *sal* (Fig. 2a). This underscores the benefit of using heterologous expression coupled with rule-based metabolic analysis to uncover the hidden chemistry in Archaea, especially those with low yields.

Identification of new classical lanthipeptides from Archaea

The structures of classical lanthipeptides **1**-**4** were elucidated through a comprehensive approach that integrated High-Resolution Mass Spectrometry (HRMS), nuclear magnetic resonance (NMR) analysis, and advanced Marfey's analysis. Following biosynthetic rule-based metabolic analysis, sallans A1 (**1**, observed $[M+2H]^{2+}$ = 437.1816, calculated $[M+2H]^{2+}$ = 437.1816, Δ = 0.0 ppm) and A2 (**2**, observed $[M+2H]^{2+}$ = 379.6688, calculated $[M+2H]^{2+}$ = 379.6682, Δ = 1.6 ppm) were found to be analogs encoded by BGC *sal*, which were predicted with one dehydration (-H_2O, -18 Da) on their corresponding core peptides (CTRYSFD for **1** and CTRYSF for **2**, Fig. 2a, Supplementary Fig. 5). The MS/MS data confirmed a C-S crosslink between Cys1 and Ser5 in all of them (b_5, -H_2O, -18 Da), with a distinction at the C-terminus due to the presence of an additional D residue in **1** (y_2, 'FD' residue, Supplementary Fig. 5). Furthermore, the 2D NMR data of purified **2** showed HMBC correlation of H_β to C_β (δ_H = 2.96, 3.31 ppm to δ_C = 35.9 ppm and δ_H = 3.07, 3.24 ppm to δ_C = 34.3 ppm) in two Ala residues confirming the thioether ring in Cys1-Ser5 (Supplementary Fig. 6 and Supplementary Table 3). Ciblan A1 (**3**, BGC *cib1*, *Haladaptatus cibarius* JCM19505) and ciblan A2 (**4**, BGC *cib2*, *H. denitrificans* DSM4425) were predicted with a similar modification on the core peptides of CibA1 (ASKLPCCDTTN) and CibA2 (SKLPCCDTTN). The modification includes two bicycle ring structures (Ser2-Cys6 and Cys7-Thr9 in A1, Ser1-Cys5 and Cys6-Thr8 in A2) and one unsaturated amino acid Dhb residue derived from the last Thr residue of each, supported by a series of b ions (b_6/b_5 for the first ring and b_9/b_8 for the second ring in ciblan A1/A2) and z_2 fragment (Supplementary Fig. 7). The NMR data of purified **3** exhibited the presence of Dhb residue (C_α= 131.0 ppm, C_β = 130.9 ppm, H_β = 6.66 ppm, C_γ = 13.4 ppm, H_γ = 1.86 ppm) derived from Thr10, along with two thioether rings in Ser2-Cys6 (δ_H = 2.97, 2.92 ppm of Ala6 to δ_C = 35.0 ppm of Ala2) and Cys7-Thr9 (δ_H = 2.98 ppm of Abu9 to δ_C = 41.2 ppm of Ala7) (Supplementary Fig. 8 and Supplementary Table 3). The stereochemistry of **2** and **3** were confirmed with L configuration in all unmodified amino acids and DL-Lan (2*S*,6*R*) or DL-MeLan (2*S*,3*S*,6*R*) residue in thioether rings (Supplementary Fig. 9 and Supplementary Table 4). Notably, **2** with only six amino acids represents the smallest natural lanthipeptide to date.

Due to their low yield, compounds **5**-**12** were primarily identified using HRMS and biosynthetic analysis. The MS/MS analysis and full reduction-desulfurization study of pellan (**5**, BGC *pel, H. pelagicus* RC-68) confirmed the presence of two dehydrations within residues Thr2, Cys4, Ser7 and Cys12, and proved interwind ring structures in the core peptide of PelA (STGCYISDKPVCI, Supplementary Fig. 10). The partial desulfurization product of archalan γ (**6**, BGC *alnγ*, *H. salinus* YJ-37-H) indicated one methylation on Cys1, residues Ser2, Cys3, and Cys8 reduced to be Ala, and an intact C-S crosslinking in Thr5-Cys9 (CSCYTRICCDIDS, Supplementary Figs. 11-15). Subsequent dithiothreitol reduction and MS^n analysis determined a disulfide bond between Cys1 and Cys8 and a thioether ring between Ser2 and Cys3 in **6** (Supplementary Figs. 16 and 17). Altogether, archalan γ (**6**) featured N-terminal methylation, a Cys1-Cys8 crosslink, and thioether crosslinks in Ser2-Cys3 and Thr5-Cys9. It expands the number of the smallest thioether ring containing lanthipeptides which only has a few reported cases (such as mersacidin, ltnα, and sacAα)[35].Archalan α2 (**7**, BGC *alnα*, *H. salinus* YJ-37-H) and medlan B (**8**, BGC *medb*, *H. mediterranei* ATCC33500) contained two thioether rings, supported by their key fragments of b_4 (-H_2O, C-S crosslink in Cys1-Thr4) and b_9 (-2H_2O, another crosslink in Cys5-Ser9) in MS/MS analysis (Supplementary Fig. 3), in line with the previously reported structure of archalan α[26]. Archalan β1-2 (**9** and **10**, BGC *alnβ*, *H. salinus* YJ-37-H) and litlan

A1-2 (**11** and **12,** BGC *lit*, *H. litoreus* JCM31109) exhibited similar crosslinks to the previously reported archalan β[26], featuring two interwind thioether rings that resulted in MS/MS fragments from the N/C-terminal overhang amino acids outside of the rings (Supplementary Figs. 4 and 18).

The chemical structures of mature archaeal lanthipeptides have a high appearance of acidic residues which may facilitate adaptation to high-cation cellular environments and maintain a stable compound structure to function efficiently in high-salt surroundings. Moreover, archaeal lanthipeptides revealed that they are typically shorter than bacterial lanthipeptides, such as nukacin ISK-1 and haloduracins (Fig. 2d). Despite being short in length, their bicycle and interwind crosslinking patterns and modifications including disulfide bond, methylation, dehydration, hydroxylation, are as intricate as those found in bacterial lanthipeptides.

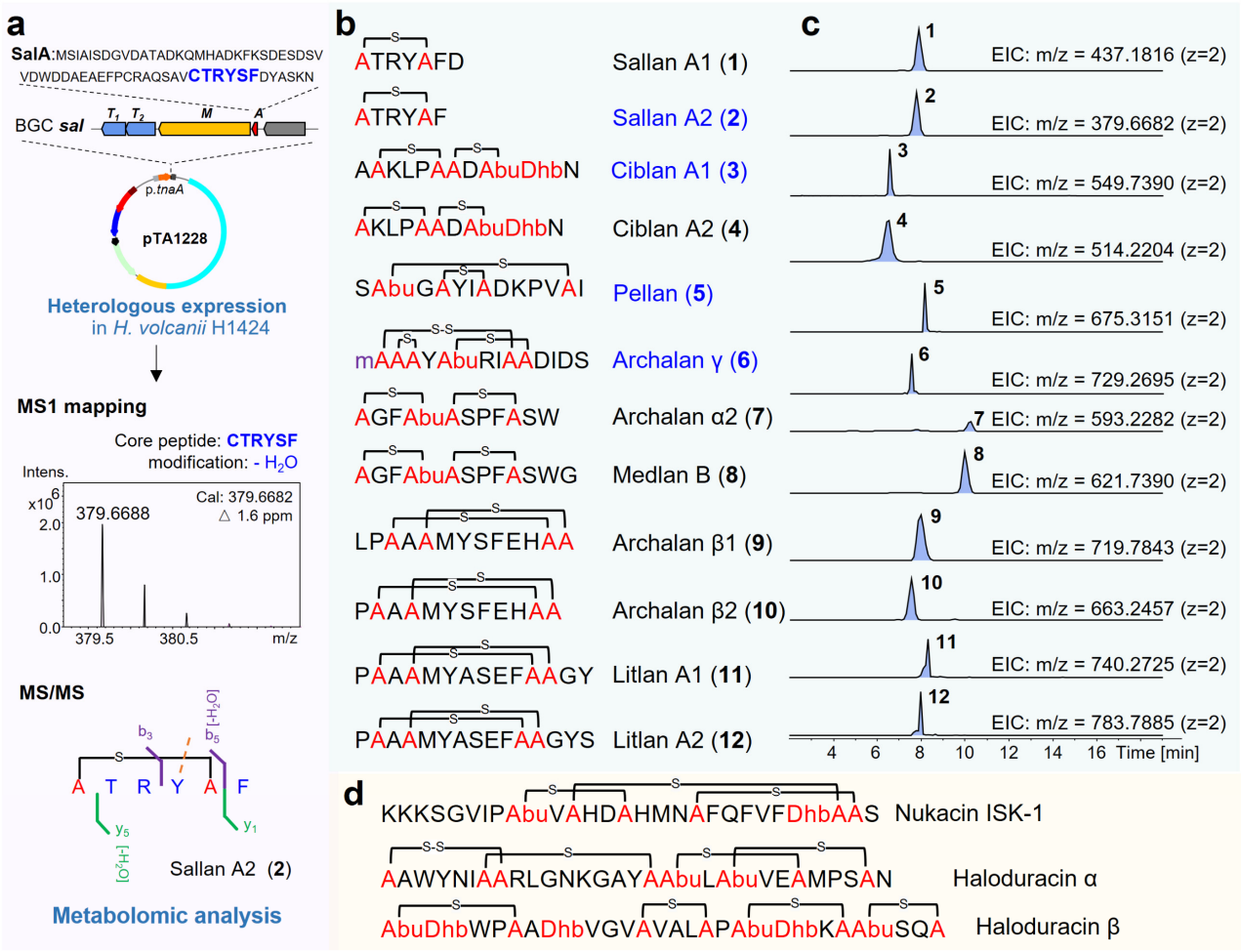


Figure 2. Deciphering novel archaeal class II lanthipeptides through heterologous expression and biosynthetic rule-based metabolic analysis. a, Workflow for the heterologous expression-based discovery of eleven lanthipeptide BGCs in *H. volcanii* H1424, in conjunction with metabolic analysis, is illustrated through the example of sallan A2 (**2**). Classical lanthipeptides (**b-c**) with different modifications ('m' in purple represents methylation) were identified via HR LC-MS. Structures confirmed through NMR analysis or chemical derivatization are marked in blue. **d**, Representative bacterial class II lanthipeptides.

