Deciphering the chemical landscape and potential ecological function of RiPPs from the untapped Archaea domain

Zhi-Man Song[1,2], Cunlei Cai[1,2], Ying Gao[1], Xiaoqian Lin[1], Gengfan Wu[1], Haoyu Liang[1], Qianlin Zhuo[1], Junliang Zhang[3], Peiyan Cai[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

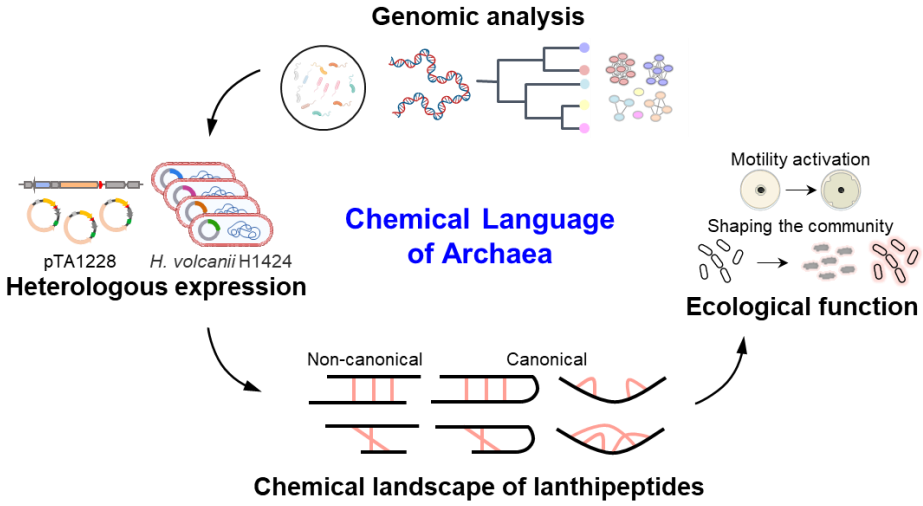
[3] School of Pharmacy, Shanghai Jiao Tong University, Shanghai 200240, China

[*]To whom correspondence may be addressed.

Email: yxpli@hku.hk

Abstract

Chemical communication is crucial in ecosystems with complex microbial assemblages. However, due to archaeal cultivation challenges, our understanding of the structure diversity and function of secondary metabolites (SMs) within archaeal communities is limited compared to the extensively studied and well-documented bacterial counterparts. Our comprehensive investigation into the biosynthetic potential of archaea, combined with metabolic analyses and heterologous expression, has unveiled the previously unexplored biosynthetic capabilities and chemical diversity of Archaeal SMs. This represents the first reported instance of heterologous expression of archaeal SMs. 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. The maturation process of this noncanonical DADC lanthipeptide deviates from classical ones by cleaving amino acids within the thioether rings. This sheds light on the biosynthetic novelty of archaeal metabolites and emphasizes their potential as an untapped resource for SMs. Interestingly, archaeal lanthipeptides exhibit haloarchaea-specific antagonistic activity and enhance the host's motility, facilitating the archaeal interaction with biotic and abiotic environments. Our discoveries broaden our understanding of archaeal chemical language and provide promising prospects for future exploration of SM-mediated interactions in this domain.



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 and often difficult to cultivate 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 [3-8]. These interactions extend to symbiosis, predation, antagonism, mutualism, and participation in essential processes like carbon, nitrogen, and sulfur cyclings[3-7]. Similar to their prokaryotic counterparts, archaea were recently reported to employ UV-protective carotenoids[9] and putative siderophores[10] for abiotic interactions. However, chemical-mediated biotic interactions are much less understood in archaea than in bacteria.

Recent studies suggest archaea may employ quorum-sensing molecules[11,12], alongside antagonistic proteins such as archaeocins[13,14] and peptides like lanthipeptide[15] to mediate chemical interactions in complex ecological niches. Moreover, archaea can use a variety of enzymes and metabolic pathways distinct from bacteria[2,16,17], representing a largely untapped source for discovering novel enzymes and metabolites. Nonetheless, due to the challenges in archaeal cultivation and genetic manipulation[15,18], the chemical landscape, biosynthesis, and ecological function of archaea metabolites remain largely exclusive[2,19,20]. By investigating the unexplored biosynthetic potential of secondary metabolites (SMs) from archaea, we aimed to gain valuable insights into their chemical diversity and chemical-mediated interactions. This exploration could reveal new bioactive SMs from archaea, invigorating the relatively uncharted field of archaeal chemical biology.

Haloarchaea, conveniently cultured in a laboratory setting, serve as ideal models for studying archaeal interactions, evolution, cell biology, and metabolic pathways[4,8,21,22]. Our previous research, based on fermentation of five wild-type strains, identified only two archaeal lanthipeptides from one strain, despite genome mining revealing the presence of thousands of biosynthetic gene clusters (BGCs) across different archaeal strains[15,18]. Heterologous expression of natural products is an efficient, powerful method for revealing silent BGCs and investigating their biosynthesis in bacteria and fungi. However, expressing archaeal biosynthetic pathways in bacteria hosts can be problematic due to distinct metabolic features, potentially leading to improper folding, low expression levels, or inactive archaeal enzymes[23,24]. The scarcity of SMs from even well-defined BGCs in archaea, alongside the challenges involved in their heterologous expression in model bacteria hosts, underscores the necessity for new strategies in discovering archaeal metabolites. Thus, it is critical to establish an archaeal heterologous expression system for natural products to reveal their chemical diversity, biological functions, and ecological significance.

Here, our study reports the heterologous expression of BGCs in archaea to expedite the discovery of archaeal chemical languages. We first systematically analyzed archaeal post-translationally modified peptide (RiPP) BGCs, the most prevalent family of archaeal SMs, aiming to uncover their biosynthetic potential, chemical landscape, and ecological functions. Our finding revealed that class II lanthipeptides found in haloarchaea are evolutionarily different from bacteria, potentially for adapting to hypersaline environments. To gain a deeper insight into the lanthipeptides, we successfully validated over twenty-four classic and noncanonical lanthipeptides originating from eleven BGCs by combining

the heterologous expression with biosynthetic rule-based metabolomic analysis[25]. Remarkably, in the maturation of these noncanonical lanthipeptides, the cleavage with core peptide region results in the opening of thioether rings, leading to the unprecedented lanthipeptide subfamily with diamino and dicarboxylic termini. The identified archaeal lanthipeptides demonstrated a narrow anti-archaeal spectrum specifically against various strains from Haloarchaea, suggesting their potential role as chemical weaponry in shaping microbial communities in hypersaline environments. Intriguingly, we discovered a new ecological function of lanthipeptides that modulated the motility of the host strain, *Haloferax volcanii* H1424, which could potentially enhance environmental adaptation by improving nutrient uptake or space occupation. This study significantly broadens our comprehension of archaeal chemical languages and opens promising avenues for future investigations on SM-mediated archaeal interactions.

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 more extensively researched bacterial domain[15]. They encode a broad range of known secondary metabolite classes, such as RiPPs, terpenes, nonribosomal peptides (NRPs), polyketides (PKs), and other metabolites. Our initial fermentation-based discovery identified a new lanthipeptide from haloarchaea, marking the first discovery of a lantibiotic and anti-archaeal secondary metabolite in the archaea domain[15]. To investigate the biosynthetic landscape of RiPPs, our current study further utilized the antiSMASH 7.0 tool[26] that is capable of analyzing both bacterial and archaeal BGCs, to 11,644 archaeal genomes from the NCBI database to delve deeper into exploring the diversity and uniqueness of these RiPPs. Our findings revealed that RiPP BGCs exhibited a wide variety and novelty across different subfamilies, such as YcaO-related RiPPs (thioamitides and thiopeptides), radical S-adenosylmethionine-dependent enzyme-modified RiPPs (ranthipeptides and sactipeptides), lanthipeptides, and lassopeptides (Fig. 1a and Supplementary Fig.1).

While other RiPP families are widely distributed throughout different phyla/classes, class II lanthipeptides are exclusively in the Haloarchaea (refer to the class Halobacteria), implying a niche-specific metabolite (Fig. 1a and Supplementary Data). However, 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 BGCs in the wild-type strain or through heterologous expression were challenging, prompting a shift in focus towards lanthipeptide. This decision was informed by the successful production of lanthipeptides in our prior research. Consequently, an updated library containing 353 putative precursors (LanAs) and 110 lanthipeptide synthase LanMs from 102 archaeal lanthipeptide BGCs is constructed in silico through antiSMASH-based genome mining method[15,26] (Supplementary Data) to decipher the chemical landscape of archaeal lanthipeptides. Within the library, a total of 67 deduplicated archaeal LanMs that are exclusively harbored 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). The phylogenetic tree analysis of archaeal LanMs demonstrated their evolutionary separation from bacterial LanMs (Supplementary Fig. 1 and Supplementary Data). The maximum likelihood tree of archaeal LanAs exhibited two major clades,

one including our previously characterized LanAs, Alnα and Alnβ, while the other clade remains to be explored (Supplementary Fig. 2). Additionally, in the library, a total of 196 unique LanA sequences have been categorized into seven families through sequence similarity network (SSN) analysis (Fig. 1b), highlighting untapped chemical space for lanthipeptide discovery.

