

# Cellular differentiation into hyphae and spores in halophilic archaea

Received: 15 June 2022

Accepted: 14 March 2023

Published online: 01 April 2023

 Check for updates

Shu-Kun Tang <sup>1,17</sup> , Xiao-Yang Zhi <sup>1,17</sup>, Yao Zhang<sup>2,17</sup>, Kira S. Makarova <sup>3,17</sup>, Bing-Bing Liu <sup>1,4,17</sup>, Guo-Song Zheng<sup>5</sup>, Zhen-Peng Zhang<sup>2</sup>, Hua-Jun Zheng <sup>6</sup>, Yuri I. Wolf <sup>3</sup>, Yu-Rong Zhao<sup>1</sup>, Song-Hao Jiang<sup>2,7</sup>, Xi-Ming Chen<sup>8</sup>, En-Yuan Li<sup>1</sup>, Tao Zhang<sup>2</sup>, Pei-Ru Chen<sup>2,7</sup>, Yu-Zhou Feng<sup>1</sup>, Ming-Xian Xiang<sup>1</sup>, Zhi-Qian Lin<sup>1</sup>, Jia-Hui Shi<sup>2,7</sup>, Cheng Chang <sup>2</sup>, Xue Zhang <sup>4</sup>, Rui Li<sup>1</sup>, Kai Lou<sup>9</sup>, Yun Wang<sup>9</sup>, Lei Chang<sup>2</sup>, Min Yin<sup>1</sup>, Ling-Ling Yang<sup>1</sup>, Hui-Ying Gao<sup>2</sup>, Zhong-Kai Zhang<sup>10</sup>, Tian-Shen Tao<sup>11</sup>, Tong-Wei Guan<sup>12</sup>, Fu-Chu He<sup>2</sup>, Yin-Hua Lu<sup>5</sup>, Heng-Lin Cui<sup>13</sup>, Eugene V. Koonin <sup>3</sup> , Guo-Ping Zhao <sup>14</sup>  & Ping Xu <sup>2,7,11,15,16</sup> 

Several groups of bacteria have complex life cycles involving cellular differentiation and multicellular structures. For example, actinobacteria of the genus *Streptomyces* form multicellular vegetative hyphae, aerial hyphae, and spores. However, similar life cycles have not yet been described for archaea. Here, we show that several haloarchaea of the family *Halobacteriaceae* display a life cycle resembling that of *Streptomyces* bacteria. Strain YIM 93972 (isolated from a salt marsh) undergoes cellular differentiation into mycelia and spores. Other closely related strains are also able to form mycelia, and comparative genomic analyses point to gene signatures (apparent gain or loss of certain genes) that are shared by members of this clade within the *Halobacteriaceae*. Genomic, transcriptomic and proteomic analyses of non-differentiating mutants suggest that a Cdc48-family ATPase might be involved in cellular differentiation in strain YIM 93972. Additionally, a gene encoding a putative oligopeptide transporter from YIM 93972 can restore the ability to form hyphae in a *Streptomyces coelicolor* mutant that carries a deletion in a homologous gene cluster (*bldKA-bldKE*), suggesting functional equivalence. We propose strain YIM 93972 as representative of a new species in a new genus within the family *Halobacteriaceae*, for which the name *Actinoarchaeum halophilum* gen. nov., sp. nov. is herewith proposed. Our demonstration of a complex life cycle in a group of haloarchaea adds a new dimension to our understanding of the biological diversity and environmental adaptation of archaea.

Archaea are a prokaryotic domain of life that structurally resemble, but are evolutionarily distinct, from bacteria. Apart from the sharp separation in the phylogenetic trees of universal genes (mostly encoding translation system components), archaea differ from

bacteria in many major features including partly unrelated DNA replication and transcription machineries, different structures of membrane lipids and cell walls, and the corresponding, distinct enzymatic machineries involved in membrane and cell wall biogenesis,

A full list of affiliations appears at the end of the paper.  e-mail: [tangshukun@ynu.edu.cn](mailto:tangshukun@ynu.edu.cn); [koonin@ncbi.nlm.nih.gov](mailto:koonin@ncbi.nlm.nih.gov); [gpzhao@sibs.ac.cn](mailto:gpzhao@sibs.ac.cn); [xuping@ncpsb.org.cn](mailto:xuping@ncpsb.org.cn)

several unique coenzymes, and unique RNA modifications<sup>1</sup>. A remarkable diversity of archaeal cell morphology matching that of bacterial cells has been discovered including lobe-shaped *Sulfolobus acidocaldarius*<sup>2</sup>, filamentous *Methanospirillum hungatei*<sup>3</sup> and *Thermofilum pendens*<sup>4</sup>, rod-shaped *Thermoproteus tenax*<sup>5</sup> and *Pyrobaculum aerophilum*<sup>6</sup>, and even square-shaped *Haloquadratum walsbyi*<sup>7</sup>. The distinct shapes of these archaeal cells are maintained by the cell wall and cytoskeleton. However, despite the expanding research into archaeal biology including some limited cellular differentiation, such as cyst formation in *Methanosarcina*<sup>8,9</sup>, so far, there have been no reports of complex cellular differentiation accompanied by major morphological changes in archaea. This apparent lack of complex cellular differentiation in archaea contrasts the diverse different forms of cellular differentiation in bacteria that includes the formation of spores in bacilli and clostridia, heterocysts and akinetes in cyanobacteria, and particularly, complex differentiated colonies in myxobacteria and actinomycetes<sup>10</sup>.