Discovery of unique lanthipeptides featuring unprecedented maturation

Apart from the classical lanthipeptides mentioned above, several unusual peptide signals from BGCs *meda* and *lar* engrossed our attention for the unexpected modification (Fig. 3 and Supplementary Figs. 19-29). They were detected in the extracts of both wild-type and recombinant strains, suggesting that they were likely encoded by these two BGCs, albeit with unexpected modifications as revealed by MS/MS data. While the MS analysis of classical lanthipeptides allows for the accurate prediction of core peptide sequences and putative modifications, the MS/MS data of these peptide signals only provided short amino acid fragments without consistent patterns, such as 'FD' fragment from the C-terminus of LarA (CPTCDF), 'Y' residue also from C-terminus and 'S' residue from N-terminus (STCTY) (Fig. 3b and Supplementary Fig. 24).

To fully characterize their structures, 3.2 mg of **13** ($[M+3H]^{3+}$ = 415.4769, calculated $[M+3H]^{3+}$ = 415.4772, Δ = 0.7 ppm) from the wild-type strain, 3.5 mg of **14** ($[M+2H]^{2+}$ = 631.2194, calculated $[M+2H]^{2+}$ = 631.2198, Δ = 0.6 ppm), and 2.0 mg of **15** ($[M+2H]^{2+}$ = 652.2301, calculated $[M+2H]^{2+}$ = 652.2251, Δ = 7.7 ppm) from the recombinant strain were purified, which were named medlan A1 (**13**), larlan A2 (**14**) and larlan A5 (**15**), respectively. Further NMR spectroscopy analysis revealed that these lanthipeptides comprised two parallel but opposite-direction peptide chains crosslinked by thioether bonds, characterizing them as a novel lanthipeptide subfamily (Supplementary Figs. 20-22 and Supplementary Table 3). These crosslinked patterns expose diamino-dicarboxylic termini, leading to the name of DADC lanthipeptides, with certain amino acids within the thioether rings being partially cleaved (Fig. 3). In addition, medlan A2 (**16**), larlans A1 (**17**), A3 (**18**), and A4 (**19**) are analogs of medlan A1 (**13**) or larlan A2 (**14**) and have been deduced to exhibit longer or shorter overhang amino acids at the N/C terminus through LC-MS/MS analysis, possibly due to hydrolysis by unknown aminopeptidase (Supplementary Figs. 23-29). As for three MeLan residues in medlans and larlans increasing the complexity of stereochemistry configuration, two additional single mutations of Thr to Ser in precursor LarA were constructed. By purifying 0.3 mg of each mutant (T2S and T4S of larlan A2) for the advanced Marfey's analysis, a clear stereo structure annotation of each MeLan residue was achieved when compared with four MeLan standards[39,40] (Supplementary Fig. 9 and Supplementary Table 4). The thioether rings between Thr2 and Cys10, Thr9 and Cys3 are characterized in the LL-MeLan (2*R*,3*R*,6*R*) configuration, while another ring between Thr4 and Cys7 in DL-MeLan (2*S*,3*S*,6*R*) configuration in larlan A2 (Fig. 3).

To find out whether other lanthipeptides also employ a similar maturation strategy, the LC-MS/MS data of both wild-type and recombinant strains were meticulously analyzed. The metabolic profile of *H. salinus* YJ-37-H, harboring six lanthipeptide BGCs, displayed mass signals with a fragmental pattern resembling larlans and medlans (Supplementary Fig. 30). To characterize their structures, 1.0 mg of compound **20** ($[M+2H]^{2+}$ = 453.6800, calculated $[M+2H]^{2+}$ = 453.6796, Δ = 0.9 ppm) was obtained for NMR and advanced Marfey's analysis (Supplementary Figs. 31 and 32). As anticipated, archalan β3 (**20**), derived from BGC *alnβ*, shares the same characteristics as DADC lanthipeptides. Other derivatives, archalans β4-7 (**21-24**), were also detected and elucidated by LC-MS/MS with additional modifications or overhang amino acids (Supplementary Figs. 33-36). The advanced Marfey's analysis revealed that two Lan residues are LL (2*R*,6*R*) configurations, and other unmodified amino acids are proteinogenic L configurations in archalan β3 (Fig. 3, Supplementary Fig. 9 and Supplementary Table 4).

Identifying this particular lanthipeptide subfamily exclusively in Archaea highlights the existence of a unique biosynthetic pathway and enzymes that are distinct from those in bacteria. Synteny analysis[41] of the DacA family revealed the presence of three conserved genes: *A*, *M*, and *H*, which might serve as the boundary of DADC lanthipeptide BGCs (Supplementary Fig. 37 and Supplementary Data 3). The *lar* BGC was then selected for the in vivo reconstitution study which resulted in the production of a trace amount of larlan A2 when co-expressed genes *larA* and *larM* (Fig. 3b, c and Supplementary Fig. 37). This observation suggested the potential presence of generic peptidases from the heterologous host. On the contrary, a significantly increased production of larlan A2 was observed when co-expressed with the unannotated gene *larH* (Fig. 3c). This implied that LarH may play a role as a regulator, cofactor, chaperon, or a yet unidentified peptidase involved in the biosynthesis of larlan A2. In another case, we observed the production of classical lanthipeptide archalan β1-2 but didn't detect DADC lanthipeptide archalan β3-7 through heterologous expression, despite including the putative peptidase within the BGC (Supplementary Figs. 4 and 37). Although both classical and DADC lanthipeptides were identified in the wild-type strain, heterologous expression only yielded classical lanthipeptides of archalan β1-2, supporting the hypothesis that putative proteases outside the BGC are involved in biosynthesis and suggesting a complex biosynthetic pathway for DADC lanthipeptides. This pathway is predicted to include additional hydrolysis steps compared to classical lanthipeptides, facilitated by unidentified generic proteases to cleave amino acids within the thioether rings (Fig. 3d).

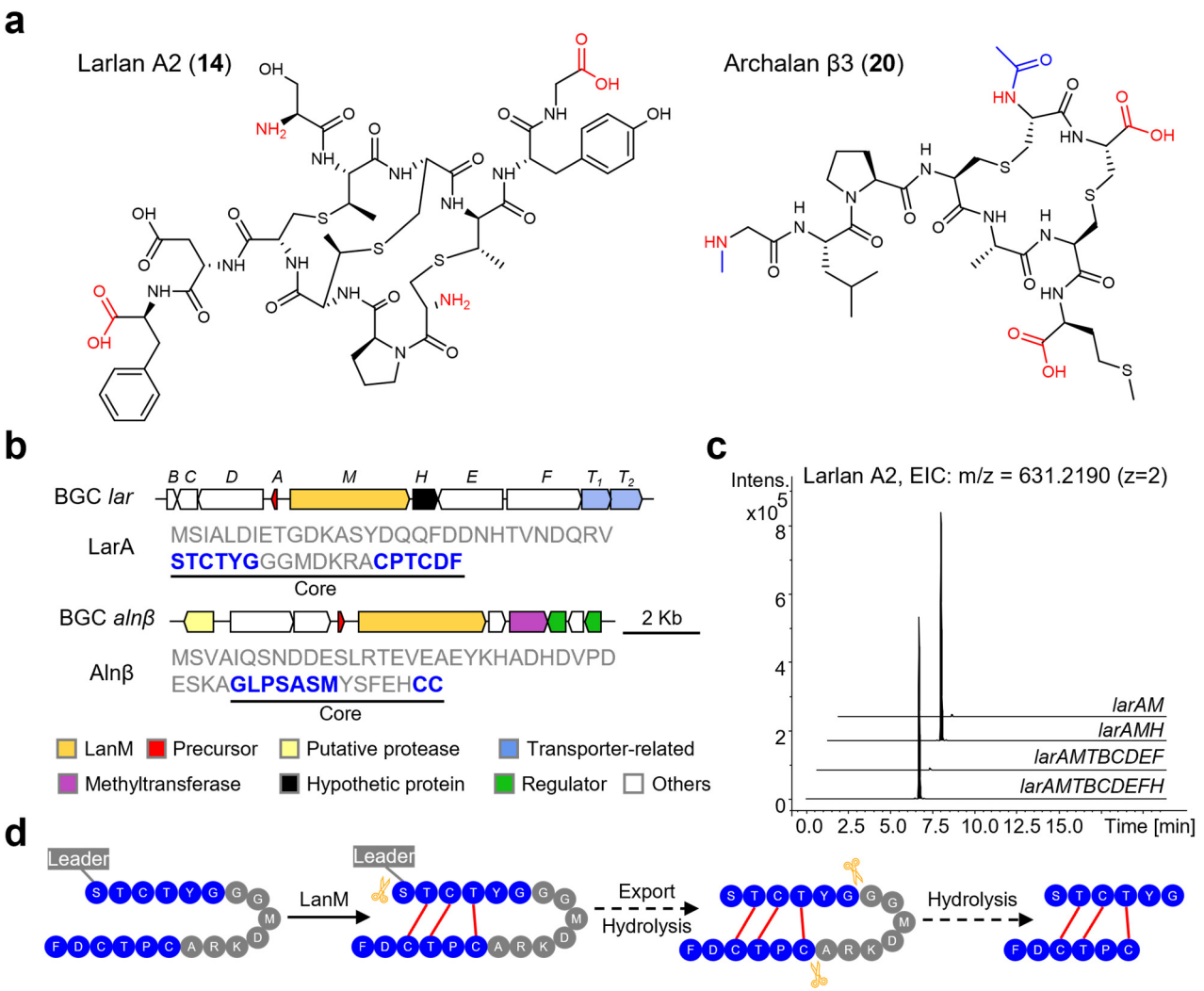


Figure 3. Novel chemistry and biosynthesis of archaeal lanthipeptides. a, Chemical structure of DADC lanthipeptides. Diamino-dicarboxylic termini are in red, and specific modifications at the N-terminus are in blue. **b**, BGCs *lar* and *alnβ* with their corresponding precursors. **c**, LC-MS analysis of in vivo reconstitution of BGC *lar* with/without gene *larH*. **d**, Proposed biosynthetic pathway of DADC lanthipeptides, demonstrated by larlan A2.