Co-evolved archaeal LanMs and LanAs exhibit environmental adaptation

Our collective analysis of the amino acid biases, phylogenetic tree, SSN, and sequence logo indicates a unique chemical distinctiveness of archaeal lanthipeptides as compared to their bacterial counterparts (Fig. 1 and Supplementary Fig. 2). Archaeal LanAs are predicted to be shorter and comprise a higher portion of C and S instead of T residues than bacterial LanAs (Fig. 1b, c and Supplementary Fig. 2), which are crucial conserved residues for thioether ring formation and cleavage sites in bacterial lanthipeptide biosynthetic pathways[27]. More importantly, the distinctive amino acid composition of archaeal LanAs and LanMs supports the hypothesis that they underwent eco-evolution in response to hypersaline environments (Fig. 1c, d). 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. 1c, d). The increased abundance of D/E residues and the reduced portion of K residue are linked to the "salt-in" strategy of haloarchaea, enabling interactions with cations in their high-salinity environment[8,28].

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 is a unique evolutionary trait of archaeal lanthipeptide biosynthetic pathways that has not been observed in other protein analyses to the best of our knowledge[28]. Considering the evolutionary divergence of LanM and all amino acid 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. 1e and Supplementary Fig. 1). 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 that are exclusively harbored in Haloarchaea, thereby playing significant ecological roles in halophilic environments.

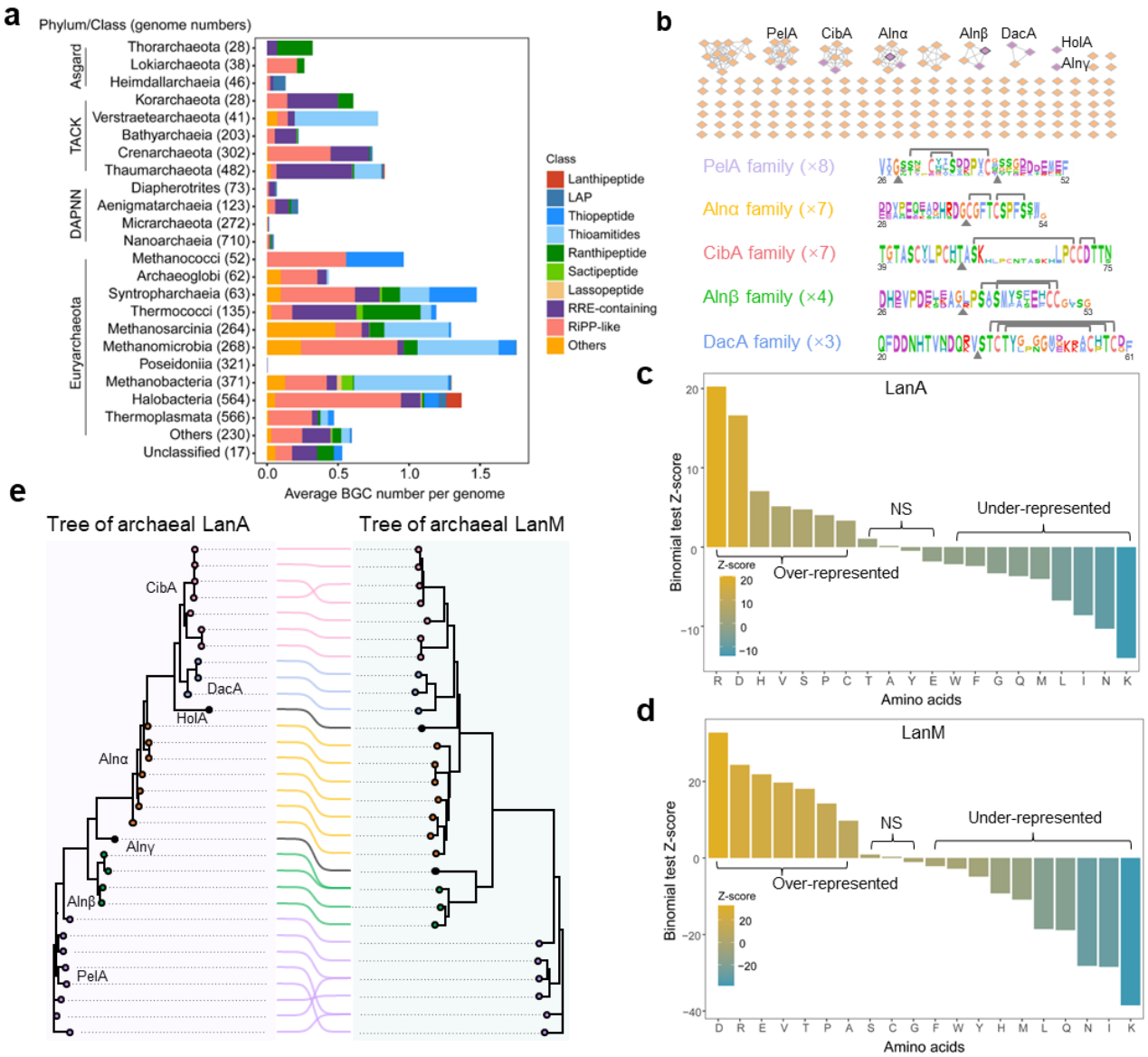


Figure 1. Genomic and protein amino acid compositional biases analysis of archaeal LanAs and LanMs. a, The distribution of RiPP BGCs in the domain of Archaea at the phylum or class level. **b**, The SSN analysis of putative archaeal LanAs and sequence logos of experimentally verified LanA families. Sequences with 100% identity are grouped into a single node, comprising 196 unique sequences. Black borders indicate two LanAs discovered in our previous study. Purple 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. **c-d**, Binomial tests were conducted to compare the proportions of all 20 amino acids in LanAs (**c**) and LanMs (**d**) between archaeal and bacterial counterparts. *Z*-scores were calculated relative to archaeal LanAs and LanMs, with $Z > 1.96$ indicating significant enrichment of a given amino acid in archaea ('Over-represented'), $Z < -1.96$ indicating significant depletion ('Under-represented'), and some amino acids without significant bias ('NS'). **e**, Co-phylogenomic analysis of LanAs (left) and LanMs (right) from verified SSN families in this study. The nodes and lines are colored according to their LanA families in **b**. Two black notes represent verified orphan notes in the SSN analysis, HolA and Alnβ.

Elucidating classical archaeal lanthipeptides by heterologous expression in *Haloferax volcanii*

Moving beyond the traditional method of natural product discovery through wild-type cultivation, we heterologously expressed eleven BGCs from various SSN families in *H. volcanii* H1424 using the pTA1228 vector[29]. The procedure was initiated by cloning our previously verified lanthipeptide BGCs[15], *alnα* and *alnβ*, which resulted in the production of corresponding analogs of the natural lanthipeptides, archalans α1 (**7**), β1 (**9**) and β2 (**10)** (Fig. 2b, c and Supplementary Figs. 3-4). It marks the initial endeavor to authenticate archaeal natural product BGCs through heterologous expression (Fig. 2), laying the groundwork for the subsequent expression of other archaeal lanthipeptide BGCs. Based on this approach, we successfully heterologously expressed eleven BGCs from the strains stocked in our lab (Supplementary Table 1), covering all seven lanthipeptide families in SSN (Fig.1). This endeavor linked seven BGCs from five major families and two orphan BGCs with their lanthipeptide products, representing the first comprehensive exploration of archaeal lanthipeptides so far (Figs.1 and 2).

Except for the unpredictable lanthipeptides from DcaA family, a total of nine classical lanthipeptides were characterized using biosynthetic rule-based metabolic analysis[25] (Fig. 2 and Supplementary Figs. 5-9), including holsalans (**1** and **2**), ciblans (**3** and **4**), penlan (**5**), archalan γ (**6**), medlan B (**8**) and litlans (**11** and **12**). It's important to note that holsalans (**1** and **2**), ciblan A2 (**4**), pellan (**5**), and litlans (**11** and **12**) were not detected in their corresponding wild-type strains, even with the prior knowledge of them obtained from heterologous expression. Other was detected in relatively low titers compared to the results from heterologous expression. This highlights the advantage of heterologous expression using an archaeal host in revealing and characterizing archaeal lanthipeptides that may not be easily detected in their native organisms.

These classical lanthipeptides **1**-**4** structures 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, Holsalans A1 (**1**, $[M+2H]^{2+}$ = 437.1816) and A2 (**2**, $[M+2H]^{2+}$ = 379.6688) were found to be analogs encoded by BGC *hol* (*Halorussus salinus* YJ-37-H), predicting with one dehydration (-H_2O, -18 Da) on their corresponding core peptides (CTRYSFD for **1** and CTRYSF for **2**, Supplementary Fig. 3). The MS/MS data supported that they all have a C-S crosslink between Cys1 and Ser5 (b_5, -H_2O, -18 Da), but differ at the C-terminus with the presence of an additional D residue in **1** (y_2, 'FD' residue, Supplementary Fig. 3). 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 Figs. 10, 35 and Supplementary Table 3). Ciblan A1 (**3**, BGC *cib*, *Haladaptatus cibarius* JCM19505) and ciblan A2 (**4**, BGC *den*, *H. denitrificans* DSM4425) were predicted with the same modification on the core peptide (ASKLPCCDTTN for A1 and SKLPCCDTTN for A2). 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, supporting 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. 4). 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 key HMBC crosslinks in Ser2-Cys6 (δ_H = 2.97, 2.92 ppm to δ_C = 35.0 ppm) and Cys7-Thr9 (δ_H = 3.19, 2.98 ppm to δ_C = 41.2 ppm) (Supplementary Figs. 10, 36 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. 31 and Supplementary Table 4). It's worth noting that **2** with only six amino acids represents the smallest natural lanthipeptide to date.