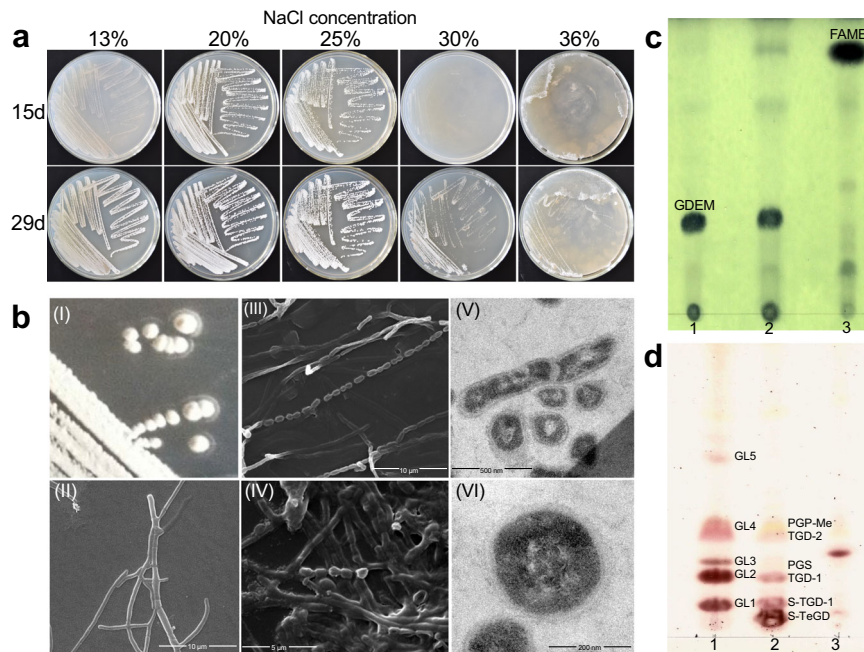
In this work, we describe a group of haloarchaea displaying cellular differentiation into hyphae and spores, isolated from salt marsh sediment. The morphogenesis of this haloarchaea visually resembles that of *Streptomyces* bacteria. Comparative genomics suggests potential gain and loss of genes that might be relevant to this type of archaeal morphogenesis. A gene encoding a Cdc48-family ATPase, and a gene cluster encoding a putative oligopeptide transporter, might be involved in cellular differentiation, as suggested by multi-omic analyses of non-differentiated mutants. Remarkably, this gene cluster from haloarchaea can restore hyphae formation of a *Streptomyces coelicolor* mutant that carries a deletion in a homologous gene cluster (*bldKA-bldKE*). These results provide new knowledge on the morphology of archaea and enrich our understanding of the biological diversity and environmental adaptation of archaea.

## Results and discussion

### A new haloarchaeal lineage with complex morphological differentiation

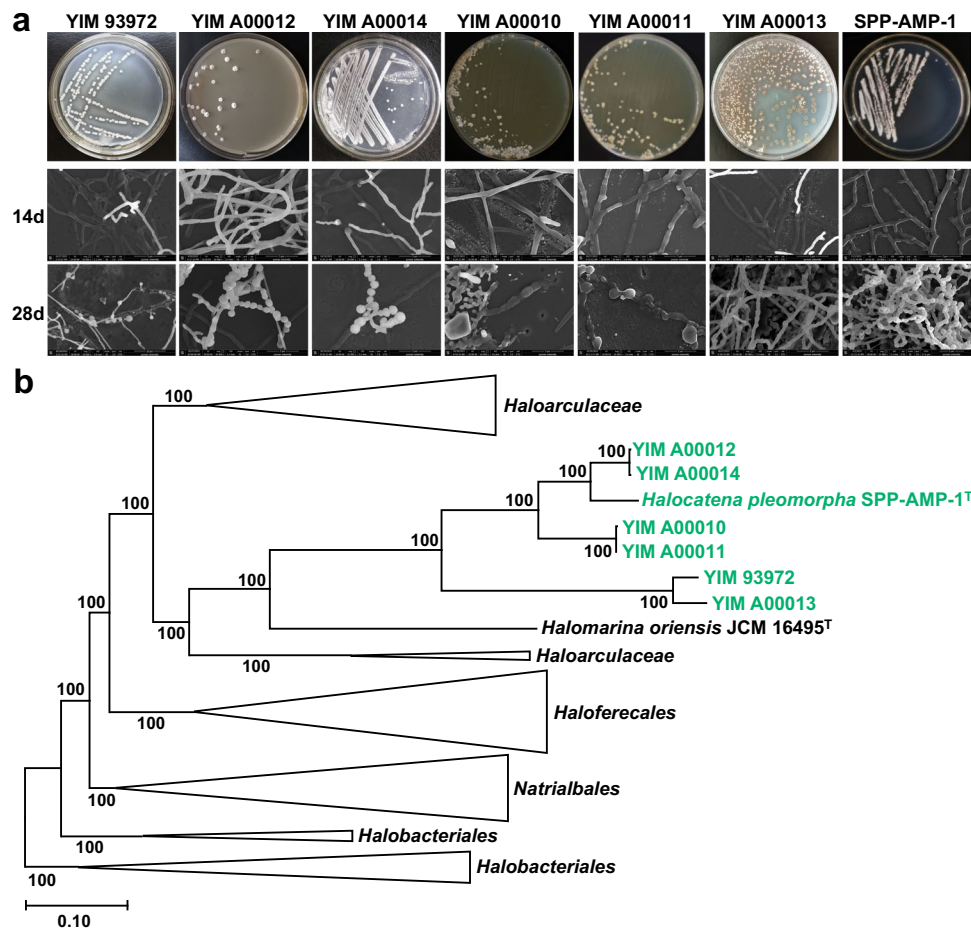
During the investigation of actinobacterial diversity from hypersaline environments, an unexpected hyper-halophilic archaeon with a distinct growth pattern was discovered serendipitously from the Qijiaoqing Salt Lake in the Xinjiang Uygur Autonomous Region of China. This novel strain, designated YIM 93972, was isolated from a soil sample using the standard dilution-plating technique at 37 °C for four weeks on Modified Gause (MG) agar plates supplemented with 20% NaCl (w/v), which previously has been mainly used to isolate halophilic actinomycetes<sup>11</sup>. The colonies of YIM 93972 comprised yellow, orange, and red branching filaments with white aerial hyphae, underwent complex cellular differentiation (Fig. 1a, Supplementary Fig. 1a), and produced spores (0.5–0.7 × 1.0–1.2 µm in size) in both solid (Fig. 1b I–III) and liquid media (Fig. 1b IV). As other spore-forming strains<sup>12</sup>, the spores of this new halophilic archaeon were resistant to heat stress up to 70 °C. In contrast, substrate hyphae (SH) were resistant up to 60 °C (Supplementary Fig. 1b). The higher resistance to heat stress of spores might enable them to survive under extreme saline conditions by their protective and repair capabilities. Optimal growth of YIM 93972 was observed at 40–45 °C, 3.8–4.3 M NaCl and 7.0–7.5 pH (Supplementary Data 1). Transmission electron microscopy (TEM) revealed the different stages of spore differentiation, from cell constriction, septa formation, and rod separation to chain formation (Fig. 1b V and VI). Thus, the cellular differentiation of this new haloarchaeon includes the growth of substrate and aerial hyphae and spore formation, which resembles the morphogenetic development of *Streptomyces*.