Archaeal lanthipeptides exhibit anti-archaeal and motility regulation activities

Given the chemical diversity and novelty of archaeal lanthipeptides, we next wanted to investigate their ecological functions. Six lanthipeptide crude extracts obtained from heterologous expression and seven purified lanthipeptides were collected for antimicrobial activity screening against nine haloarchaeal and six bacterial strains (Fig. 4). To exclude the potential influence of halocin produced by the heterologous host[42], the same gradient fractions from the host containing an empty vector were tested for comparison. Specifically, extracts from the heterologous expression of BGCs *medb, cib2,* and *alnα-γ* exhibited significant anti-haloarchaeal activity at 100 μg mL^{-1}, even though the lanthipeptides were in low proportion (Fig. 4a). The mature products from BGCs *alnα* and *medb*, archalan α (**25**), archalan α2 (**7**) and medlan B (**8**), were analogs exhibiting a similar anti-haloarchaeal spectrum (Fig. 4a). The purified lanthipeptide, archalan α (**25**), demonstrated strong inhibition against *H. larsenii* JCM 13917, *Haladaptatus cibarius* DSM 19505, *H. amylolyticus* JCM 18367, and *Halomicroarcula salina* JCM 18369 at 100 μg mL^{-1} (Fig. 4b). Notably, the MIC value of **25** against *H. salina* JCM 18369 was 6.25 μg mL^{-1} (Supplementary Fig. 38). On the contrary, the DADC lanthipeptide archalan β3 (**20**) exhibited lower activity than the classical lanthipeptide archalan β1-2 (**9**, **10**) produced by heterologous expression of BGC *alnβ*. The specific anti-haloarchaeal activity observed in archaeal lanthipeptides may be attributed to the characteristics of their habitats, which predominantly host Haloarchaea.

To study the in vivo function of lanthipeptide BGCs, stab-inoculation screens were conducted to assess their impact on the host's motility. Surprisingly, three recombinant strains were found to modulate the motility as either activator or inhibitor (Fig. 4c). Strains containing BGCs *cib2* and *pel*, responsible for the production of ciblan A2 (**4**) and pellan (**5**), exhibited a significantly larger motility zone than the control (Fig. 4c, d, Supplementary Fig. 39 and Supplementary Data 4). Subsequently, strain with BGC *cib2* was selected for time-course motility assays to gain a deeper insight, indicating that BGC *cib2* significantly activated the host's motility zone in the beginning without affecting growth and all strains displayed the Bull's eye pattern in the late stage (Supplementary Fig. 39). Conversely, BGC *alnα*, responsible for the production of archalan α (**25**), inhibited the motility may be attributed to its antagonistic activity, which could impede host's growth (Fig. 4b, c, and Supplementary Fig. 39). No obvious motility regulation assay in BGC *cib1*, in contrast to BGC *cib2* with identical precursor peptides, may be attributed to structure difference of their key products (Fig. 2 and Supplementary Fig. 4). These findings point to a new ecological role of lanthipeptides, potentially reducing the host's reaction time or increasing that of other strains in response to stimuli, by regulating motility activity[23,24,43].

A comprehensive investigation combining morphological and transcriptomic analysis was conducted to elucidate the mechanisms by which archaeal lanthipeptides enhance motility. Microscopic analysis of recombinant strains demonstrated that lanthipeptides could elevate the frequency of rod-shaped cell morphology, a characteristic associated with hypermotility[23,24,43] (Fig. 4 e-g and Supplementary Data

5). Additionally, the transcriptomic analysis revealed that the archaellin gene *flgA1*, a major flagellin gene responsible for the motility of *H. volcanii*[44], was upregulated in strains possessing BGCs *cib2* or *pel* (Fig. 4h, i and Supplementary Data 6). These studies support the hypothesis that lanthipeptides may promote motility by inducing a rod-shaped cell morphology and upregulating the expression of the archaellin gene *flgA1*. These lanthipeptide-mediated regulations could potentially confer advantages to their host in environmental adaptation, such as occupying space, enhancing nutrient acquisition, and evading predation (Fig. 4j).

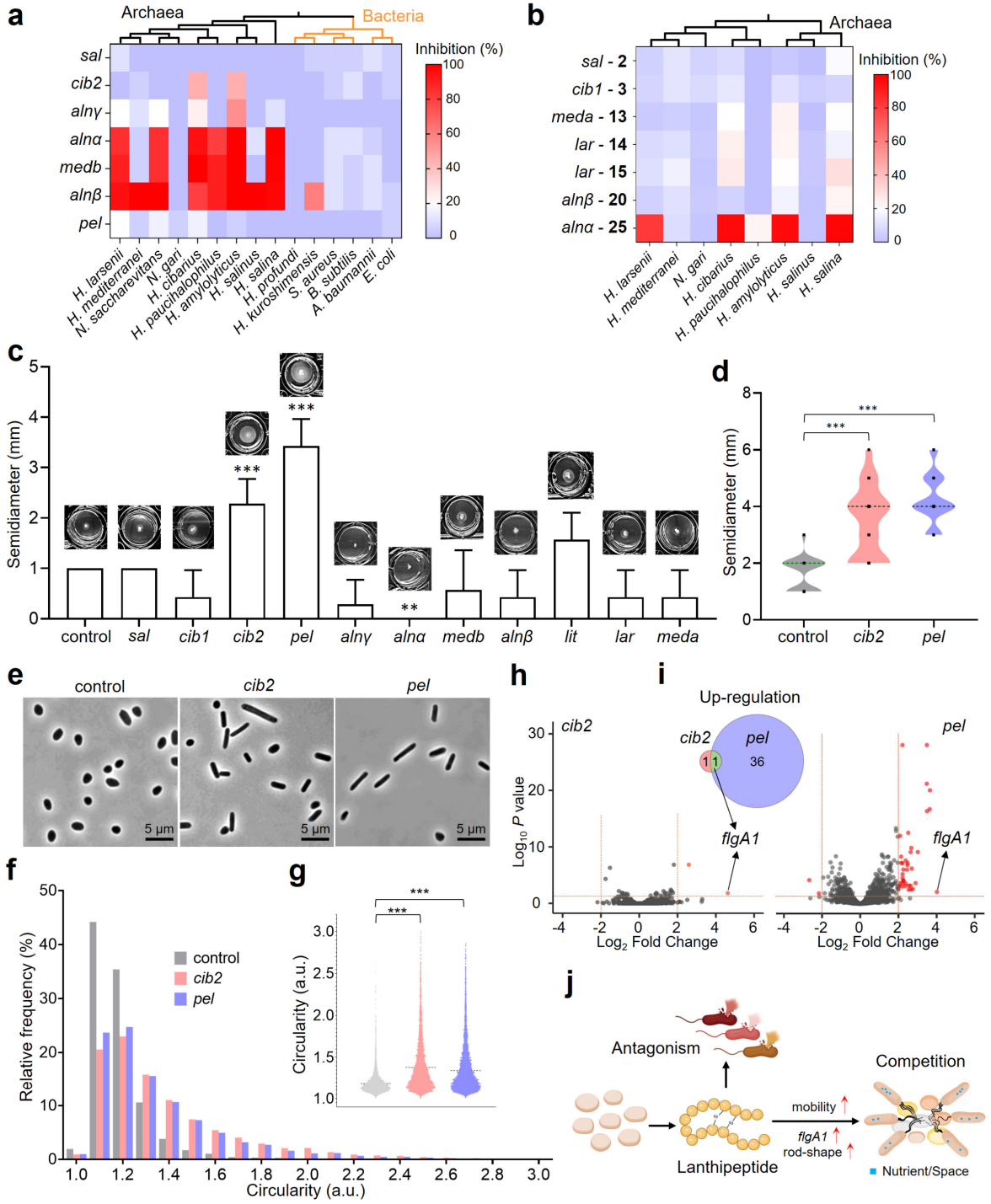


Figure 4. Ecological functions of archaeal lanthipeptides. The antimicrobial activity of the

fractionated crude extracts of recombinant strains with different BGCs (**a**) and purified archaeal lanthipeptides at 100 μg mL^{-1} (**b**). Compounds **2**, **3**, **13**, **14**, **15**, **20**, and **25** correspond to sallan A2, ciblan A1, medlan A1, larlan A2, larlan A5, archalan β3, and archalan α, which are purified lanthipeptides encoded by the BGCs labeled in front of them. **c**, The motility assay screening of *H. volcanii* H1424 with empty vector (control) or different lanthipeptide BGCs in 24 well plates with a diameter of 1.56 cm/well. *** $P < 0.001$, ** $P < 0.05$. Each group has three biological replicates. Bars show the mean ± standard deviation (SD), n = 7. **d**, The violin plot of the motility assay of *H. volcanii* H1424 with empty vector, BGCs *cib2* and *pel*, respectively. The median and quartile ranges are shown as horizontal lines in black and green respectively. Black dots are the aligned experiment data, *** $P < 0.001$, n = 21. Each group has six biological repeats. **e**, Phase-contrast images of *H. volcanii* H1424 with empty vector, BGCs *cib2* and *pel*, respectively. Scale bars are 5 μm. Cell circularity (**g**) and frequency distribution (**f**) were calculated from the phase-contrast microscopy of recombinant strains. Each group has three biological replicates. Lines display the means of each group (n = 5,690, 5,553, and 6,264 for control, *cib2*, and *pel* groups, respectively), *** $P < 0.001$. **h**, Volcano plots illustrate the differential expression genes (DEGs) in strains with BGC *cib2* (left) or *pel* (right) compared to the strain with an empty vector. **i**, Venn diagram depicting the overlap DEGs detected in both strains. **j**, A proposed mode for the ecological function of archaeal lanthipeptides in antagonistic interaction and nutrient/space competition.

Discussion

Archaea, which predominantly survive in extreme environments, are reported to have unique metabolic pathways with eco-evolutionary strategies distinct from bacteria and fungi[2,6]. A notable portion of archaeal BGCs consists of RiPPs and terpenes[26], as identified *in silico*, contrasting with the prevalent NRPs and PKs BGCs in bacteria[45]. However, given the lack of identified SMs, our understanding of archaeal chemical language, remains markedly limited compared to our comprehensive knowledge of bacteria and fungi, particularly regarding their biosynthesis and ecological roles[46,47]. While RiPPs exhibit a high diversity and widespread distribution in archaea, class II lanthipeptides have been found to exist exclusively in halophilic archaea[26-28]. This observation implies archaeal lanthipeptides to be niche-specific metabolites adapted to hypersaline environments, as evidenced by the amino acid composition bias and phylogenetic analysis of the hallmark enzyme LanMs and their corresponding precursor LanAs in this study. We then employed a combined strategy involving the biosynthetic rule-based metabolomic analysis and the first use of heterologous expression of archaeal BGCs in Haloarchaea to unveil the chemical landscape of archaeal lanthipeptides. The heterologous expression approach substantially expands the repertoire of archaeal SMs and encourages further exploration of other unique and silent BGCs from the domain of Archaea, such as diverse and widely distributed YcaO/rSAM-modified RiPPs and NRPs. Our discovery of diverse lanthipeptides from archaea expedites the discovery and biosynthesis study of archaeal chemical languages, laying the stage for their chemical ecology study.