Due to their extremely low yield, compounds **5-12** were primarily identified using HRMS and biosynthetic analysis. The MS1 analysis and full reduction-desulfurization product of pellan (**5**, BGC *spel, H. pelagicus* RC-68) confirmed the presence of two dehydrations within residues Thr2, Cys4, Ser7 and Cys12, suggesting interwind or bicycle ring structures in the core peptide (STGCYISDKPVCI, Supplementary Fig. 5). Furthermore, the MS/MS fragments of m/z = 1236.5314 $[M+H]^+$ and m/z = 1073.4679 $[M+H]^+$ with the loss of only 'I' and 'IY' residues within the rings supported the interwind ring structure in Thr2-Cys12 and Cys4-Ser7 instead of bicycle structure in Thr2-Cys4 and Cys12-Ser7 (Supplementary Fig. 5). Archalan γ (**6**, BGC *alnγ*, *H. salinus* YJ-37-H) gave a prominent doubly charged peak (m/z = 729.2666 $[M+2H]^{2+}$) showing the linear <u>DIDS</u> motif in the C-terminus, consistent with the predicted core peptide (CSCYTRICC<u>DIDS</u>, Supplementary Figs. 11-12). Partial desulfurization product and tandem mass analysis indicated one methylation on Cys1, along with residues Cys1, Ser2, Cys3, and Cys8 reduced to be Ala, revealing an intact C-S crosslinking in Thr5-Cys9 (Supplementary Figs. 13-14). Subsequently, the dithiothreitol reduction and MS^n analysis determined the presence of a disulfide bond between Cys1 and Cys8 and a thioether ring between Ser2 and Cys3 (Supplementary Figs. 15-16). Archalan γ (**6)** contained N-terminal methylation, a Cys1-Cys8 crosslink, and thioether crosslinks in Ser2-Cys3 and Thr5-Cys9 (Supplementary Figs. 12-16). This structure expands the number of the smallest thioether ring containing lanthipeptides which only has a few reported cases (such as mersacidin, ltnα, and sacAα)[27,30].

Archalan α1 (**7**, BGC *alnα*, *H. salinus* YJ-37-H) and medlan B (**8**, BGC *medb*, *H. mediterranei* ATCC33500) harbored two thioether rings supporting 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 line with previously reported structure of archalan α[15]. Archalans β1-2 (**9** and **10**, BGC *alnγ*, *H. salinus* YJ-37-H) and litlans A1-2 (**11** and **12,** BGC *lit*, *H. litoreus* JCM31109) harbored similar crosslinks as previously reported archalan β[15] with two interwind thioether rings causing MS/MS fragments of the N/C-terminal overhang amino acids outside of the rings (Supplementary Figs. 8 and 9). These above chemical structures of archaeal lanthipeptides revealed that they are typically shorter than bacterial lanthipeptides, such as nukacin ISK-1, Halα, and Halβ (Fig. 2). Despite being short in length, their cross-linking patterns and modifications are intricate similar to those found in bacterial lanthipeptides, including disulfide bond, methylation, dehydration, hydroxylation, bicycle and interwind ring structures. Moreover, 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.

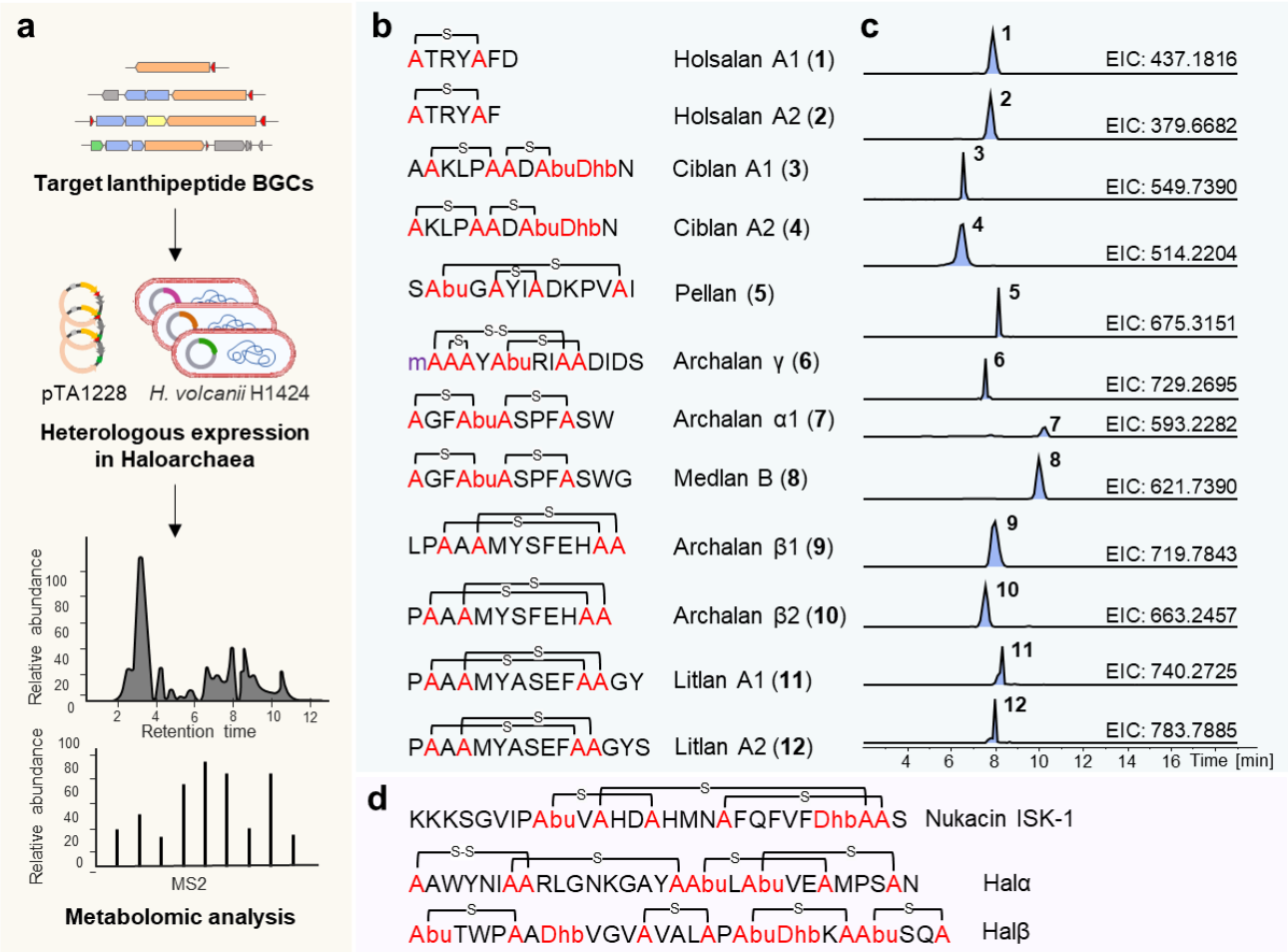


Figure 2. Deciphering novel archaeal class II lanthipeptides through heterologous expression and biosynthetic rule-based metabolic analysis. a, Workflow for heterologous expression eleven lanthipeptide BGCs in *H. volcanii* H1424. Classical lanthipeptides (**b**) with different modifications ('m' in purple represents methylation) were identified via HR LC-MS and NMR analysis. **d**, Representative classical bacterial class II lanthipeptides.

Noncanonical DADC lanthipeptides with diamino-dicarboxylic termini resulting from cleaving amino acids within the thioether rings

Apart from the classical lanthipeptides mentioned above, several unusual peptide signals from BGCs *meda* and *lar* engrossed our attention for the unpredictable modification (Fig. 3 and Supplementary Figs. 17-24). They appeared in the extracts of both wild-type and heterologous expression strains, suggesting that they were likely encoded by these two BGCs, albeit with unexpected modifications as revealed by MS/MS data. While traditional lanthipeptides allow for the accurate prediction of core peptide sequences and putative modifications based on MS analysis, 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 Figs. 18-23). Nonetheless, these data still provided valuable clues to lock their corresponding precursor peptides with these fragmental analyses and instill confidence in the purification process.

To fully characterize their structures, 3.2 mg of **13** (m/z = 415.4769 $[M+3H]^{3+}$), 3.5 mg of **14** (m/z = 631.2194 $[M+2H]^{2+}$), and 2.0 mg of **15** (m/z = 652.2301 $[M+2H]^{2+}$) 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 three thioether bonds. These crosslinked patterns expose diamino-dicarboxylic termini (Supplementary Figs. 10, 37-39 and Supplementary Table 3), leading to the name of DADC lanthipeptides, with certain amino acids within the thioether rings being partially cleaved (Fig. 3 and Supplementary Figs. 18-19). In addition, medlan A2 (**16**), larlans A1 (**17**), A3 (**18**), and A4 (**19**) are deduced to have longer or shorter overhang amino acids in the N/C terminus through MS/MS analysis, possibly caused by the aminopeptidase (Supplementary Figs. 20-23). These unique characteristics classify these lanthipeptides as a novel subfamily.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) for the advanced Marfey's analysis, a clear stereo structure annotation of each MeLan residue was achieved when compared with four MeLan standard samples[31] (Supplementary Fig. 31 and Supplementary Table 4). Ultimately, 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 and Supplementary Fig. 31).