Analysis of the 16S rRNA gene indicated that YIM 93972 was closely related to *Halocatena pleomorpha* SPP-AMP-1<sup>†</sup>, with 96.65% identity, and the two species are likely to represent a distinct lineage within



**Fig. 1 | Characteristics of haloarchaeal strain YIM 93972. a** Salt tolerance of strain YIM 93972 grown in ISP 4 medium after 15 and 29 days ( $n = 3$ ). **b** Colony morphology (I) and scanning electron micrographs (EM) of strain YIM 93972 cultivated on solid and liquid ISP 4 medium containing 25% NaCl ( $n = 3$ ). Scanning EM images of substrate mycelia (II; bar, 10 µm), aerial mycelia (III; bar, 10 µm), and long rod-like spore chains (IV; spore size, 0.5–0.7 × 1.0–1.2 µm). Scanning EM images of mycelia grown in liquid ISP 4 medium (IV; bar, 5 µm). Transmission EM images (V and VI) of cells collected from ISP 4 plate after sporulation (bar, 200 nm). **c** Thin-layer chromatographic analysis of whole-organism methanolsates. Lanes: 1, *Haloflex*

*volcanii* CGMCC 1.2150<sup>†</sup>; 2, YIM 93972; 3, *Escherichia coli* K12. **d** Analysis of the polar lipid composition of YIM 93972 using one-dimensional thin layer chromatography. Lanes: 1, YIM 93972; 2, *Halomarina orientis* JCM 16495<sup>†</sup>; 3, *Halomarina salina* ZS-57-S<sup>†</sup>. The origin is at the bottom. Abbreviations: FAMES, fatty acid methyl esters; GDEMs, glycerol diether moieties; GLs, glycolipids; PGP-Me, phosphatidylglycerol phosphate methyl ester; PGS, phosphatidylglycerol-sulfate; TGD-1, galactosyl mannosyl glucosyl diether; TGD-2, glucosyl mannosyl glucosyl diether; S-TeGD, sulfated galactosyl mannosyl galactofuranosyl glucosyl diether; S-TGD-1, sulfated galactosyl mannosyl glucosyl diether.



**Fig. 2 | Morphology and phylogeny of morphogenetic haloarchaea.** **a** Colony morphologies (first row) and scanning electron micrographs (second and third rows) of 7 morphogenetic *Halobacteria* grown in ISP 4 medium after 14 and 28 days ( $n = 3$ ). **b**, Phylogenetic placement of morphogenetic isolates (green) within *Halobacteria*. The Maximum Likelihood phylogenetic tree (IQ-Tree<sup>59</sup>, LG + F + R10

model) is based on 268 orthologous genes that are universal among 130 *Halobacteria* genomes and have at most 4 additional paralogs in any of these. The tree is rooted at the *Halobacterium-Halodesulfurarchaeum* clade following<sup>60</sup>. Branch support values were estimated using aBayes. Scale, substitutions per amino acid site.

the family *Halobacteriaceae*. Although YIM 93972 displays morphological features resembling those of actinomycetes, it is otherwise a typical archaeon that, in particular, contains glycerol diether moieties (GDEMs) and phosphatidylglycerol phosphate methyl ester (PGP-Me) in membrane lipids and an S-layer cell wall (Fig. 1c, d, Supplementary Fig. 2, Supplementary Data 1). In addition, genome analysis and proteomic data indicate that, like many extreme halophilic archaea, YIM 93972 produces halorhodopsin (Supplementary Fig. 3), a 7-transmembrane protein that function as a light-driven ion transporter. However, YIM 93972 lacks the genes encoding enzymes involved in the biosynthesis of bacterioruberin (in particular, lycopene elongase, halo.01227), a C<sub>50</sub> carotenoid that has been identified in several extremely halophilic archaea, the production of which is regulated by halorhodopsin<sup>13,14</sup>.

To investigate the ecological distribution of YIM 93972 and its relatives, we collected 15 soil samples from the Aiding Lake, Large South Lake, Dabancheng East Lake, and Uzun Brac Lake of Xinjiang Uygur Autonomous Region and amplified the 16S rRNA genes from environmental DNAs using a pair of PCR primers designed based on the 16S rRNA gene sequence of YIM 93972. In two soil samples from Aiding Salt Lake, 47 sequences of 16S rRNA gene were obtained and shown to form a clade together with YIM 93972 (OTU13 in Supplementary Fig. 4a, sequences in Supplementary Data 18). Thus, YIM 93972 and related haloarchaea are widespread in the Salt Lake

environment. Additionally, an attempt was made to isolate archaea related to YIM 93972 from the same soil samples from these Salt Lakes using the same method on Modified Gause agar plates. Finally, three strains YIM A00010, YIM A00011, and YIM A00014 from the Aiding Salt Lake as well as two strains YIM A00012 and YIM A00013 from the Uzun Brac Salt Lake were confirmed as novel archaeal strains with morphological differentiation (Fig. 2a).

The phylogenetic tree of 16S rRNA showed that *H. pleomorpha* SPP-AMP-1<sup>T</sup> isolated recently from a man-made saltpan site in India<sup>15</sup> and *Halomarina orientis* JCM 16495<sup>T</sup> isolated from a seawater aquarium in Japan<sup>16</sup> were the two known species most closely related to the six new strains with morphological differentiation (Supplementary Fig. 4b). Cells of *H. orientis* JCM 16495<sup>T</sup> were irregular coccoid or discoid shapes, but those of *H. pleomorpha* SPP-AMP-1<sup>T</sup> were pleomorphic in addition to rod-shaped. These two strains and the other two pleomorphic strains (*Haloferax volcanii* CGMCC 1.2150<sup>T</sup><sup>17</sup> and *Haloplanus salinarum* JCM 31424<sup>T</sup><sup>18</sup>) were collected for morphological comparison. On the same ISP 4 medium containing 20% NaCl, only *H. pleomorpha* SPP-AMP-1<sup>T</sup> exhibited differentiated morphology similar to that of YIM 93972 (Fig. 2a, Supplementary Fig. 5 and 6). Collectively, these observations demonstrated that a distinct group of morphologically differentiating (hereafter morphogenetic, for brevity) haloarchaea is widespread in various hypersaline environments.