The discovery of a new lanthipeptide subfamily, DADC lanthipeptides, underscores the distinct characteristics of archaeal metabolic pathways and enzymes. These lanthipeptides are marked by the unprecedented diamino-dicarboxylic termini, resulting from the unknown cleavage of amino acids within the thioether rings. We propose that DADC lanthipeptide undergoes a multiple-step maturation

process, initially forming classical lanthipeptides and then cleaving amino acids within the thioether rings through a generic protease outside the BGCs. This hypothesis is corroborated by the findings from the in vivo reconstitution of BGCs *lar* and *alnβ* and their corresponding natural products. The precise biosynthetic pathways and enzymes of DADC lanthipeptides remain elusive due to the challenges encountered in the enzymatic study of Haloarchaea. Nonetheless, our research sheds light on the complexity and novelty of archaeal SMs, emphasizing the untapped potential of archaea in natural products and enzyme discovery.

To gain first insight into the ecological function of archaeal lanthipeptides, we assessed in vitro antimicrobial tests and in vivo motility assay. Their specific antagonism against haloarchaea may stem from the prevalence of haloarchaea over normal bacteria in halophilic habitats[48]. However, it is important to note that this study, with a panel of limited indicators, potentially overlooks antagonistic interactions with other archaeal or bacterial species. Subsequently, we carried out the motility regulation assay, which is believed to contribute to nutrient acquisition, reproductive efficiency, antibiotic resistance, and evasion of predation[23,24,43]. The discovery that lanthipeptide BGCs *cib2* and *pel* were involved in activating motility reveals a previously unknown ecological function of RiPPs. This newfound role may assist the host organism in adapting to environmental stimuli. Notably, our microscopic and transcriptomic analyses provide evidence that lanthipeptides enhanced the host's motility by inducing a shift in cell shape towards a rod-like morphology and upregulating the expression of archaellin gene *flgA1*. These findings underscore the potential role of lanthipeptides in influencing two crucial factors for archaeal motility[23,24,44], despite our limited understanding of regulation mechanisms due to the difficulties in genetically manipulating the wild-type host.

In conclusion, this study presents the first heterologous expression of archaeal BGCs to reveal the chemical landscape of lanthipeptides in archaea. Furthermore, we have demonstrated a newfound ecological function of archaeal lanthipeptides in regulating in vivo motility, with the mechanism involving the transformation of rod-shaped cell morphology and the upregulation of archaellin expression. Our discovery of biofunctional lanthipeptides from Archaea provides new insights into their ecological functions, which are anticipated to spur research into the less-explored field of archaeal chemical biology and chemical ecology. These findings not only signify a major advancement in decoding the archaeal chemical language but also highlight the potential of archaea as a rich source of bioactive SMs and unique enzymes.

Methods

BGC analysis

Using antiSMASH bacterial version 7.0[33], archaeal BGCs were identified from the archaeal genomes available on NCBI RefSeq and Genbank databases (accessed in Aug. 2023). The genomes were initially deduplicated based on 99% average nucleotide identity using dRep v3.4.5[49]. The completeness and contamination of genomes were assessed using CheckM2[50], and genomes that met the quality requirements (completeness >80% and contamination <5%) were selected for further analysis. Representative genomes were then selected prioritizing higher assembly levels and genome size. The BGC class was determined based on the product annotation provided by antiSMASH. BGC features

were extracted using BiG-SLiCE v2.0.0 with default settings[51]. The similarity of archaeal BGCs to 2,502 BGCs from MiBiG 3.1[34] was quantified by calculating the minimal cosine distance. Pairwise cosine distances between archaeal BGCs or between archaeal BGCs and MiBiG BGCs were computed using the SciPy library in Python 3.8[52]. Based on those cosine distances, hierarchical clustering with average linkage was adopted for clustering BGCs into GCCs at a distance cutoff of 0.8 using Scikit-learnv1.0.2[53]. The heatmap showing BGC distribution was plotted using *Complex Heatmap* package in R 4.1.2[54]. The BGC synteny was analyzed by Clinker[55] with default parameters.

Sequence similarity network analysis of archaeal LanAs

The SSN analysis was conducted as in our previous study[26]. Archaeal class II lanthipeptide BGCs were extracted according to the product annotation from antiSMASH. The precursor peptides of class II lanthipeptide, LanAs, were predicted from the ten open reading frames adjacent to *lanM*, with lengths ranging from 10 to 100 amino acids and containing Cys and Thr/Ser residues within 30 amino acids of the C-terminus. A total of 354 LanAs was submitted to the EFI - Enzyme Similarity Tool[56] with an *E*-value of 1.0×10^{-1} and an Alignment Score Threshold of 7. Sequences with 100% identity were grouped into a single node, leading to 196 unique sequences. The resulting SSN was visualized using Cytoscape 4.1[57].

Phylogenomic and amino acid composition analysis of archaeal LanAs and LanMs

A total of 12,398 bacterial and 110 archaeal LanMs (Supplementary Data 2) were firstly deduplicated at 80% similarity using MMseqs[58] with the following parameters: -c 0.8 --min-seq-id 0.8 -e 10 --cluster-reassign 1. Alignments were then performed using MUSCLE[59] and visualized with FastTree[60] using default parameters. LanA sequences were aligned using MAFFT v7[61] with parameter: --maxiterate 1000. The phylogenomic trees were conducted by IQ-TREE v2[62] with the following parameters: -m LG+G4 -B 1000 -T 4. The visualization of phylogenetic trees of archaeal LanAs and LanMs was done using iTOL v6[63]. The co-phylogenetic analysis of archaeal LanAs and LanMs was performed using Phytools 2.0[64]. To compare the compositional bias of each amino acid in archaeal and bacterial lanthipeptides at protein and substrate levels, the amino acid composition of LanAs and LanMs was calculated through an in-house Python v.3.12.7 script (https://github.com/GAOYingHKU/Binomial-test-Z-score/) and plotted with ggplot2 (v.3.5.1)[36,65]. *Z*-scores were calculated relative to archaeal LanAs and LanMs, where $Z > 2.0$ indicated significant enrichment of a given amino acid in archaea ('Over-represented'), $Z < -2.0$ indicated significant depletion ('Under-represented'), and some amino acids showed no significant biases ('NS').

Cultivation of haloarchaea strains and extraction of lanthipeptide metabolites

All the wild-type strains listed in Supplementary Table 1 were cultured in modified DSMZ medium 1520 (NaCl 250.0 g L^{-1}, KCl 3.0 g L^{-1}, trisodium citrate 3.0 g L^{-1}, $MgSO_4 \cdot 7\ H_2O$ 20.0 g L^{-1}, yeast extract (Difco) 1.0 g L^{-1}, casamino acids (Difco) 7.5 g L^{-1}, glucose 1.0 g L^{-1}, Na-succinate 10.0 g L^{-1}, sucrose 20.0 g L^{-1}) on a rotary shaker (37 °C and 200 r.p.m.). After cultivating for 2 days, 10% cell culture was inoculated to a large volume fermentation at 37 °C and 200 r.p.m. for another 5-7 days. The recombinant strains were cultured in complete Hv-YPC medium (yeast extract 5 g L^{-1}, peptone 1.0 g L^{-1}, casamino acids 1.0 g L^{-1}, NaCl 160.0 g L^{-1}, $MgCl_2 \cdot 6H_2O$ 20.0 g L^{-1}, $MgSO_4 \cdot 7H_2O$ 23.0 g L^{-}

473 [1], KCl 4.7 g L^{-1}, 13.3 mM pH 7.5 Tris-HCl and other 3 mM $CaCl_2$ adding after cooling) or selectivity
474 Hv-Ca medium (casamino acids 17.0 g L^{-1}, NaCl 160.0 g L^{-1}, $MgCl_2 \cdot 6H_2O$ 20.0 g L^{-1}, $MgSO_4 \cdot 7H_2O$
475 23.0 g L^{-1}, KCl 4.7 g L^{-1}, 13.3 mM pH 7.5 Tris-HCl, and other 3 mM $CaCl_2$, thiamine 0.9 μg L^{-1} and
476 biotin 0.1 μg L^{-1} adding after cooling) as previously reported[37,38]. The single colonies were initially
477 cultivated in Hv-Ca medium at 45 °C for 24 h, and then 5 mL of the suspension was inoculated to a
478 250 mL flask containing 80 mL Hv-YPC medium. Culture broth of each strain was collected at -80 °C
479 after 3 - 7 days of fermentation. For lanthipeptide extraction, 1 mL culture broth was extracted by
480 Diaion® HP-20 resin (Sigma-Aldrich), followed by washing three times with 1 mL of ddH_2O and
481 eluting with 0.5 mL MeOH. The resulting extracts were dried by a vacuum concentrator, redissolved
482 in MeOH, and analyzed by HR LC-MS as detailed in Supplementary Information.

483 **Heterologous expression and reconstitution of lanthipeptide BGCs in *H. volcanii* H1424**

484 Lanthipeptide BGCs were amplified from the genomic DNA of respective wild-type strains using
485 primers designed with appropriate overhangs (Supplementary Table 3). The corresponding PCR
486 products were inserted into the plasmid pTA1228 via Gibson assembly. The recombinant vectors were
487 first propagated in *E. coli* DH5α (DE3) and then introduced into *H. volcanii* H1424 via PEG-mediated
488 transformation approach guided by the Handbook[66]. The recombinant *H. volcanii* H1424 strains were
489 selected in Hv-Ca agar plates and single colonies were grown in Hv-YPC liquid medium for 2 days
490 (45 °C and 200 r.p.m.).