To find whether other lanthipeptides also employ a similar maturation strategy, the LC-MS/MS data of both wild-type and heterologous expression strains were meticulously analyzed. The metabolic profile of *H. salinus* YJ-37-H, harboring six lanthipeptide BGCs, were found to display a set of mass signals with a similar fragmental pattern to larlans and medlans (Supplementary Figs. 25-30). To characterize their structures, 1.2 mg of compound **20** (m/z = 453.6796 $[M+2H]^{2+}$) was obtained for NMR and advanced Marfey's analysis (Fig. 3, Supplementary Figs. 9, 25, 39, and Supplementary Tables 3, 4). 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. 27-30). 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 (Fig. 3a, Supplementary Fig. 31 and Supplementary Table 4).

Identifying this particular lanthipeptide subfamily exclusively in archaea highlights the existence of a unique biosynthetic pathway and enzymes distinct from those in bacteria. Synteny analysis[32] 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. 32 and Supplementary Table 5). The high-yield *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* without peptidase gene (Fig. 3b, c and Supplementary Fig. 32). It suggested the possibility of putative genomic peptidases from the 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 LarH may play a role as a regulator, cofactor, or a yet unidentified peptidase involved in the biosynthesis of larlan A2. In another case, we detected the production of classical lanthipeptide archalans β1-2 but didn't detect DADC lanthipeptide archalans β3-7 through heterologous expression despite including the putative peptidase within the BGC (Supplementary Figs. 8 and 32). Both classical archalan β and DADC lanthipeptide

archalans β3-7 were identified in the wild-type host; however, heterologous expression only yielded classical lanthipeptides archalan β. This supports the hypothesis that the involvement of putative proteases outside BGC in the biosynthesis process, suggesting a more complex biosynthetic pathway for DADC lanthipeptides. This pathway includes additional hydrolysis steps compared to classical lanthipeptides, facilitated by unidentified generic proteases to cleave amino acids within the thioether rings. (Fig. 3d).

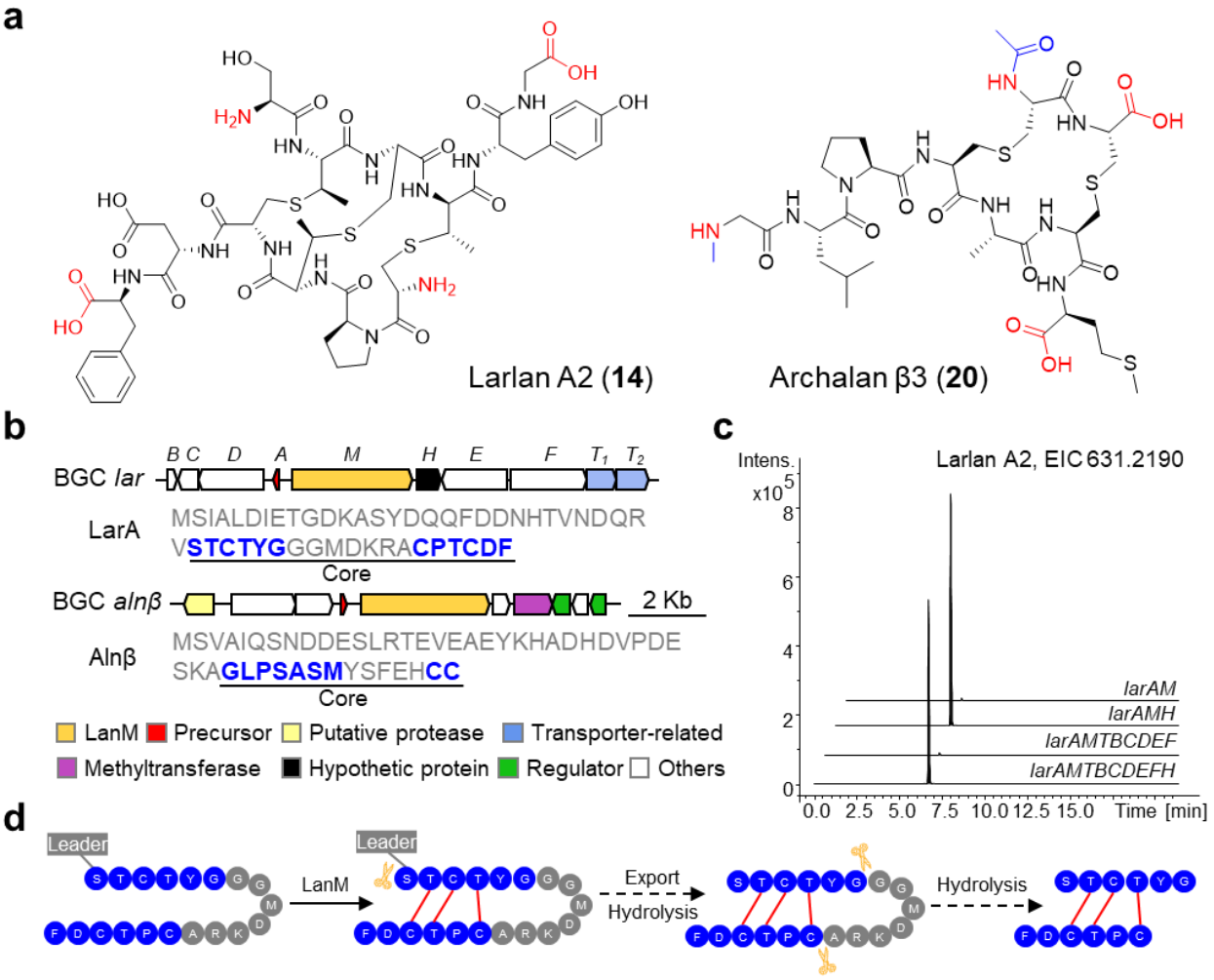


Figure 3. Chemical structure and proposed biosynthetic pathway of DADC lanthipeptides. a, Chemical structure of DADC lanthipeptides. Diamino-dicarboxylic termini are red, and specific modifications in the N-terminus are blue. **b**, BGCs *lar* and *alnβ* with their corresponding precursors. **c**, LC-MS 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

Six lanthipeptide crude extracts obtained from heterologous expression and seven purified archaeal lanthipeptides were collected for the antimicrobial activity screening against nine haloarchaea strains from five different genera and six bacteria strains including two halophilic bacteria (Fig. 4). All the crude extracts from heterologous expression strains were processed through a C18 solid-phase extraction (SPE) cartridge to accumulate the target lanthipeptides. The same gradient fractions of the

empty plasmid-containing strain were utilized for comparison to account for the potential influence of halocins[33,34]. Specifically, extracts from BGCs *medb, den,* and *alnα-γ* exhibited obvious anti-haloarchaeal activity at 100 μg mL^{-1} even though the lanthipeptides in a low proportion (Fig. 4a). One of these purified lanthipeptides, 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}. Notably, the MIC value of archalan α against *H. salina* JCM 18369 was 3.125 μg mL^{-1} (Supplementary Fig. 33). It has a strong inhibition (>50%) against *H. amylolyticus* JCM 18367 and *H. salina* JCM 18369 even at a concentration of 0.78 μg mL^{-1} (Fig. 4b and Supplementary Fig. 33). The narrow spectrum of 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 eleven lanthipeptide BGCs, stab-inoculation screens were conducted to assess their impact on the host's motility, with the empty vector-containing strain as a control. Surprisingly, three heterologous expression strains were found to modulate the host motility as either activator or inhibitor (Fig. 4c). Strains with BGCs *den* and *spel* exhibited a significantly larger motility zone than the control, suggesting a potential linkage of these genes or products to the observed hypermotility phenomenon (Fig. 4c, d). Subsequently, BGC *den* was selected for time-course motility assays to gain a deeper insight, indicating that BGC *den* significantly activated the host's motility zone in the beginning without affecting growth, with all strains displaying the Bull's eye pattern in the late stage (Supplementary Fig. 34). BGC *alnα* inhibited the motility may be attributed to its antagonistic activity (Fig. 4b and Supplementary Fig. 34), which could hinder the growth of host. These findings point to an unknown ecological role of archaeal lanthipeptides, potentially reducing the host's reaction time or increasing that of other strains in response to stimuli[35,36]. This could contribute to environmental adaptation by occupying space, enhancing nutrient acquisition, evading predation, *etc* (Fig. 4e).