## Genomic features of morphogenetic haloarchaea

To gain further insight into the biology of morphogenetic haloarchaea (referring specifically to haloarchaea with hyphae differentiation and spore formation), the complete genomes of YIM 93972 and other five morphogenetic strains were sequenced. The genome of YIM 93972 is composed of a major chromosome (2,676,592 bp, 57.2% G + C content, 2803 predicted genes), a minor chromosome (844,905 bp, 54.7% G + C content, 722 predicted genes), and three plasmids (Supplementary Fig. 7; Supplementary Data 2a). Altogether, the genome of YIM 93972 encompassed 3744 predicted protein-coding genes, 47 tRNAs, and two rRNA operons (the two 16S rRNA genes shared 99.8% identity) located on the major and minor chromosomes. Although all six genomes of the morphogenetic haloarchaeal strains consisted of two chromosomes of different sizes, otherwise, we identified many distinct genomic features (Supplementary Data 2, Supplementary Data 2b–f). In particular, the genome sizes of YIM A00010 and YIM A00011 (5.97 Mb and 5.82 Mb, respectively) are substantially larger than those of the other four strains (3.23 Mb on average). Among the six strains, only YIM A00013 had two rRNA operons like YIM 93972, but the genome of the former carried fewer plasmids and encompassed fewer genes than the latter.

We then constructed a set of 14,870 clusters of haloarchaeal orthologous genes (haloCOGs; Supplementary\_data\_file\_1). Phylogenomic analysis of 268 low-paralogy haloCOGs that are universally conserved in 130 genomes of class *Halobacteria* (Supplementary Data 3) showed that YIM 93972 and its morphogenetic relatives including *H. pleomorpha*, formed a clade with *Halomarina orientis*, within the family *Halobacteriaceae* (Fig. 2b, Supplementary\_data\_file\_2). The morphogenetic haloarchaea shared a core of 2,053 strictly conserved orthologous genes, but lacked 336 genes that are common in other *Halobacteria* (Supplementary Data 3, Supplementary Data 4a). For 3208 of the 3744 predicted genes of YIM 93972, an ortholog was identified in at least one non-morphogenetic haloarchaeon. After mapping the patterns of gene presence-absence in haloCOGs onto the species tree, we obtained a maximum likelihood reconstruction of the history of gene gains and losses in class *Halobacteria*. The ancestor of the morphogenetic haloarchaea was estimated to contain 3903 haloCOGs, of which 1420 were represented in the tree branch that separates the morphogenetic clade from the common ancestor with the sister clade, whereas 339 new genes were gained (Supplementary Data 4b).

Several gene losses appeared to be relevant for the morphogenesis of these new haloarchaea, including the hyphae formation, and lack of cell motility (Supplementary Data 4c). These include the loss of *CetZ* (*FtsZ3*) which is involved in controlling the rod-like cell shape in non-morphogenetic haloarchaea<sup>19</sup>, the archaellum and another halobacterial type IV pili system, typified by the *PilB3* locus of *Haloferax volcanii*<sup>20</sup>, which is represented in *Halobacteria* and *Methanomicrobia*<sup>20</sup>. Rod-shaped cells were not observed in morphogenetic haloarchaea, suggesting a functional link between the loss of *cetZ* and complex hyphae differentiation. Conversely, in morphogenetic haloarchaea, the halo.02163 family of MreB-like proteins is expanded, whereas the halo.02581 family of MreB-like proteins that is common in other haloarchaea is lost. Given that MreB is the key cell shape determinant in bacteria<sup>21</sup>, these findings further emphasize specific changes in cytoskeleton organization that are likely to be relevant for the pleomorphism of these archaea (Supplementary Fig. 8).

Among the gains, there are 61 genes that are strictly conserved in all 7 morphogenetic haloarchaea and absent in other *Halobacteria* in our genome set. These gained genes included a distinct homolog of Cdc48, an AAA+ ATPase (halo.06695, *ORF\_0238*). The proteins of this superfamily play key roles in a variety of cellular processes, including cell-cycle regulation, DNA replication, cell division, the ubiquitin-proteasome-system and others<sup>22, 23</sup> and might contribute to the