491 **Structure elucidation of archaeal lanthipeptides**

492 ^{1}H, ^{13}C, ^{1}H-^{1}H COSY, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC, TOCSY, and NOESY NMR spectra for medlan
493 A1 (**13**), larlan A2 (**14**), larlan A5 (**15**) and archalan β3 (**20**) were recorded on Avance DRX 600 FT-
494 NMR spectrometer (600 and 150 MHz for ^{1}H and ^{13}C NMR, respectively) using DMSO-d_6. Chemical
495 shifts were reported using the DMSO-d_6 resonance as the internal standard for ^{1}H-NMR DMSO-d_6: δ
496 = 2.50 p.p.m. and ^{13}C-NMR DMSO-d_6: δ = 39.6 p.p.m. Sallan A2 (**2**) and ciblan A1 (**3**) used MeOH-
497 d_4 as the solvent and the internal standard were referenced to: ^{1}H-NMR MeOH-d_4: δ = 3.31, 4.87 p.p.m.
498 and ^{13}C-NMR MeOH-d_4: δ = 49.0 p.p.m.

499 **Motility assay**

500 The motility screening for recombinant strains was set up according to previously reported
501 methodology[23,24]. Briefly, 2 μL of mid-log phase strain cultures (OD_{600} = 0.1, 50 μL in 96-well plates
502 of Corning, Thermo Scientific Varioskan Flash) were stabbed into the plates containing modified Hv-
503 Ca + 0.4% agar motility medium supplemented with additional trace elements (per liter: 5.0 g EDTA,
504 0.8 g $FeCl_3$, 0.05 g $ZnCl_2$, 0.01 g $CuCl_2$, 0.01 g $CoCl_2$, 0.01 g H_3BO_3, 1.6 g $MnCl_2$, 0.01 g Ni_2SO_4, and
505 0.01 g H_2MoO_4). The plates were incubated in a sealed plastic bag at 45 ℃ for 1 day. The time-course
506 motility assays were then conducted at 45 ℃ for 2 - 4 days. The containers used were 6/12/24 - well
507 plates (NEST, China), with approximately 8/4/2 mL of medium added per well, respectively.

Light microscopy

Phase-contrast microscopy was performed using a 100x oil immersion objective lens (Zeiss, LSM980 system, Germany) to examine the cell morphology of three recombinant strains (containing the empty vector pTA1228, pTA1228- BGC *cib2* or pTA1228- BGC *pel*) after 15 hours of cultivation in a shaker (45 °C, 200 r.p.m). Specifically, 1 μL of each strain was placed onto slides coated with a 1% agarose pad containing 18% BSW, and a coverslip was used to cap the samples. Each group has three biological replicates to ensure consistency. ImageJ software[67] was used to distinguish and measure the perimeter and area values of different cells (Supplementary Data 4). The circularity values were calculated using the formula:

$$Circularity = \frac{Perimeter^2}{4\pi \times Area}$$

For the tolerance of ImageJ, the circularity values were filtered no greater than 3.0 and no less than 1.0. The frequency of circularity values was then visualized using GraphPad Prism 8 software.

Transcriptomic analysis

The cell cultures of recombinant strains carrying the empty plasmid pTA1228, pTA1228-BGC *cib2* or pTA1228-BGC *pel* in Hv-Ca medium were harvested after 20 hours of incubation (45°C and 200 r.p.m.). Each group has three biological replicates. The total RNA of all the samples was extracted using RNAprep Pure Cell/Bacteria Kit (Tiangen Biotech, Beijing, China) according to the manufacturer's instructions. RNA qualification was examined by 1% agarose gel electrophoresis and Aligent 5400 (Agilent Technologies, Santa Clara, CA, USA). RNA libraries were constructed using the Fast RNA-seq Lib Prep Kit V2 (ABclonal, Cat. RK20306) and sequenced with Illumina nova X Plus. Sequence reads were trimmed for adaptor sequence, low-quality sequence and rRNA sequence using kneaddata v0.12.0 (https://github.com/biobakery/kneaddata) with parameters: -db SILVA_128_LSUParc_SSUParc_ribosomal_RNA --trimmomatic-options ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:4:20 MINLEN:60 --run-trim-repetitive --reorder --bowtie2-options '--very-sensitive --dovetail' --bypass-trf. Trimmed sequence reads were mapped to the reference genome sequence of *H. volcanii* (GCF 010692905.1) along with GTF annotations using Bowtie 2.3.5.1 with default parameters[68]. The differential expression genes (DEGs; $p < 0.05$, absolute regulation fold change ≥ 2) between control and experiment groups were analyzed with DESeq2 (version 1.38.3)[69]. The DEGs were depicted through the utilization of volcano plots using the EnhancedVolcano in R 4.2.1.

Ethics approval and Consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

All genomes used in this research were obtained from the NCBI Assembly RefSeq database (https://www.ncbi.nlm.nih.gov/assembly) and Genebank database (https://www.ncbi.nlm.nih.gov/genbank/). The HRMS data generated in this study have been deposited in MassIVE under the accession code MSV000095989. The transcriptomic data of this study has been deposited in NCBI with the accession number GSE281021. Data supporting the results of this study is available in the paper, as well as supplementary information and supplementary data.

Code availability

The code used in this study for amino acid composition bias analysis of LanAs and LanMs assigned a citable DOI through Zenodo (DOI10.5281/zenodo.13705935) and can be accessed at https://github.com/GAOYingHKU/Binomial-test-Z-score/.

Competing interests

The authors declare no competing interests.

Funding

This work is partially funded by a Shenzhen Basic Research General Programme (JCYJ20210324122211031), the Research Grants Council of Hong Kong (27107320 and 17115322), and the Hong Kong Branch of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (SMSEGL20SC02) to Y.-X.L. The funding sources had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

Author contributions

Z.-M.S. and Y.-X.L. designed the research and prepared the manuscript. Z.-M.S., C.C., H.L., J.Z. and Q.Z. performed research. Z.-M.S. and Y.-X.L. characterized the structures of compounds. Z.-M.S. and C.C. performed the LC-MS analysis. Z.-M.S., Y.G., X.L., D.Z., and G.W. did the genomic analysis. Y.G. analyzed the transcriptomic data. Z.-M.S., Y.Q., and H.J. performed the phase-contrast microscopic analysis. Z.-M.S., C.C., Y.G., X.L., Y.Q., D.Z., G.W., H.L., Q.Z., J.Z., P.Y. C., H.J. and Y.-X.L. analyzed data and discussed the results. W.L. and Y.-X.L. founded the research. Y.-X.L. supervised the research. All authors read and approved the final manuscript.

Acknowledgments

The authors would like to express their gratitude to Prof. Thorsten Allers (School of Life Sciences, University of Nottingham, UK) and Prof. Likui Zhang (College of Environmental Science and Engineering, Yangzhou University, China) for their generous sharing of the plasmid pTA1228, host strain *H. volcanii* H1424 and corresponding protocol. Thank Prof. Shihui Dong (College of Chemistry and Chemical Engineering, Lanzhou University, China) for kindly providing MeLan and Lan standard samples and Yi-Man Eva FUNG, Jo Yip, and Bonnie Yan from the Department of Chemistry, University of Hong Kong for their help in MS and NMR analysis. Thank the microscopic imaging service at the Centre for PanorOmic Sciences (University of Hong Kong) and Zhuohan Li (Department of Chemistry, University of Hong Kong) for their help in microscopic analysis. Some figures were created with BioRender.com.