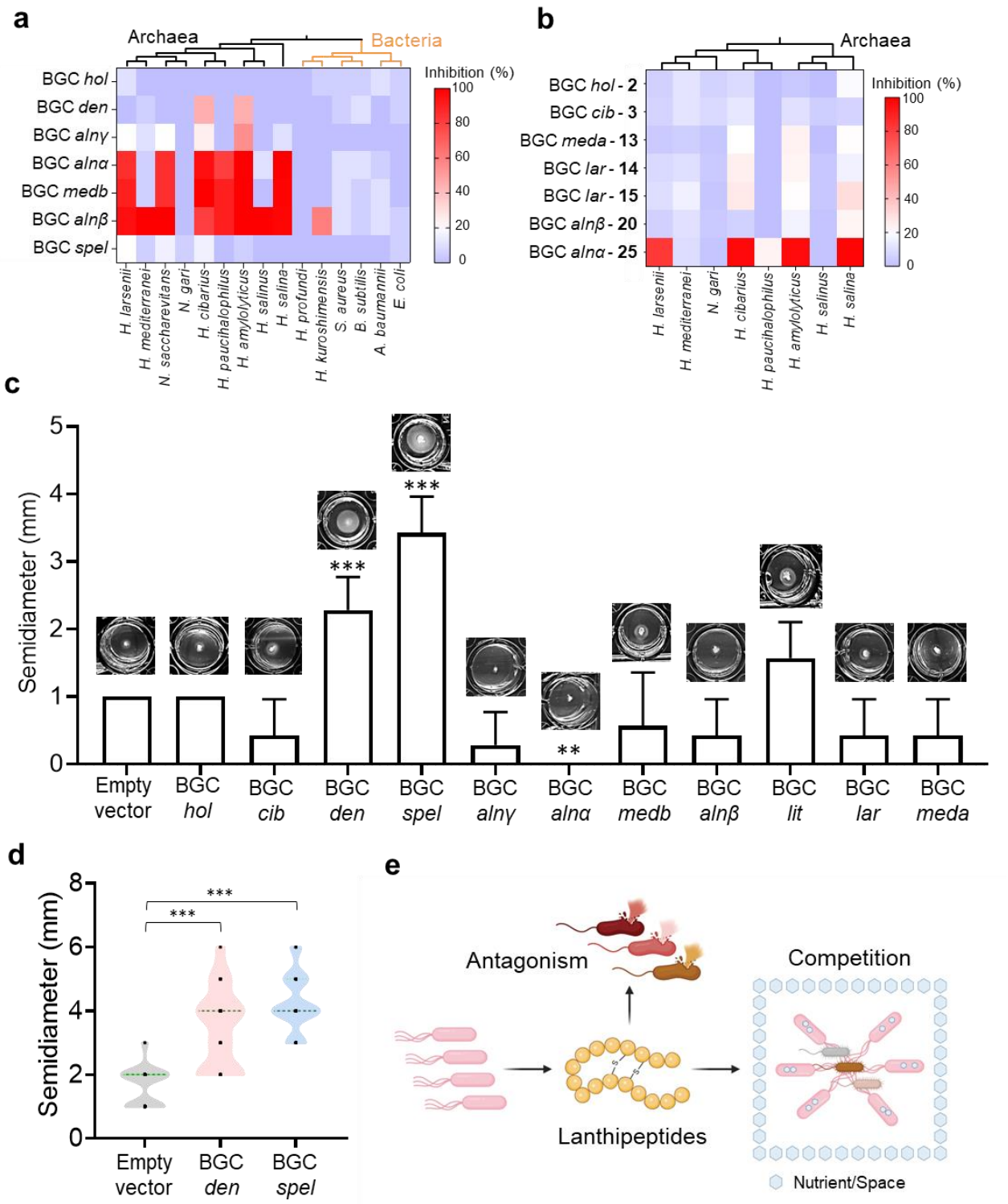


Figure 4. Biological functions of archaeal lanthipeptides. The antimicrobial activity of the fractionated crude extracts of heterologous expression strains (**a**) and purified archaeal lanthipeptides at 100 μg mL^{-1} (**b**). Compounds **2**, **3**, **13**, **14**, **15**, **20**, and **25** correspond to halslan 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 or different lanthipeptide BGCs in 24 well plates with a diameter of 1.56 cm/well. *** $P < 0.001$, ** $P < 0.05$. 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 *den* and *spel*, 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. **e**, 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[2,8]. A notable portion of archaeal BGCs consists of terpenes and RiPPs[15,37], contrasting with the prevalent nonribosomal peptide synthases and polyketide synthases BGCs in bacteria[15,37,38]. While RiPPs exhibit a high diversity and wide distribution in archaea, class II lanthipeptides have been predicted to exist exclusively in halophilic archaea[8,15,18,39-41]. This observation implies archaeal lanthipeptides to be niche-specific metabolites adapted to hypersaline environments, as evidenced by the amino acid biases and phylogenetic analysis of the hallmark enzyme LanMs and their corresponding LanAs in this study. However, lack of identified SMs, our understanding of archaeal chemical language, such as RiPPs, remains markedly limited compared to our comprehensive knowledge of bacteria and fungi, in terms of their biosynthesis and corresponding ecological roles[42-44]. To bridge this knowledge gap, we employed a combined strategy involving the biosynthetic rule-based metabolomic analysis and the first use of the heterologous expression of archaeal BGCs in Haloarchaea to elucidate 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 wildly distributed YcaO/rSAM-modified RiPPs and nonribosomal peptides. Our discovery of diverse lanthipeptides from archaea expedites the discovery and biosynthesis study of archaeal chemical languages, laying the stage for their functional study in vivo.

The unique characteristics of archaeal metabolic pathways and enzymes are highlighted by the discovery of a new lanthipeptide subfamily, DADC lanthipeptides. Despite the long discovery history of lanthipeptide, DADC lanthipeptides with diamino-dicarboxylic termini resulting from the cleavage of amino acids within the thioether rings haven't been identified before. We propose that DADC lanthipeptide undergoes a multiple-step maturation process, to form classical lanthipeptide first and then cleave amino acids within the thioether rings by the generic protease located outside the BGCs. This hypothesis is supported 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 our inability to obtain the fully modified precursor peptides and 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 product and enzyme discovery and in the pursuit of understanding microorganism interactions.

To look first insight into the ecological function of archaeal lanthipeptides, we assessed their in vitro antimicrobial test and in vivo motility assay. Their specific antagonism against haloarchaea may stem from the prevalence of haloarchaea over traditional bacteria in halophilic habitats[5,8,45-47]. However, it

is important to note that this study, with a panel of limited indicators, potentially overlooks antagonistic interactions with other species of archaea or bacteria. This narrow focus may also limit the breadth of insights relevant to other ecological interactions in diverse microbial communities. Motility activity is believed to enhance nutrient acquisition, reproductive efficiency, antibiotic resistance, and evasion of predation[21,22,48]. The discovery that BGCs *den*, *spel*, and *alnα* are involved in motility regulation activities reveals a previously unknown ecological function of RiPPs. This role could potentially assist the host in its environmental adaptation. However, due to the difficulties in genetically manipulating the wild-type host, the in vivo functions of lanthipeptides and their mechanisms remain unclear. Nonetheless, our discovery of bioactive archaeal lanthipeptides provides new insights into the ecological function of archaeal metabolites, which are anticipated to spur research into the less-explored field of archaeal chemical biology and chemical ecology.

In conclusion, this study presents the first heterologous expression of archaeal BGCs by exploring the landscape and biosynthesis of lanthipeptides in Haloarchaea, revealing a new lanthipeptide subfamily characterized by unforeseen diamino-dicarboxylic termini that may require additional hydrolysis steps than classical lanthipeptides for amino acid cleavage within the thioether rings. Furthermore, we have demonstrated a new ecological function of archaeal lanthipeptides in regulating in vivo motility, shedding light on their role in resource competition, in addition to their involvement in in vitro antagonistic interaction. These findings not only reveal a major advancement in deciphering the archaeal chemical language but also highlight the potential of archaea as a rich source of bioactive SMs and unique enzymes, promising exciting prospects for future research in this field.

Methods

General materials and experimental procedures

Biochemicals and media components for bacterial cultures were purchased from standard commercial sources. All strains used in this study are listed in Supplementary Table 1. 1D and 2D NMR spectra were recorded at 298K on a Bruker Avance DRX 600 FT-NMR spectrometer (600 and 150 MHz for ^{1}H and ^{13}C NMR, respectively). High-resolution liquid chromatography-mass spectrometry (HR LC-MS) analyses were performed on UltiMate 3000 UHPLC Systems with a Waters Acquity UPLC BEH C18 column (1.7 μm, 300 Å, 2.1 × 150 mm) coupled to Bruker impact™ II Mass Spectrometer except for the heterologous expression analysis of DADC lanthipeptides with 100 mm column. The column was maintained at 40 °C and run at a 0.2 mL min $^{-1}$ flow rate, using 0.1% formic acid in H_2O as solvent A and 0.1% formic acid in ACN as solvent B. A gradient was employed for chromatographic separation starting at 5% B for 2 minutes, then 5% to 95% B for 15 minutes, washed with 95% for 4 minutes, and finally held at 5% B for 1 minute. All the samples were analyzed in positive polarity, using data-dependent acquisition mode. The HR LC-MS data were analyzed with Bruker Compass DataAnalysis 4.3. Low-resolution (LR) LC-MS analyses were performed on the Waters ACQUITY H-Class UPLC system coupled with an ACQUITY SQ detector 2 mass spectrometer. Separations were carried out on the same column and methods as HR LC-MS unless otherwise stated.

Cultivation and extraction

All the wild-type strains listed in Table 1 were cultured in 5 mL of modified DSMZ medium 1520 (composition: 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 g L^{-1}, Sucrose 20 g L^{-1}) on a rotary shaker (37 °C and 200 r.p.m.). For host strain, *H. volcanii* H1424, were cultured in complete Hv-YPC (composition: yeast extract 5 g L^{-1}, peptone 1 g L^{-1}, casamino acids 1 g L^{-1}, NaCl 160 g L^{-1}, $MgCl_2 \cdot 6H_2O$ 20 g L^{-1}, $MgSO_4 \cdot 7H_2O$ 23 g L^{-1}, KCl 4.7 g L^{-1}, 13.3 mM pH 7.5 Tris-HCl and other 3mM $CaCl_2$ adding after cooling) or selectivity Hv-Ca medium (composition: casamino acids 17 g L^{-1}, NaCl 160 g L^{-1}, $MgCl_2 \cdot 6H_2O$ 20 g L^{-1}, $MgSO_4 \cdot 7H_2O$ 23 g L^{-1}, KCl 4.7 g L^{-1}, 13.3 mM pH 7.5 Tris-HCl, and other 3mM $CaCl_2$, Thiamine 0.89 μg L^{-1} and Biotin 0.11 μg L^{-1} adding after cooling) as previously report. After cultivating at 45 °C for 2 days, 5 mL of the suspension was transferred into a 250 mL flask containing 80 mL medium. After 3, 5, and 7 days of fermentation, 20 mL of each culture was collected at -80 °C. For extraction, 1 mL culture broth of each was extracted by Diaion® HP-20 resin (Sigma-Aldrich), followed by washing three times with 1mL of H_2O and eluting with 0.5 mL MeOH. After being dried by a vacuum concentrator, all crude extracts were redissolved in methanol and subjected to HR LC-MS.