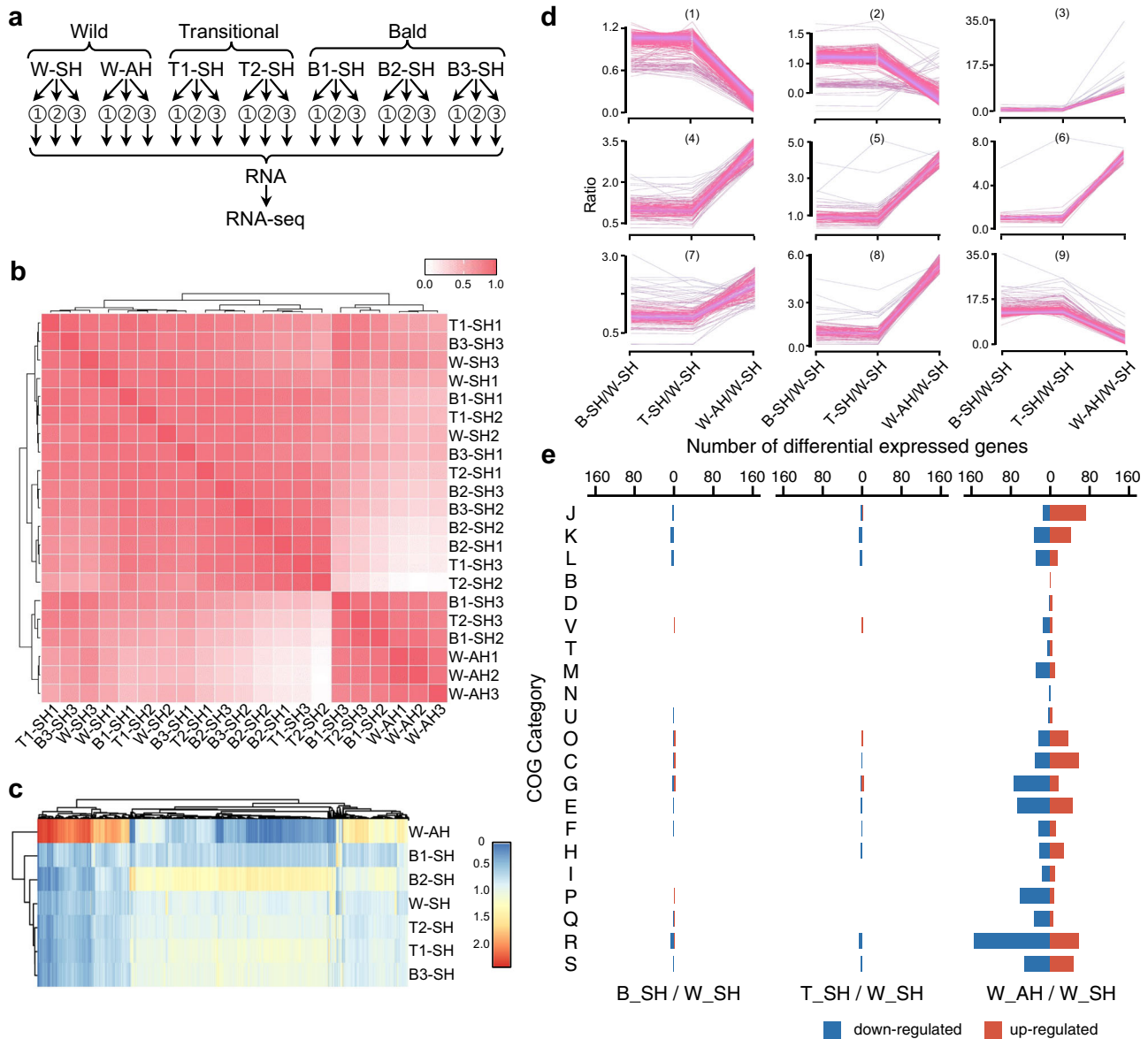
morphological differentiation. Another notable case is a M6 family metalloprotease (halo.06530 and halo.06177, *ORF\_0582* and *ORF\_1236*), a distant homolog of InhA1, a major component of the exosporium in *Bacillus* spores<sup>24</sup>. In most archaea, the S-layer is the exclusive cell envelope component that plays a crucial role in surface recognition and cell shape maintenance<sup>25, 26</sup>. The morphogenetic haloarchaea shared a distinct of S-layer-forming glycoprotein (halo.06692, *ORF\_1400*), which might be related to the unique morphology. Furthermore, proteomic data showed that YIM 93972 produced two varieties of S-layer proteins encoded by *ORF\_1400* and *ORF\_1704* (Supplementary Fig. 9a–d). For most of the remaining genes gained and, in some cases, duplicated in the morphogenetic haloarchaea, there was no functional prediction or only a generic prediction, although notably many of these genes encoded predicted transmembrane or secreted proteins (Supplementary Data 4b, d). Several of these proteins contain aspartate-rich repeats that could coordinate Ca<sup>2+</sup> ions, which are known to activate spore germination in sporulating actinobacteria<sup>27</sup>.

Sporulating bacteria employ a distinct set of small molecules for energy storage and spore protection. We identified genes present in all 7 genomes of morphogenetic haloarchaea that are involved in trehalose biosynthesis and utilization<sup>28</sup>, dipicolinic acid biosynthesis<sup>29</sup> and poly(R)-hydroxyalkanoic acid biosynthesis<sup>30</sup> (Supplementary Data 3). Comparative omics data indicated that genes involved in the biosynthesis of small molecules implicated in sporulation, namely, poly(R)-hydroxyalkanoic acid synthesis (*ORF\_2608*) and 4-hydroxy-tetrahydrodipicolinate (*ORF\_1392*), were upregulated 3.6 and 7.8 folds at the mRNA, and 1.3 and 1.6-fold at the protein level, respectively. This observation suggests that poly(R)-hydroxyalkanoic acid could be a key storage molecule in the spores. Polyhydroxyalkanoates also accumulate in hyphae and spores in some actinobacterial species<sup>30</sup>. However, all these genes are also widespread in non-morphogenetic *Halobacteria*. Another notable gene module present in all genomes of morphogenetic haloarchaea consists of a MoxR family ATPase and a von Willebrand factor type A (vWA) domain containing protein, which is strongly associated with bacterial multicellularity<sup>31</sup>.

## Random mutagenesis and comparative multi-omics uncover several morphogenetic genes

Seeking to identify the genetic basis of cellular differentiation in morphogenetic haloarchaea, we performed nitroso-guanidine (NTG) chemical mutagenesis on spores of YIM 93972. After four generations of serial cultivation from the original 1110 mutant colonies, we obtained five transitional (weak aerial hyphae) and three bald (no aerial hyphae) mutants (Supplementary Fig. 10a, b). Analysis of the genome sequences of these 8 mutants identified two mutations in protein-coding gene sequences and one mutation upstream of a gene in all three bald mutants; and one mutation upstream of a gene in three of the five transitional mutants (Supplementary Fig. 10c, Supplementary Data 5). In the three bald mutants, the two intragenic mutations were the non-synonymous G1067A substitution in *ORF\_0238* (ATPase of the AAA+ class, Cdc48 family), and synonymous C309T substitution in *ORF\_0964* (RNA methyltransferase, SPOUT superfamily), whereas the third mutation was the T136G substitution in the non-coding region upstream of *ORF\_2797* (ParB-like DNA binding protein). In five transitional mutants, the mutation shared by three mutants (T2, T4, and T5) was C69T in the intergenic region upstream of *ORF\_1717* (Uncharacterized protein).