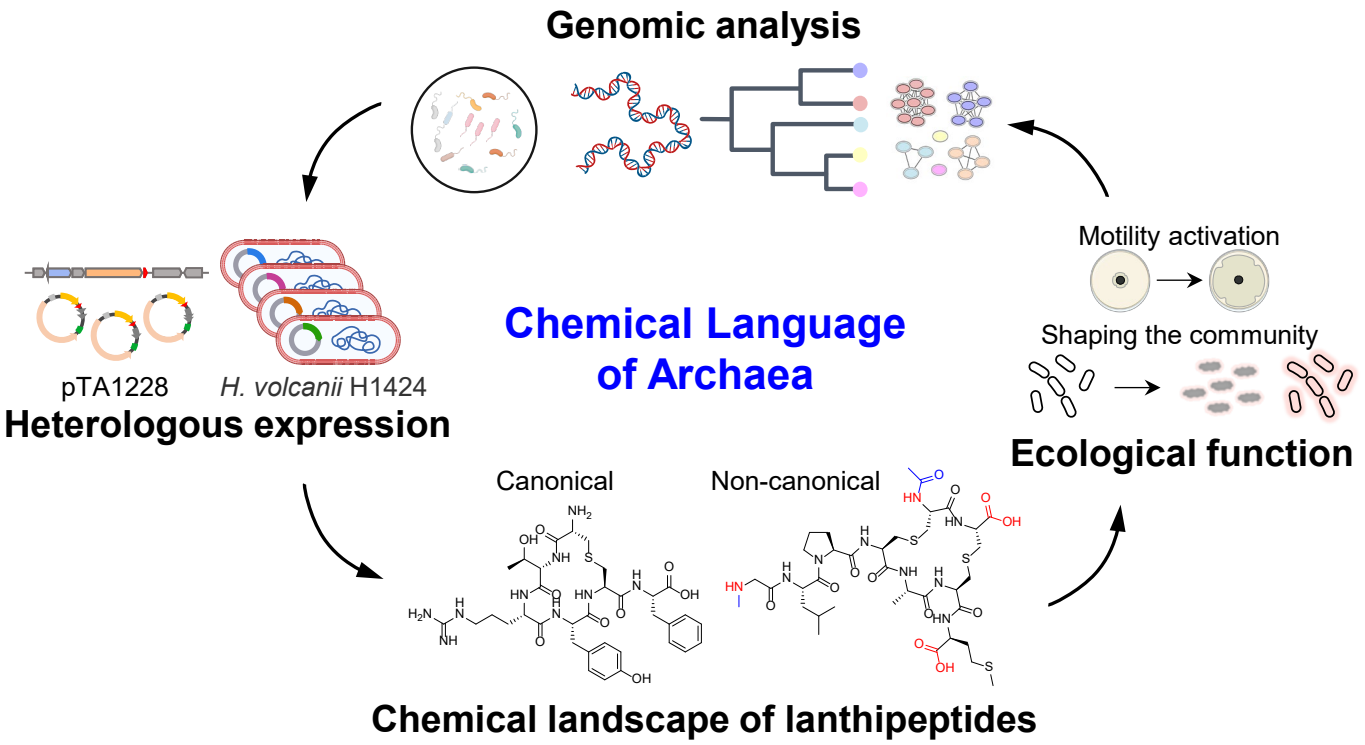
References

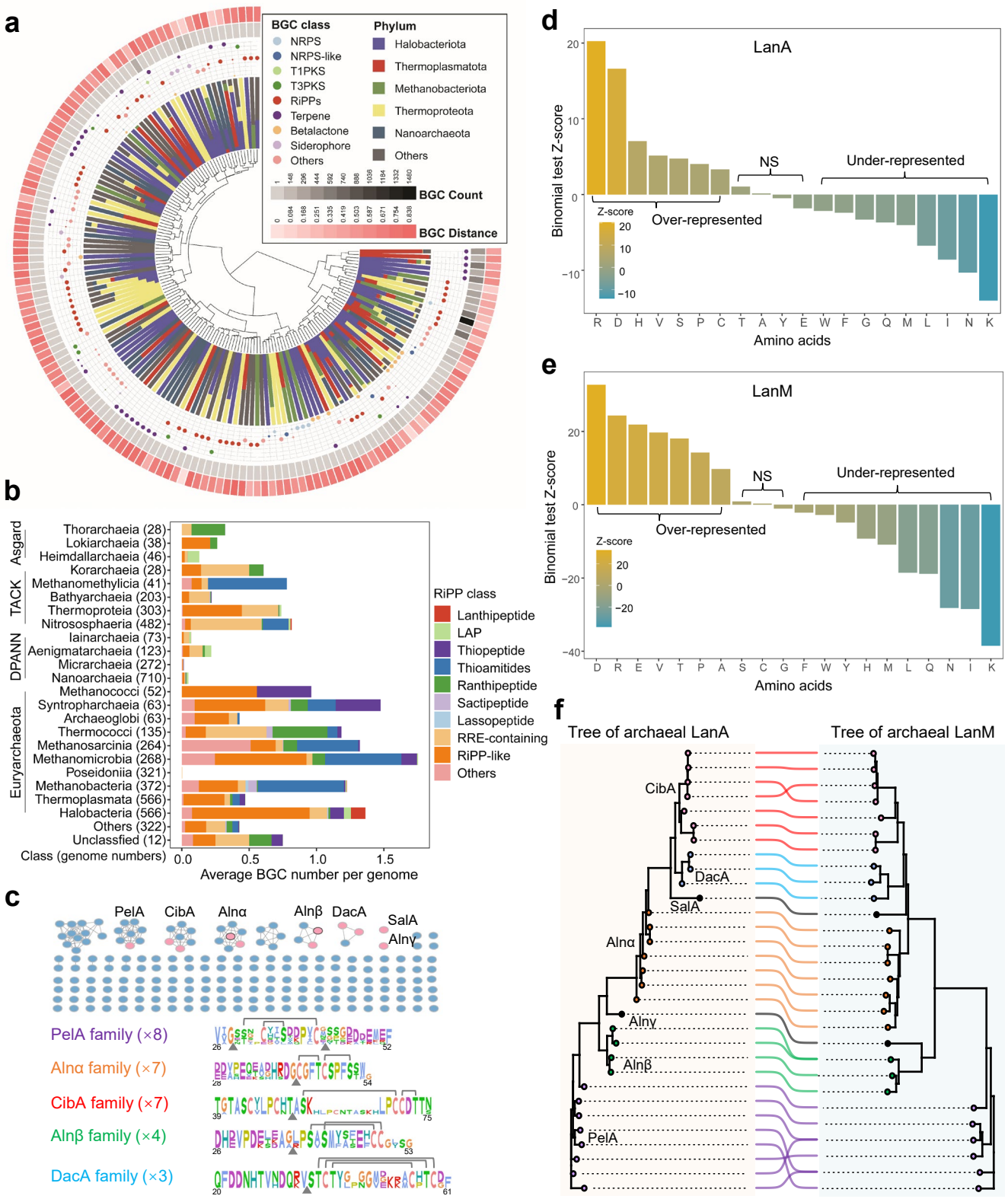
1. Woese, C.R., Kandler, O. & Wheelis, M.L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**, 4576-4579 (1990).
2. Baker, B.J. et al. Diversity, ecology and evolution of Archaea. *Nat Microbiol* **5**, 887-900 (2020).
3. Moissl-Eichinger, C. et al. Archaea are interactive components of complex microbiomes. *Trends Microbiol* **26**, 70-85 (2018).
4. Crits-Christoph, A. et al. Functional interactions of archaea, bacteria and viruses in a hypersaline endolithic community. *Environ Microbiol* **18**, 2064-2077 (2016).
5. Weiland-Bräuer, N. Symbiotic interactions of archaea in animal and human microbiomes. *Curr Clin Microbiol* **10**, 161-173 (2023).
6. Dong, Y. et al. Eco-evolutionary strategies for relieving carbon limitation under salt stress differ across microbial clades. *Nat Commun* **15**, 6013 (2024).
7. Coker, O.O., Wu, W.K.K., Wong, S.H., Sung, J.J.Y. & Yu, J. Altered gut archaea composition and interaction with bacteria are associated with colorectal cancer. *Gastroenterology* **159**, 1459-1470 e1455 (2020).
8. Xiong, L. et al. A new type of DNA phosphorothioation-based antiviral system in archaea. *Nat Commun* **10**, 1688 (2019).
9. Schonknecht, G. et al. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* **339**, 1207-1210 (2013).
10. Oesterhelt, D. & Stoeckenius, W. Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nat New Biol* **233**, 149-152 (1971).
11. Niessen, N. & Soppa, J. Regulated iron siderophore production of the halophilic archaeon *Haloferax volcanii*. *Biomolecules* **10** (2020).
12. Zhang, G. et al. Acyl homoserine lactone-based quorum sensing in a methanogenic archaeon. *ISME J* **6**, 1336-1344 (2012).
13. van Wolferen, M., Orell, A. & Albers, S.V. Archaeal biofilm formation. *Nat Rev Microbiol* **16**, 699-713 (2018).
14. Greening, C. et al. Minimal and hybrid hydrogenases are active from archaea. *Cell* **187**, 3357-3372 e3319 (2024).
15. Lloyd, C.T. et al. Discovery, structure and mechanism of a tetraether lipid synthase. *Nature* **609**, 197-203 (2022).
16. Liu, A. et al. Functional elucidation of TfuA in peptide backbone thioamidation. *Nat Chem Biol* **17**, 585-592 (2021).
17. Nayak, D.D., Mahanta, N., Mitchell, D.A. & Metcalf, W.W. Post-translational thioamidation of methyl-coenzyme M reductase, a key enzyme in methanogenic and methanotrophic Archaea. *Elife* **6**, e29218 (2017).
18. Garcia, A.A., Chadwick, G.L., Liu, X.L. & Welander, P.V. Identification of two archaeal GDGT lipid-modifying proteins reveals diverse microbes capable of GMGT biosynthesis and modification. *Proc Natl Acad Sci U S A* **121**, e2318761121 (2024).
19. Wu, K. et al. Isolation of a methyl-reducing methanogen outside the Euryarchaeota. *Nature* **632**, 1124-1130 (2024).
20. Tang, Y. et al. The β-d-manno-heptoses are immune agonists across kingdoms. *Science* **385**, 678-684 (2024).
21. Kumar, V. et al. Halocins, natural antimicrobials of Archaea: Exotic or special or both? *Biotechnol*