Sequence similarity network analysis of archaeal LanAs and BGC similarity analysis

The network process is similar to our previously reported lanthipeptides[15]. Briefly, archaeal genomes were downloaded from NCBI RefSeq and Genbank databases (accessed in Aug. 2023) and analyzed by antiSMASH 7.0 with default parameters[49]. Class II lanthipeptide BGCs were extracted according to the product annotation from antiSMASH. The precursor peptides of class II lanthipeptide were predicted from the ten open reading frames adjacent to *lanM* with lengths between 10 - 100 amino acids and Cys and Thr/Ser within 30 amino acids of the C-terminus. A total of 354 LanAs was submitted to the EFI - Enzyme Similarity Tool[50] with an *E*-value of 1.0×10^{-1} and an Alignment Score Threshold of 7. Sequences with 100% identity are grouped into a single node, comprising 196 unique sequences. The resulting SSN was visualized using Cytoscape 4.1[51]. The BGC similarity was analyzed by Clinker[52] with default parameters.

Phylogenomic analysis of archaeal LanAs and LanMs

Bacterial and archaeal LanMs were firstly deduplicated with 80% similarity using MMseqs[53] with the following parameters: -c 0.8 --min-seq-id 0.8 -e 10 --cluster-reassign 1. Alignments were then performed using MUSCLE[54] and submitted to FastTree[55] with default parameters for protein alignment. LanAs were aligned using MAFFT version 7[56] with default parameters and --maxiterate 1000. The phylogenomic trees were conducted by IQ-TREE version 2[57] with the following parameters: -m LG+G4 -B 1000 -T 4. Phytools 2.0 was used for the co-phylogenetic analysis of LanAs and LanMs[58]. iTOL v6 was used for visualization[59].

Amino acid composition analysis

To compare the compositional bias of each amino acid between archaeal and bacterial lanthipeptides in protein and substrate levels, the amino acid composition of LanAs and LanMs was calculated through an in-house Python v.3.10.5 script (https://github.com/GAOYingHKU/Binomial-test-Z-score/)

461 and analyzed using principal component analysis (PCA) with ggplot2 (v.3.4.2)[28,60]. *Z*-scores were
462 calculated relative to archaeal LanAs and LanMs, with $Z > 1.96$ indicating significant enrichment of a
463 given amino acid in archaea ('Over-represented'), $Z < -1.96$ indicating significant depletion ('Under-
464 represented'), and some amino acids without significant bias ('NS').

465 **Reconstitution of lanthipeptide BGCs in *H. volcanii* H1424**

466 Lanthipeptide BGCs were amplified from the genomic DNA of corresponding wild-type strains with
467 appropriate overhangs in the primer (Supplementary Table 3). Different versions with different gene
468 deletions were designed to determine the gene functions within the *med* and *lar* BGCs. The
469 corresponding PCR products were inserted into the plasmid pTA1228 via Gibson assembly. The
470 resulting clones were propagated in *E. coli* DH5α (DE3) before transforming in *H. volcanii* H1424.
471 The recombinant *H. volcanii* H1424 cells were selected in Hv-Ca agar plates and grown in Hv-YPC
472 liquid medium for 3, 5, and 7 days (45 °C and 200 r.p.m.). The product extraction and detection
473 processes are as described above. For the detection of noncanonical lanthipeptides, another UHPLC
474 column, Waters ACQUITY Premier Peptide BEH C18 Column (1.7 μm, 300 Å, 2.1 × 100 mm), was
475 used for the HR LC-MS analysis for the saturation of former one.

476 **Production, identification, and purification of archaeal lanthipeptides**

477 Archalan β3

478 The fermentation and production of archalan β3 were performed as previously reported [15]. A total of
479 50 L culture broth of *H. salinus* was collected after cultivation for 5 days at 37 °C and extracted by 10%
480 HP-20 resin (v/v) overnight. The rest wash and elution steps were the same extraction methods as
481 above. The crude extract evaporated and passed through an open ODS (Octadecylsilane) column. This
482 column was then eluted with a gradient of MeOH-H_2O (10 - 50% at intervals of 10% and 100%). The
483 target compound was observed in the fraction 40% MeOH by LC-MS analysis. Following this, the
484 fraction was evaporated and loaded to a Sephadex LH-20 column (Solarbio, China) in a solvent
485 mixture of 60% MeOH/H_2O system to separate and concentrate archalan β3 and its analogues. Those
486 fractions with target compound were merged, evaporated, and subjected to an HPLC system (Waters,
487 Parsippany, NJ, United States) on Phenomenex Kinetex XB-C18 column (250 mm × 10 mm, 5 μm,
488 100 Å) with 32% MeOH/ H_2O containing constant 0.05% trifluoroacetic acid (TFA) at the flow rate
489 of 4 mL min^{-1}. archalan β3 (~1 mg) was appeared at 40.0 - 44.0 min.

490 Holsalan A1-2

491 The holsalans were only produced by the heterologous expression in *H. volcanii* H1424 at 45 °C. The
492 recombinant H1424 culture broth was collected and extracted by HP20 as above. Holsalans were eluted
493 by 16% ACN (+ 0.05% TFA) isocratic elution on the HPLC system with a Phenomenex Kinetex XB-
494 C18 column (250 mm × 10 mm, 5 μm, 100 Å) at the flow rate of 4 mL min^{-1}. Holsalans A1 and A2
495 appeared at 30.8 - 33.5 and 36.0-38.0 min respectively.

496 Ciblan A1

497 Ciblan A1 was produced and extracted from the wild-type strain *H. cibarius* DSM 19505 fermentation
498 broth after 5 days at 37 °C. The isolation of ciblan A1 was eluted with Phenomenex Biphenyl column

(250 mm × 10 mm, 5 μm, 100 Å) with 10% to 14% ACN (+ 0.05% TFA) in 5 - 45 min at the flow rate of 3 mL min^{-1}. Ciblan A1 appeared at 19.9 - 23 min.

Medlan A1-2

For larger-scale fermentation, wild-type *H. mediterranei* was cultured for 3 days under the same conditions mentioned above. The fraction with target compounds was obtained using the methods as archalan β3 going through the ODS open column and LH-20 column. The purification of melan A1-2 was on HPLC system with Phenomenex Biphenyl column (250 mm × 10 mm, 5 μm, 100 Å) with 40% MeOH/ H_2O containing constant 0.05% TFA at the flow rate of 4 mL min^{-1}. Medlan A1 (~4.5 mg) and A2 (**2**) appeared at 20.5 – 24.0 and 14.8 - 15.6 min respectively.

Larlan A2 and 5

The purification of larlans was from the crude extract of heterologous expressed genes *lar AMT* to *lar 7-15* in recombinant *H. volcanii* H1424. Larlan A2 was eluted with 15% ACN (+ 0.05% TFA) on the HPLC system with a Phenomenex Biphenyl column (250 mm × 10 mm, 5 μm, 100 Å) at the flow rate of 4 mL min^{-1}. Peak larlan A2 appeared in 31.8-35.2min. Larlan A5 was eluted with 39% MeOH (+ 0.05% TFA) on the HPLC system with a Phenomenex Kinetex XB - C18 column (250 mm × 10 mm, 5 μm, 100 Å) at the flow rate of 3 mL min^{-1}. Peak larlan A5 appeared in 56.0 -62.0 min.

Two different single mutants of LarA

Each MeLan residues in the larlan A have four different possibilities, LL-Melan, L-*allo*-L-MeLan, and D-*allo*-L-MeLan. Therefore, two single mutants of the Thr in the ring had been constructed and transferred in *H. volcanii* H1424. Around 200 mL culture broth of each mutant strain was extracted by HP20 for HPLC separation. They were purified in different conditions. LarA T2S was purified with a gradient of 15- 20% ACN/H_2O (+ 0.05% TFA) from 5- 45min in a Phenomenex Kinetex XB-C18 column (250 mm × 4.6 mm, 5 μm, 100 Å) at 1 mL min^{-1} and appeared in 29.2 - 30.2 min. LarA T4S was purified with a gradient of 33- 40% MeOH/H_2O (+ 0.05% TFA) from 5- 35min in a Phenomenex Kinetex XB-C18 column (250 mm × 10 mm, 5 μm, 100 Å) at the flow of 3 mL min^{-1} and appeared in 34.2 - 35.1 min. LarA T9S was purified with a gradient of 20- 60% ACN/H_2O (+ 0.05% TFA) from 5- 35min in a Phenomenex Kinetex XB-C18 column (250 mm × 4.6 mm, 5 μm, 100 Å) at 1 mL min^{-1} and appeared in 9.6-10.1 min.

Structure elucidation of archaeal lanthipeptides

^1H, ^{13}C, ^1H-^1H COSY, ^1H-^{13}C HSQC, ^1H-^{13}C HMBC, TOCSY, and NOESY NMR spectra for archalan β3, medlan A1, and larlan A2-5 were recorded on Avance DRX 600 FT-NMR spectrometer (600 and 150 MHz for ^1H and ^{13}C NMR, respectively) using DMSO-d_6. Chemical shifts were reported using the DMSO-d_6 resonance as the internal standard for ^1H-NMR DMSO-d_6: δ = 2.50 p.p.m. and ^{13}C-NMR DMSO-d_6: δ = 39.6 p.p.m. Holsalan A2 and ciblan A1 were using CD_3OD and the internal standard for ^1H-NMR CD_3OD: δ = 3.31, 4.87 p.p.m. and ^{13}C-NMR CD_3OD: δ = 49.0 p.p.m.