Given the lack of an experimental genetic system for YIM 93972, we performed whole transcriptomic (Fig. 3a, Supplementary Data 6–9) and quantitative proteomic analyses (Fig. 4a, Supplementary Fig. 11, Supplementary Data 10–12) for two transitional and three bald mutants to further characterize the potential association of the identified mutations with the mutant phenotypes. We identified the G1067A substitution in *ORF\_0238* in the transcriptome of all



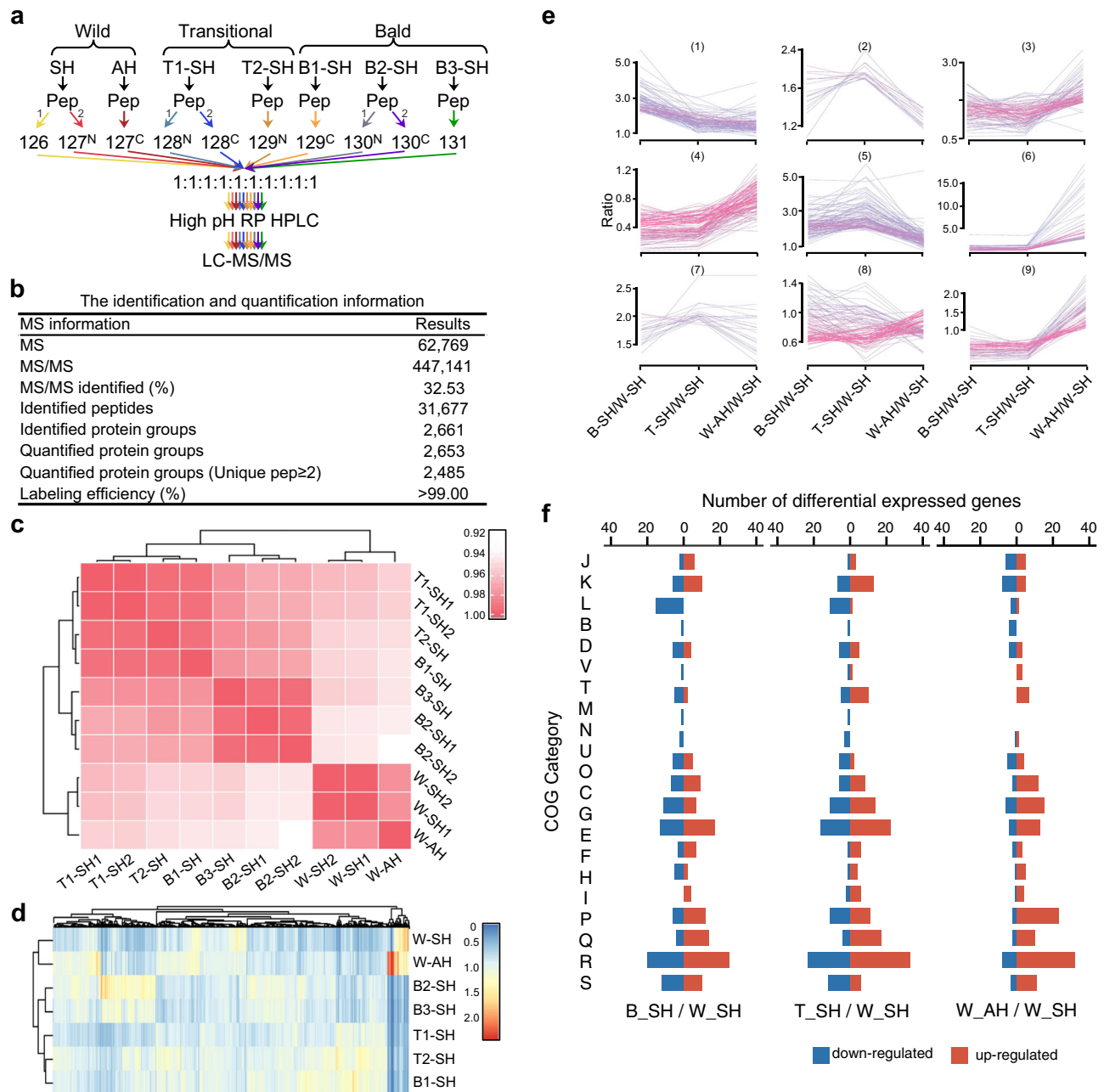
**Fig. 3 | Transcriptomic analysis of strain YIM 93972 wild type and non-differentiating mutants.** **a** Schematic workflow for transcriptomic analysis. We collected cultures representative for aerial hyphae (AH) and substrate hyphae (SH) from wild-type colony, and the morphological mutants. Two transitional (T) and three bald (B) colonies were chosen for the mutants. Each sample included 3 technical repetitions. Total 16 culture plates were randomly pooled for W-SH and W-AH sample, respectively. Total 32 culture plates were randomly pooled for each mutant sample. **b** Spearman's correlation coefficients for transcriptomic profiling of 21 samples. The x and y axes represent the log<sub>2</sub>-transformed gene intensities in each two-sample comparison. **c** Heat maps of differentially expressed genes. Red indicates upregulated genes; blue, downregulated genes. Statistical differences between two groups were analyzed using two-tailed unpaired *t*-tests. *P* < 0.05 was considered statistically significant. **d** The expression clusters of total differentially expressed genes. All dysregulated genes were clustered into different expression groups according to the fold change based on the gene expression value at the

mRNA level in different samples by Mfuzz (v2.58.0)<sup>75</sup>. **e** Functional classification of dysregulated genes according to COG functional categories. The COG categories are listed as follows. J: Translation, ribosomal structure and biogenesis; K: Transcription; L: Replication, recombination and repair; B: Chromatin structure and dynamics; D: Cell cycle control, cell division, chromosome partitioning; V: Defense mechanisms; T: Signal transduction mechanisms; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; U: Intracellular trafficking, secretion, and vesicular transport; O: Posttranslational modification, protein turnover, chaperones; C: Energy production and conversion; G: Carbohydrate transport and metabolism; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; R: General function prediction only; S: Function unknown.

transitional mutants (except for two biological replicates of mutant T2) and bald mutants but not the wild-type transcriptome (Supplementary Fig. 10d, Supplementary Data 7). Furthermore, *ORF\_0238* was downregulated 3.9 fold (*p* < 0.001) at the protein level in the bald mutants (Supplementary Data 11-12). Phylogenetic analysis of the Cdc48 family indicated that *ORF\_0238* belonged to a distinct branch specific for morphogenetic haloarchaea (Supplementary Fig. 12).