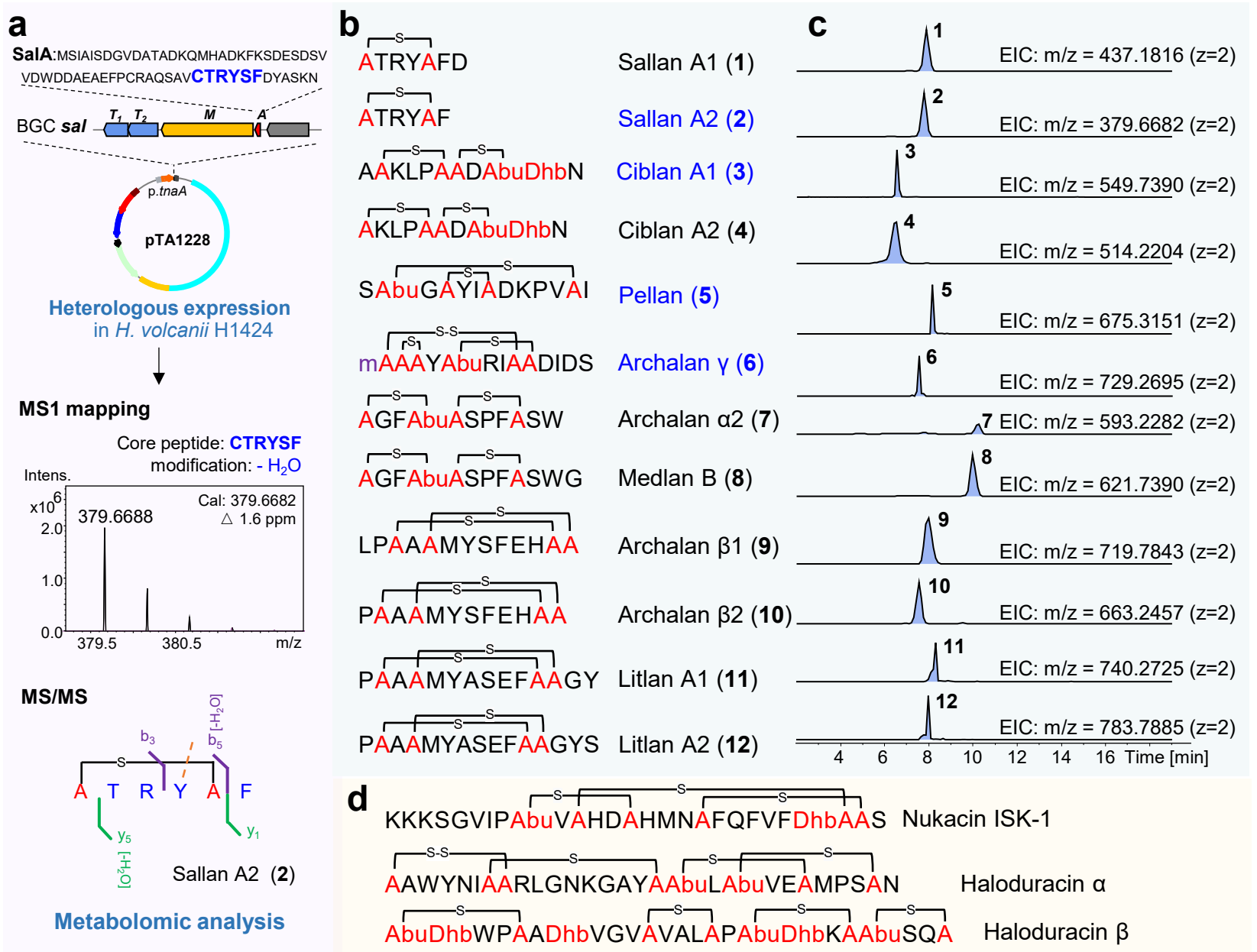
Adv **53**, 107834 (2021).
22. Makarova, K.S. et al. Antimicrobial peptides, polymorphic toxins, and self-nonself recognition systems in archaea: an untapped armory for intermicrobial conflicts. *mBio* **10**, e00715-19 (2019).
23. Schiller, H. et al. Identification of structural and regulatory cell-shape determinants in *Haloferax volcanii*. *Nat Commun* **15**, 1414 (2024).
24. Duggin, I.G. et al. CetZ tubulin-like proteins control archaeal cell shape. *Nature* **519**, 362-365 (2015).
25. Tang, S.K. et al. Cellular differentiation into hyphae and spores in halophilic archaea. *Nat Commun* **14**, 1827 (2023).
26. Liang, H. et al. Genomic and metabolic analyses reveal antagonistic lanthipeptides in archaea. *Microbiome* **11**, 74 (2023).
27. Castro, I. et al. The lanthipeptide biosynthetic clusters of the domain Archaea. *Microbiol Res* **253**, 126884 (2021).
28. Walker, M.C. et al. Precursor peptide-targeted mining of more than one hundred thousand genomes expands the lanthipeptide natural product family. *BMC Genomics* **21**, 387 (2020).
29. Scott, T.A. et al. Widespread microbial utilization of ribosomal β-amino acid-containing peptides and proteins. *Chem* **8**, 2659-2677 (2022).
30. Zweerink, S. et al. Activity-based protein profiling as a robust method for enzyme identification and screening in extremophilic Archaea. *Nat Commun* **8**, 15352 (2017).
31. Sugiyama, R. et al. The biosynthetic landscape of triceptides reveals radical sam enzymes that catalyze cyclophane formation on Tyr- and His-containing motifs. *J Am Chem Soc* **144**, 11580-11593 (2022).
32. Cheng, Z. et al. Rule-based omics mining reveals antimicrobial macrocyclic peptides against drug-resistant clinical isolates. *Nat Commun* **15**, 4901 (2024).
33. Blin, K. et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res* **51**, W46-W50 (2023).
34. Terlouw, B.R. et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res* **51**, D603-D610 (2023).
35. Repka, L.M., Chekan, J.R., Nair, S.K. & van der Donk, W.A. Mechanistic understanding of lanthipeptide biosynthetic enzymes. *Chem Rev* **117**, 5457-5520 (2017).
36. Baker, B.A. et al. Expanded phylogeny of extremely halophilic archaea shows multiple independent adaptations to hypersaline environments. *Nat Microbiol* **9**, 964-975 (2024).
37. Haque, R.U., Paradisi, F. & Allers, T. *Haloferax volcanii* for biotechnology applications: challenges, current state and perspectives. *Appl Microbiol Biotechnol* **104**, 1371-1382 (2020).
38. Brendel, J. et al. A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-derived RNAs (crRNAs) in *Haloferax volcanii*. *J Biol Chem* **289**, 7164-7177 (2014).
39. Wang, M. et al. Genome mining of sulfonated lanthipeptides reveals unique cyclic peptide sulfotransferases. *Acta Pharm Sin B* **14**, 2773-2785 (2024).
40. Luo, Y., Xu, S., Frerk, A.M. & van der Donk, W.A. Facile method for determining lanthipeptide stereochemistry. *Anal Chem* **96**, 1767-1773 (2024).
41. Gilchrist, C.L.M. & Chooi, Y.H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473-2475 (2021).
42. Kavitha, P.K., Lipton, A., A R, S. & Aishwarya, M.s. Growth characteristics and halocin production by a new isolate, *Haloferax volcanii* KPS1 from Kovalam solar saltern (India). *Research*

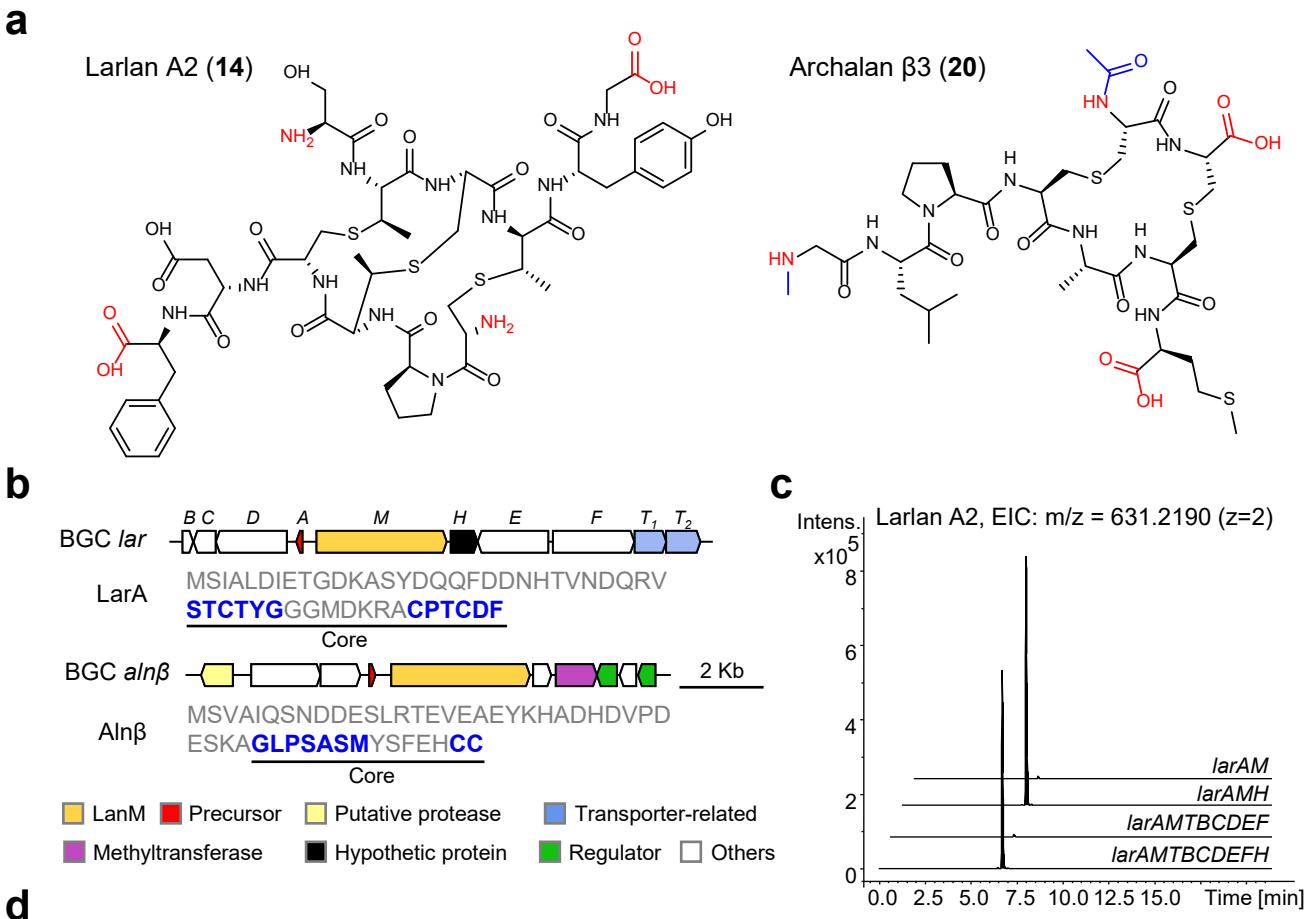
Journal of Biological Sciences **6**, 257-262 (2011).
43. Quax, T.E.F. et al. Structure and function of the archaeal response regulator CheY. *P Natl Acad Sci USA* **115**, E1259-E1268 (2018).
44. Tripepi, M., Imam, S. & Pohlschroder, M. *Haloferax volcanii* flagella are required for motility but are not involved in PibD-dependent surface adhesion. *J Bacteriol* **192**, 3093-3102 (2010).
45. Paoli, L. et al. Biosynthetic potential of the global ocean microbiome. *Nature* **607**, 111-118 (2022).
46. Hudson, G.A. & Mitchell, D.A. RiPP antibiotics: biosynthesis and engineering potential. *Curr Opin Microbiol* **45**, 61-69 (2018).
47. Ongpipattanakul, C. et al. Mechanism of action of ribosomally synthesized and post-translationally modified peptides. *Chem Rev* **122**, 14722-14814 (2022).
48. Shu, W.S. & Huang, L.N. Microbial diversity in extreme environments. *Nat Rev Microbiol* **20**, 219-235 (2022).
49. Olm, M.R., Brown, C.T., Brooks, B. & Banfield, J.F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* **11**, 2864-2868 (2017).
50. Chklovski, A., Parks, D.H., Woodcroft, B.J. & Tyson, G.W. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* **20**, 1203-1212 (2023).
51. Kautsar, S.A., van der Hooft, J.J.J., de Ridder, D. & Medema, M.H. BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. *Gigascience* **10**, giaa154 (2021).
52. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261-272 (2020).
53. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research* **12**, 2825-2830.
54. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).
55. van den Belt, M. et al. CAGECAT: The CompArative GEne Cluster Analysis Toolbox for rapid search and visualisation of homologous gene clusters. *BMC Bioinformatics* **24**, 181 (2023).
56. Zallot, R., Oberg, N. & Gerlt, J.A. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **58**, 4169-4182 (2019).
57. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).
58. Hauser, M., Steinegger, M. & Soding, J. MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics* **32**, 1323-1330 (2016).
59. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).
60. Price, M.N., Dehal, P.S. & Arkin, A.P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
61. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066 (2002).
62. Minh, B.Q. et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530-1534 (2020).
63. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* **52**, W78-W82 (2024).