Advanced Marfey's analysis for purified archaeal lanthipeptides

Amino acid configurations of purified archaeal lanthipeptides were determined using the advanced

Marfey's method[61,62]. Briefly, lanthipeptides (~ 0.2 mg) were dissolved in 1 mL of 6 N HCl (with 5% (v/v)) heated at 100 °C for 16 h. The hydrolysates were evaporated, redissolved in 100 μL of water, and aliquoted into two portions. Each portion and standard sample were treated with 20 μL of $NaHCO_3$ (1 M) and 50 μL of 1-fluoro-2,4-dinitrophenyl-5-L-leucinamide (L-FDLA) or D-FDLA (1 M) at 40 °C for 2 h, then quenched with 20 μL HCl (1 M) and dried under compressed air. The mixtures were dissolved in 400 μL of MeOH for LR LC-MS analysis and HR LC-MS. Separations were carried out on a Waters Acquity UPLC BEH C18 column (2.1 mm × 150 mm ID, 1.7 μm) by using a gradient elution mode at a flow rate of 0.2 mL min^{-1}: 5 - 80% ACN/H_2O with constant 0.1% formic acid in 2 - 20 min, 80 - 100% in 20 - 25 min, then isocratic 100% for 4 min. The stereochemistry was determined by comparing the retention time of L/D-FDLA derivatized amino acids.

The hydrolysate of synthetic LL-Melan, L-*allo*-L-MeLan, and D-*allo*-L-MeLan from Dong's lab[31] and nisin were used as a reference to confirm the stereochemistry of the MeLan and Lan residues of these lanthipeptides by HR LCMS. For the configuration determination of residue MeLan, different site mutation products' hydrolysates were obtained and analyzed by HR LC-MS using the gradient from 40 - 50% ACN/H_2O (0.1% formic acid) in 2 - 25 min with a Waters ACQUITY Premier Peptide BEH C18 column (1.7 μm, 300 Å, 2.1 × 150 mm).

Bioassay test for archaeal lanthipeptides

Except for purified archaeal lanthipeptides, crude extracts of other remaining classical lanthipeptides we identified by LC-MS/MS analysis were also tested the bioassay using the host strain *H. volcanii* H1424 with plasmid pTA1228 as control. The crude extract of heterologous expressed lanthipeptide BGCs and the control strains were first cultivated for 5 days in a 200 mL YPC medium and then extracted by HP20 with 50% and 100% MeOH elution. The 100% fraction with target compounds (detected by LC-MS) went through a 1 mL C18 SPE column eluted by 0, 5, 10, 20, 30, 50, 70, and 100% MeOH. The target in 70% and 100% C18 fractions were finally redissolved by dd H2O and centrifuged, MeOH precipitation and centrifuge to avoid Halocins.

Ten haloarchaea strains, including *H. larsenii* JCM 13917, *H. mediterranei* ATCC 33500, *Natrinema saccharevitans* JCM 12889, *N. gari* JCM 14663, *Haladaptatus cibarius* DSM 19505, *H. paucihalophilus* JCM 13897, *H. amylolyticus* JCM 18367, *H. salinus* YJ-37-H, *Halomicroarcula salina* JCM 18369, two halophilic bacterial strains, *Halobacillus profundi* JCM 14154 and *H. kuroshimensis* JCM 14155, and four bacterial pathogens, Methicillin-resistance *Staphylococcus aureus* ATCC 43300, *Acinetobacter baumannii* ATCC 19606, *Bacillus subtilis* 168, *Escherichia coli* ATCC 25922 were used as the indicator strain for the archaeal lanthipeptide bioassay test. Haloarchaea strains were cultured in modified DSMZ medium 589 (with 20 g L^{-1} sucrose). Archaea were adjusted to 5.0×10^5 c.f.u. mL^{-1} in the wells of 96-well microtiter plates (Corning, USA), mixed with varying concentrations of compounds and incubated in the shaking incubator (37 °C and 200 r.p.m.) for 2 - 7 days. According to the Japan Collection of Microorganisms, two halophilic bacteria strains were cultivated in 464 medium. Other bacterial strains were grown in LB medium. After overnight (37 °C and 200 r.p.m.) cultivation, they were adjusted to 5.0×10^5 c.f.u. mL^{-1} in the wells of 96-well microtiter plates, different compounds were mixed with and incubated in the static incubator (37 °C and 200 r.p.m.) for 24 h. The volume was 100 μL in each well. The crude extract samples were prepared in a

stock solution of 10 μg μL^{-1} in dd H_2O. The cell growth was monitored at OD_{600} (Varioskan Flash, Thermo Scientific, USA), and vitality was calculated by dividing the difference in OD_{600} detected at the end and beginning of the test by the value of the control group. The inhibition value was obtained by subscribing to the viability value of 100%. Figures were performed by GraphPad Prism 8. The minimal inhibitory concentration of the pure compound (MIC) against each strain was defined as the lowest compound concentration at which archaeal or bacterial growth was not observed.

Motility assay

The hypermotility screening for heterologous expression strains was set up according to previously reported methodology[21,22,48]. Briefly, 2 μL of mid-log phase strain cultures (OD_{600} = 0.1, 50 μL, Thermo Scientific Varioskan Flash) were stabbed into the plates containing modified Hv-Ca + 0.4% agar motility medium supplement with additional trace elements (per liter: 5 g EDTA, 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 0.01 g H_2MoO_4) and incubated in a closed plastic bag at 45 ℃ for 1 day. The time-course motility assays were conducted at 45℃ for 2 - 4 days. For the 6/12/24 - well plates, around 8/4/2 mL/well medium was added, respectively.

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/). Data supporting the results of this study is available in the paper, as well as supplementary information and supplementary data. The HRMS data generated in this study have been deposited in MassIVE under the accession code MSV000095989.

Code availability

The code for amino acid composition bias analysis of LanAs and LanMs assigned a citable DOI through Zenodo (DOI10.5281/zenodo.13705935).

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., and G.W. did the genomic analysis. Z.-M.S., C.C., Y.G., H.L., X.L., G.W., Q.Z., J.Z., P.Y. C., 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 coresponding protocol. Thank Prof. Shihui Dong (College of Chemistry and Chemical Engineering, Lanzhou University, China) for kindly providing four MeLan 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.

Reference

Uncategorized 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). https://doi.org:10.1073/pnas.87.12.4576

2 Baker, B. J. *et al.* Diversity, ecology and evolution of Archaea. *Nat Microbiol* **5**, 887-900 (2020). https://doi.org:10.1038/s41564-020-0715-z

3 Moissl-Eichinger, C. *et al.* Archaea Are Interactive Components of Complex Microbiomes. *Trends Microbiol* **26**, 70-85 (2018). https://doi.org:10.1016/j.tim.2017.07.004

4 Crits-Christoph, A. *et al.* Functional interactions of archaea, bacteria and viruses in a hypersaline endolithic community. *Environ Microbiol* **18**, 2064-2077 (2016). https://doi.org:10.1111/1462-2920.13259

5 von Hoyningen-Huene, A. J. E. *et al.* The archaeome in metaorganism research, with a focus on marine models and their bacteria-archaea interactions. *Front Microbiol* **15**, 1347422 (2024). https://doi.org:10.3389/fmicb.2024.1347422

6 Weiland-Bräuer, N. Symbiotic Interactions of Archaea in Animal and Human Microbiomes. *Curr Clin Microbiol* **10**, 161-173 (2023). https://doi.org:10.1007/s40588-023-00204-7

7 Lian, J., Zou, D. Y., Trebuch, L. M., Duan, C. H. & Li, M. Exploring the interactions between algae and archaea. *Mar Life Sci Tech* (2024). https://doi.org:10.1007/s42995-024-00217-1

8 Dong, Y. *et al.* Eco-evolutionary strategies for relieving carbon limitation under salt stress differ across microbial clades. *Nat Commun* **15**, 6013 (2024). https://doi.org:10.1038/s41467-024-50368-z

9 Grivard, A. *et al.* Archaea Carotenoids: Natural Pigments with Unexplored Innovative Potential. *Mar Drugs* **20** (2022). https://doi.org:10.3390/md20080524

10 Niessen, N. & Soppa, J. Regulated Iron Siderophore Production of the Halophilic Archaeon Haloferax volcanii. *Biomolecules* **10** (2020). https://doi.org:10.3390/biom10071072