Taken together, these observations suggest that *ORF\_0238* might be functionally important for sporulation and morphological differentiation in YIM 93972.

To further characterize the genetic basis of hyphal differentiation, we compared the transcriptome and quantitative proteome of the wild type from both aerial hyphae and substrate hyphae with those of morphogenetic mutants from substrate hyphae only (given the



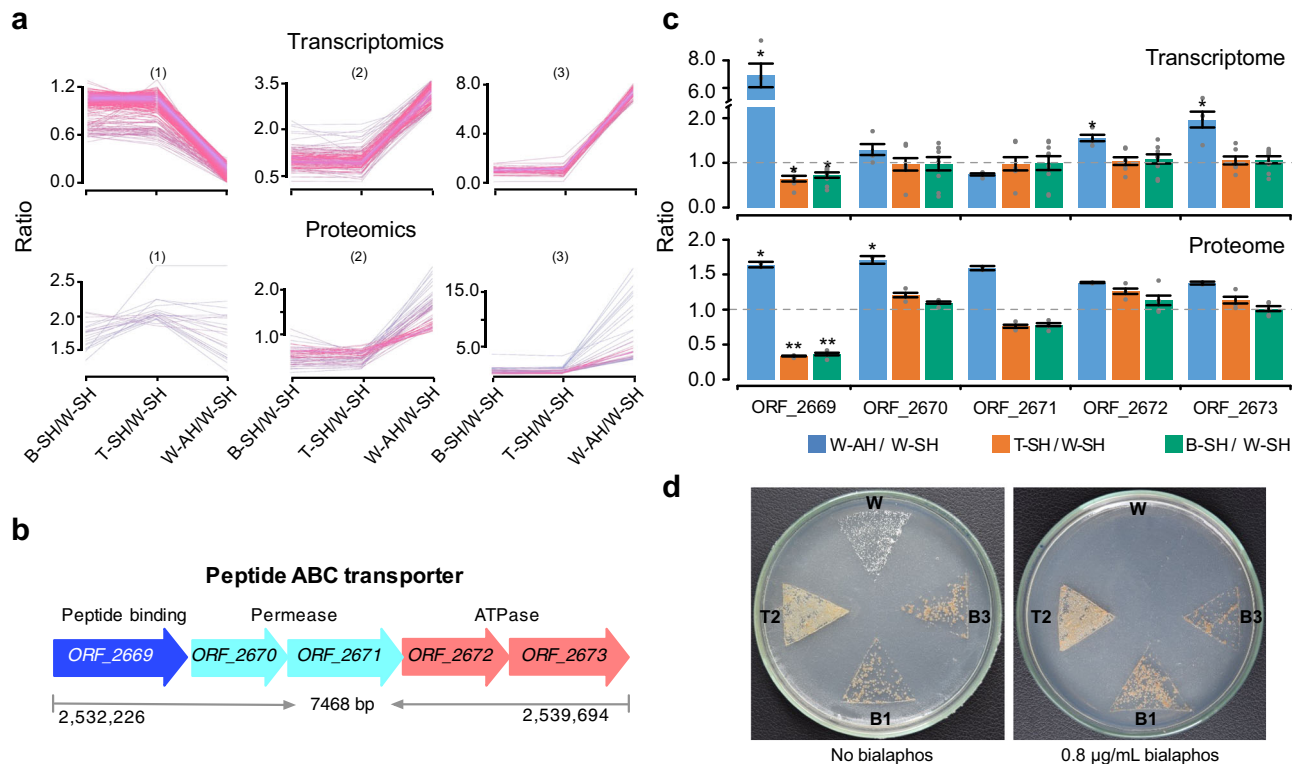
**Fig. 4 | Proteomic analysis of strain YIM 93972 wild type and non-differentiating mutants. a** Overview of TMT methodology for multiplexed comparative analysis. Abbreviations: AH, aerial hyphae; SH, substrate hyphae; W, wild type; T, transitional mutant; B, bald mutant. Two transitional and three bald colonies were selected for the mutants. Each group included two technical repetitions, W-SH1 and W-SH2 in wild group, T1-SH1 and T1-SH2 in transitional group, and B2-SH1 and B2-SH2 in bald group, respectively. Total 16 culture plates were randomly pooled for W-SH and W-AH sample, respectively. Total 32 culture plates were randomly pooled for each mutant sample. **b** Summary of MS identification and quantitation. **c** Spearman's correlation coefficients for proteome profiling of ten

samples. The  $x$  and  $y$  axes represent the  $\log_2$ -transformed protein intensities in each two-sample comparison. **d** Heat maps of differentially expressed proteins. Red indicates upregulated proteins; blue, downregulated proteins. Statistical differences between two groups were analyzed using two-tailed unpaired Significance  $A^{70}$ .  $P < 0.05$  was considered statistically significant. **e** The expression clusters of total differentially expressed proteins. All dysregulated proteins were clustered into different groups according to the fold change based on the protein expression value in different samples by Mfuzz (v2.58.0)<sup>75</sup>. **f** Functional classification of dysregulated genes according to COG functional categories. The annotations of COG categories are listed in the legend of Fig. 3.

absence of aerial hyphae) (Figs. 3a, b and 4a–c, Supplementary Fig. 11a–e). Quantitative analysis revealed 107 genes (Fig. 3c–d, Supplementary Data 9, Supplementary Data 19) and 118 proteins (Fig. 4d, e, Supplementary Data 12, Supplementary Data 20) that were significantly upregulated in the aerial hyphae of wild type compared to the substrate hyphae of both wild and mutant types. The differentially expressed genes and proteins play major roles in translation, transcription, metabolism, and ion transport (Figs. 3e and 4f,

Supplementary Data 13). In particular, ATP-binding cassette (ABC) peptide transporter operon (*ORF\_2669-ORF\_2673* on chromosome I) was significantly upregulated in wild-type aerial hyphae (Fig. 5a, b, Supplementary Data 14). In this operon, the peptide-binding protein (*ORF\_2669*) was significantly upregulated 6.91 and 1.6 fold at the mRNA and protein levels, respectively (Fig. 5c, Supplementary Data 21).