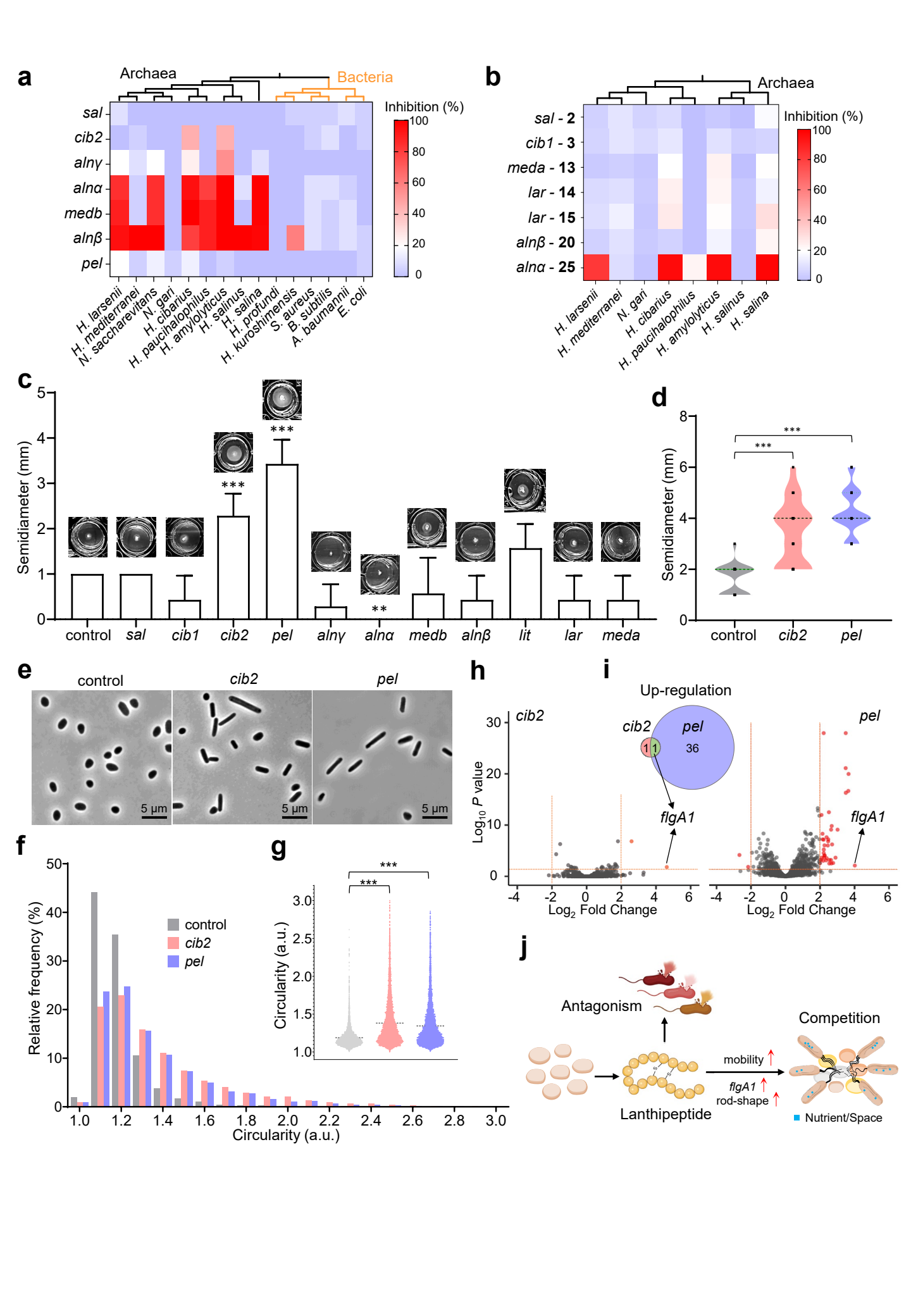
723 64. Revell, L.J. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and
724 other things). *PeerJ* **12**, e16505 (2024).
725 65. Wickham, H. ggplot2: elegant graphics for data analysis. (Springer-Verlag New York, 2016).
726 66. Dyall-Smith, M. The Halohandbook: protocols for haloarchaeal genetics. *Haloarchaeal Genetics*
727 *Laboratory, Melbourne* **14** (2008).
728 67. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**,
729 676-682 (2012).
730 68. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-
731 359 (2012).
732 69. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-
733 seq data with DESeq2. *Genome Biol* **15**, 550 (2014).


Genomic analysis
Chemical Language of Archaea
pTA1228
H. volcanii H1424
Heterologous expression
Motility activation
Shaping the community
Ecological function
Canonical
Non-canonical
Chemical landscape of lanthipeptides


a
BGC class
NRPS
NRPS-like
T1PKS
T3PKS
RiPPs
Terpene
Betalactone
Siderophore
Others
Phylum
Halobacteriota
Thermoplasmatota
Methanobacteriota
Thermoproteota
Nanoarchaeota
Others
BGC Count
BGC Distance
b
Asgard
TACK
DPANN
Euryarchaeota
Thorarchaeia (28)
Lokiarchaeia (38)
Heimdallarchaeia (46)
Korarchaeia (28)
Methanomethylicia (41)
Bathyarchaeia (203)
Thermoproteia (303)
Nitrososphaeria (482)
Iainarchaeia (73)
Aenigmatarchaeia (123)
Micrarchaeia (272)
Nanoarchaeia (710)
Methanococci (52)
Syntrophorchaeia (63)
Archaeoglobi (63)
Thermococci (135)
Methanosarcinia (264)
Methanomicrobia (268)
Poseidoniia (321)
Methanobacteria (372)
Thermoplasmata (566)
Halobacteria (566)
Others (322)
Unclassfied (12)
Class (genome numbers)
0.0
0.5
1.0
1.5
Average BGC number per genome
RiPP class
Lanthipeptide
LAP
Thiopeptide
Thioamitides
Ranthipeptide
Sactipeptide
Lassopeptide
RRE-containing
RiPP-like
Others
c
PelA
CibA
Alnα
Alnβ
DacA
SalA
Alnγ
PelA family (×8)
Alnα family (×7)
CibA family (×7)
Alnβ family (×4)
DacA family (×3)
d
LanA
Binomial test Z-score
Z-score
Over-represented
NS
Under-represented
Amino acids
R D H V S P C T A Y E W F G Q M L I N K
e
LanM
Binomial test Z-score
Z-score
Over-represented
NS
Under-represented
Amino acids
D R E V T P A S C G F W Y H M L Q N I K
f
Tree of archaeal LanA
Tree of archaeal LanM
CibA
DacA
SalA
Alnα
Alnγ
Alnβ
PelA


a
SalA:MSIAISDGVDATADKQMHADKFKSDESDSV
VDWDDAEAEFPCRAQSAVCTRYSFDYASKN
BGC sal
T1 T2 M A
p.tnaA
pTA1228
Heterologous expression
in H. volcanii H1424
MS1 mapping
Core peptide: CTRYSF
modification: - H2O
Intens.
x10^6
379.6688
Cal: 379.6682
△ 1.6 ppm
2.0
1.0
0.0
379.5
380.5
m/z
MS/MS
b3
b5 [-H2O]
A T R Y A F
y5
y1
[-H2O]
Sallan A2 (2)
Metabolomic analysis
b
ATRYAFD
Sallan A1 (1)
ATRYAF
Sallan A2 (2)
AAKLPAADAbuDhbN
Ciblan A1 (3)
AKLPAADAbuDhbN
Ciblan A2 (4)
SAbuGAYIADKPVAI
Pellan (5)
mAAAYAbuRIAADIDS
Archalan γ (6)
AGFAbuASPFASW
Archalan α2 (7)
AGFAbuASPFASWG
Medlan B (8)
LPAAAMYSFEHAA
Archalan β1 (9)
PAAAMYSFEHAA
Archalan β2 (10)
PAAAMYASEFAAGY
Litlan A1 (11)
PAAAMYASEFAAGYS
Litlan A2 (12)
c
1 EIC: m/z = 437.1816 (z=2)
2 EIC: m/z = 379.6682 (z=2)
3 EIC: m/z = 549.7390 (z=2)
4 EIC: m/z = 514.2204 (z=2)
5 EIC: m/z = 675.3151 (z=2)
6 EIC: m/z = 729.2695 (z=2)
7 EIC: m/z = 593.2282 (z=2)
8 EIC: m/z = 621.7390 (z=2)
9 EIC: m/z = 719.7843 (z=2)
10 EIC: m/z = 663.2457 (z=2)
11 EIC: m/z = 740.2725 (z=2)
12 EIC: m/z = 783.7885 (z=2)
4 6 8 10 12 14 16 Time [min]
d
KKKSGVIPAbuVAHDAHMNAFQFVFDhbAAS Nukacin ISK-1
AAWYNIAARLGNKGAYAAbuLAbuVEAMPSAN Haloduracin α
AbuDhbWPAADhbVGVAVALAPAbuDhbKAAbuSQA Haloduracin β


a
Larlan A2 (14)
Archalan β3 (20)
b
BGC lar
B C D A M H E F T1 T2
LarA
MSIALDIETGDKASYDQQFDDNHTVNDQRV
STCTYGGGMDKRACPTCDF
Core
BGC alnβ
2 Kb
Alnβ
MSVAIQSNDDESLRTEVEAEYKHADHDVPD
ESKAGLPSASMYSFEHCC
Core
LanM
Precursor
Putative protease
Transporter-related
Methyltransferase
Hypothetic protein
Regulator
Others
c
Intens.
x10^5
Larlan A2, EIC: m/z = 631.2190 (z=2)
larAM
larAMH
larAMTBCDEF
larAMTBCDEFH
0.0 2.5 5.0 7.5 10.0 12.5 15.0 Time [min]
d
Leader
LanM
Export
Hydrolysis
Hydrolysis


a
Archaea
Bacteria
Inhibition (%)
sal
cib2
alnγ
alnα
medb
alnβ
pel
H. larsenii
H. mediterranei
N. saccharevitans
N. gari
H. cibarius
H. paucihalophilus
H. amylolyticus
H. salinus
H. salina
H. profundi
H. kuroshimensis
S. aureus
B. subtilis
A. baumannii
E. coli
b
Archaea
Inhibition (%)
sal - 2
cib1 - 3
meda - 13
lar - 14
lar - 15
alnβ - 20
alnα - 25
c
Semidiameter (mm)
control
sal
cib1
cib2
pel
alnγ
alnα
medb
alnβ
lit
lar
meda
d
Semidiameter (mm)
control
cib2
pel
e
control
cib2
pel
5 μm
f
Relative frequency (%)
control
cib2
pel
Circularity (a.u.)
g
Circularity (a.u.)
h
cib2
pel
Log10 P value
Log2 Fold Change
i
Up-regulation
cib2
pel
11
36
flgA1
j
Antagonism
Competition
Lanthipeptide
mobility
flgA1
rod-shape
Nutrient/Space