659 11 Charlesworth, J. C., Beloe, C., Watters, C. & Burns, B. P. Quorum sensing in archaea: recent advances and
660 emerging directions. *Biocommunication of Archaea*, 119-132 (2017).
661 12 Zhang, G. *et al.* Acyl homoserine lactone-based quorum sensing in a methanogenic archaeon. *ISME J* **6**,
662 1336-1344 (2012). https://doi.org:10.1038/ismej.2011.203
663 13 Kumar, V. *et al.* Halocins, natural antimicrobials of Archaea: Exotic or special or both? *Biotechnol Adv* **53**
664 (2021). https://doi.org:ARTN 107834
665 10.1016/j.biotechadv.2061.107834
666 14 Thompson, T. P. & Gilmore, B. F. Exploring halophilic environments as a source of new antibiotics. *Crit Rev*
667 *Microbiol* **50**, 341-370 (2024). https://doi.org:10.1080/1040841X.2023.2197491
668 15 Liang, H. *et al.* Genomic and metabolic analyses reveal antagonistic lanthipeptides in archaea. *Microbiome*
669 **11**, 74 (2023). https://doi.org:10.1186/s40168-023-01521-1
670 16 Greening, C. *et al.* Minimal and hybrid hydrogenases are active from archaea. *Cell* **187**, 3357-3372 e3319
671 (2024). https://doi.org:10.1016/j.cell.2024.05.032
672 17 Koch, N. G. & Budisa, N. Evolution of Pyrrolysyl-tRNA Synthetase: From Methanogenesis to Genetic Code
673 Expansion. *Chem Rev* (2024). https://doi.org:10.1021/acs.chemrev.4c00031
674 18 Castro, I. *et al.* The lanthipeptide biosynthetic clusters of the domain Archaea. *Microbiol Res* **253**, 126884
675 (2021). https://doi.org:10.1016/j.micres.2021.126884
676 19 Kohtz, A. J. *et al.* Cultivation and visualization of a methanogen of the phylum Thermoproteota. *Nature*
677 (2024). https://doi.org:10.1038/s41586-024-07631-6
678 20 Wu, K. *et al.* Isolation of a methyl-reducing methanogen outside the Euryarchaeota. *Nature* (2024).
679 https://doi.org:10.1038/s41586-024-07728-y
680 21 Schiller, H. *et al.* Identification of structural and regulatory cell-shape determinants in Haloferax volcanii.
681 *Nat Commun* **15**, 1414 (2024). https://doi.org:10.1038/s41467-024-45196-0
682 22 Duggin, I. G. *et al.* CetZ tubulin-like proteins control archaeal cell shape. *Nature* **519**, 362-365 (2015).
683 https://doi.org:10.1038/nature13983
684 23 Zweerink, S. *et al.* Activity-based protein profiling as a robust method for enzyme identification and
685 screening in extremophilic Archaea. *Nat Commun* **8**, 15352 (2017). https://doi.org:10.1038/ncomms15352
686 24 Sysoev, M. *et al.* Bioprospecting of Novel Extremozymes From Prokaryotes-The Advent of Culture-
687 Independent Methods. *Front Microbiol* **12**, 630013 (2021). https://doi.org:10.3389/fmicb.2021.630013
688 25 Cheng, Z. *et al.* Rule-based omics mining reveals antimicrobial macrocyclic peptides against drug-resistant
689 clinical isolates. *Nat Commun* **15**, 4901 (2024). https://doi.org:10.1038/s41467-024-49215-y
690 26 Blin, K. *et al.* antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures
691 and visualisation. *Nucleic Acids Res* **51**, W46-W50 (2023). https://doi.org:10.1093/nar/gkad344
692 27 Repka, L. M., Chekan, J. R., Nair, S. K. & van der Donk, W. A. Mechanistic Understanding of Lanthipeptide
693 Biosynthetic Enzymes. *Chem Rev* **117**, 5457-5520 (2017). https://doi.org:10.1021/acs.chemrev.6b00591
694 28 Baker, B. A. *et al.* Expanded phylogeny of extremely halophilic archaea shows multiple independent
695 adaptations to hypersaline environments. *Nat Microbiol* **9**, 964-975 (2024).
696 https://doi.org:10.1038/s41564-024-01647-4
697 29 Haque, R. U., Paradisi, F. & Allers, T. Haloferax volcanii for biotechnology applications: challenges, current
698 state and perspectives. *Appl Microbiol Biotechnol* **104**, 1371-1382 (2020). https://doi.org:10.1007/s00253-
699 019-10314-2
700 30 Viel, J. H. & Kuipers, O. P. Mutational Studies of the Mersacidin Leader Reveal the Function of Its Unique
701 Two-Step Leader Processing Mechanism. *ACS Synth Biol* **11**, 1949-1957 (2022).
702 https://doi.org:10.1021/acssynbio.2c00088
703 31 Wang, M. *et al.* Genome mining of sulfonated lanthipeptides reveals unique cyclic peptide sulfotransferases.

Acta Pharm Sin B **14**, 2773-2785 (2024). https://doi.org:10.1016/j.apsb.2024.02.016

32 Gilchrist, C. L. M. & Chooi, Y. H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473-2475 (2021). https://doi.org:10.1093/bioinformatics/btab007

33 Ghanmi, F. *et al.* Antagonistic interactions and production of halocin antimicrobial peptides among extremely halophilic prokaryotes isolated from the solar saltern of Sfax, Tunisia. *Extremophiles* **20**, 363-374 (2016). https://doi.org:10.1007/s00792-016-0827-9

34 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* **P. Kavitha, A.P. Lipton, A.R. Sarika and M.S. Aishwarya, 2011. 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.**, 257-262 (2011). https://doi.org:10.3923/rjbsci.2011.257.262

35 Piskovsky, V. & Oliveira, N. M. Bacterial motility can govern the dynamics of antibiotic resistance evolution. *Nat Commun* **14** (2023). https://doi.org:ARTN 5584
10.1038/s41467-023-41196-8

36 Tokarova, V. *et al.* Patterns of bacterial motility in microfluidics-confining environments. *Proc Natl Acad Sci U S A* **118** (2021). https://doi.org:10.1073/pnas.2013925118

37 Dong, X. *et al.* A vast repertoire of secondary metabolites potentially influences community dynamics and biogeochemical processes in cold seeps. *Sci Adv* **10**, eadl2281 (2024). https://doi.org:10.1126/sciadv.adl2281

38 Paoli, L. *et al.* Biosynthetic potential of the global ocean microbiome. *Nature* **607**, 111-118 (2022). https://doi.org:10.1038/s41586-022-04862-3

39 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). https://doi.org:10.1186/s12864-020-06785-7

40 Li, C. *et al.* Mining and Biosynthesis of Bioactive Lanthipeptides From Microorganisms. *Front Bioeng Biotechnol* **9**, 692466 (2021). https://doi.org:10.3389/fbioe.2021.692466

41 Costa, T., Cassin, E., Moreirinha, C., Mendo, S. & Caetano, T. S. Towards the Understanding of the Function of Lanthipeptide and TOMM-Related Genes in Haloferax mediterranei. *Biology (Basel)* **12** (2023). https://doi.org:10.3390/biology12020236

42 Ramawat, K. G. & Goyal, S. in *Co-Evolution of Secondary Metabolites* (eds Jean-Michel Mérillon & Kishan Gopal Ramawat) 3-17 (Springer International Publishing, 2020).

43 Netzker, T. *et al.* Microbial communication leading to the activation of silent fungal secondary metabolite gene clusters. *Frontiers in Microbiology* **6** (2015). https://doi.org:ARTN 299
10.3389/fmicb.2015.00299

44 Tyc, O., Song, C., Dickschat, J. S., Vos, M. & Garbeva, P. The Ecological Role of Volatile and Soluble Secondary Metabolites Produced by Soil Bacteria. *Trends Microbiol* **25**, 280-292 (2017). https://doi.org:10.1016/j.tim.2016.12.002

45 Shu, W. S. & Huang, L. N. Microbial diversity in extreme environments. *Nat Rev Microbiol* **20**, 219-235 (2022). https://doi.org:10.1038/s41579-021-00648-y

46 Martinez, G. M., Pire, C. & Martinez-Espinosa, R. M. Hypersaline environments as natural sources of microbes with potential applications in biotechnology: The case of solar evaporation systems to produce salt in Alicante County (Spain). *Curr Res Microb Sci* **3**, 100136 (2022). https://doi.org:10.1016/j.crmicr.2022.100136

47 Ghai, R. *et al.* New abundant microbial groups in aquatic hypersaline environments. *Sci Rep* **1**, 135 (2011). https://doi.org:10.1038/srep00135

48 Collins, M. *et al.* Mutations Affecting HVO_1357 or HVO_2248 Cause Hypermotility in Haloferax volcanii, Suggesting Roles in Motility Regulation. *Genes (Basel)* **12** (2020). https://doi.org:10.3390/genes12010058

49 Blin, K. *et al.* antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* **49**, W29-W35 (2021). https://doi.org:10.1093/nar/gkab335

50 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). https://doi.org:10.1021/acs.biochem.9b00735

51 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003). https://doi.org:10.1101/gr.1239303

52 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). https://doi.org:10.1186/s12859-023-05311-2

53 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). https://doi.org:10.1093/bioinformatics/btw006

54 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004). https://doi.org:10.1093/nar/gkh340

55 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010). https://doi.org:10.1371/journal.pone.0009490

56 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). https://doi.org:10.1093/nar/gkf436

57 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). https://doi.org:10.1093/molbev/msaa015

58 Revell, L. J. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**, e16505 (2024). https://doi.org:10.7717/peerj.16505

59 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). https://doi.org:10.1093/nar/gkae268

60 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

61 Harada, K. *et al.* Application of D,L-FDLA derivatization to determination of absolute configuration of constituent amino acids in peptide by advanced Marfey's method. *Tetrahedron Lett* **37**, 3001-3004 (1996). https://doi.org:Doi 10.1016/0040-4039(96)00484-4

62 Fujii, K., Ikai, Y., Oka, H., Suzuki, M. & Harada, K. A nonempirical method using LC/MS for determination of the absolute configuration of constituent amino acids in a peptide: Combination of Marfey's method with mass spectrometry and its practical application. *Anal Chem* **69**, 5146-5151 (1997). https://doi.org:DOI 10.1021/ac970289b