It has been shown that *Streptomyces coelicolor* employs a distinct ATP-binding cassette (ABC) transporter to import oligopeptides that



**Fig. 5 | Sporulation-related genes of strain YIM 93972. a** Typical gene expression profiles in varied differentiating conditions at the mRNA and protein levels. Abbreviations: AH, aerial hyphae; SH, substrate hyphae; W, wild type; T, transitional mutant; B, bald mutant. **b** Gene organization of the peptide ABC transporter in strain YIM 93972. Arrows indicate the size and directions of ORFs. **c** Expression levels of the peptide ABC transporter encoding genes at the mRNA and protein level. In transcriptomic data, gene expression at the mRNA level between two groups was calculated by TPM value and shown as means  $\pm$  MSD from W-AH/W-SH ( $n = 3$ ), T-SH/W-SH ( $n = 6$ ), and B-SH/W-SH ( $n = 9$ ). In proteomic data, gene

expression at the protein level between two groups was calculated by intensity and shown as means  $\pm$  MSD from W-AH/W-SH ( $n = 2$ ), T-SH/W-SH ( $n = 3$ ), and B-SH/W-SH ( $n = 4$ ). Statistical differences between two groups were calculated using two-tailed unpaired *t*-tests (inter-group) at the mRNA level and Significance A<sup>70</sup> (intra-group) at the protein level, respectively.  $P < 0.05$ ;  $^{**}P < 0.01$ ;  $^{***}P < 0.001$ . The fold change and *p*-value were shown in Supplementary Data 4 for Fig. 5c. **d** The morphological mutants are resistant to the toxic peptide bialaphos ( $n = 3$ ). The strains listed above were grown for 14 days on ISP 4 agar plate containing 0 (left plate) and 0.8 (right plate)  $\mu\text{g/mL}$  bialaphos.

serve as signals for aerial mycelium formation<sup>32</sup>. To test whether YIM 93972 similarly used ABC transporters for importing peptides into the mycelia, we tested the resistance of transitional and bald mutants to bialaphos, an antibiotic that enters cells via oligopeptide permeases<sup>33</sup>. We found that only the wild type was sensitive to bialaphos (Fig. 5d) whereas all mutants were resistant to this antibiotic up to concentrations less than 1.0  $\mu\text{g/mL}$  (Supplementary Fig. 13). Despite the low protein sequence similarity with the *S. coelicolor* oligopeptide permease (Fig. 6a), we hypothesized that the *ORF\_2669-ORF\_2673* operon of YIM 93972 is involved in the aerial mycelium and spore formation. Deletion of the oligopeptide-transport operon (*bldKA-KE*) caused a bald phenotype (only present substrate mycelium) in *S. coelicolor*<sup>34</sup>. We introduced the entire *ORF\_2669-ORF\_2673* operon of YIM 93972 into the *S. coelicolor* M145  $\Delta bldKA-KE/pIB139$  mutant (Supplementary Data 15, 16) and found that aerial mycelium formation and sporulation were restored (Fig. 6b). Thus, the *ORF\_2669-ORF\_2673* operon of YIM 93972 might be involved in aerial mycelium formation by importing signaling oligopeptides, and could be functionally equivalent to the *bldKA-KE* operon of *S. coelicolor*.

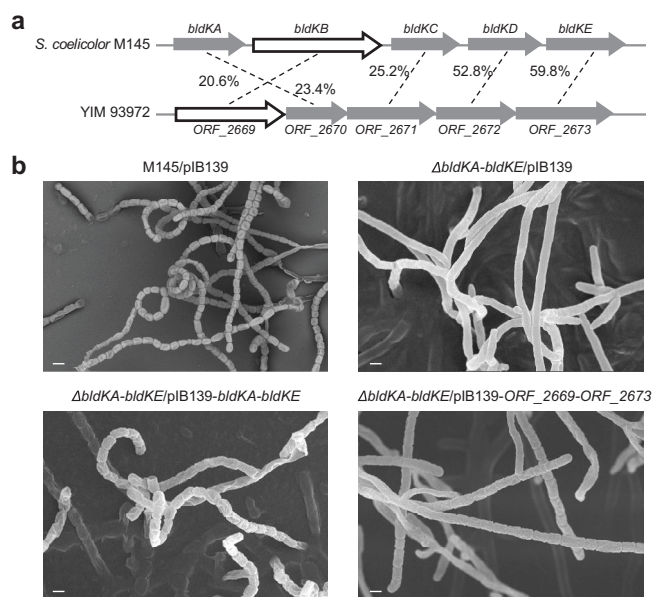
In Actinobacteria, Cyanobacteria and the sporulating Firmicutes, multiple transcriptional regulators (TRs) are involved in complex cell differentiation<sup>35–38</sup>. In our transcriptomic and proteomic data, 47 TRs were found to be significantly up- or down-regulated in aerial hyphae compared to substrate hyphae (Supplementary Data 17). These included four TRs regulating phosphate uptake (*ORF\_0890*, *ORF\_0891*, and *ORF\_1046*) as well as *ORF\_1998*, the cell fate regulator YlbF (YheA/YmcA/DUF963 family), which were all significantly upregulated. The

ortholog of *ORF\_1998* is involved in competence development and sporulation in *Bacillus subtilis*<sup>39</sup>. Another upregulated gene, *ORF\_1932*, encodes the global TR Bola, which is a key player in biofilm development<sup>39–41</sup>. Furthermore, the AbrB family TRs encoded in the same operon with genes involved in poly(R)-hydroxyalkanoic acid biosynthesis, the energy storage molecule in spores, was also upregulated. These findings are compatible with the major roles of distinct TRs in the cellular differentiation of morphogenetic archaea.

In summary, we report the biological, genetic and biochemical characteristics of a haloarchaeon that displays complex cellular differentiation. Although the mechanism of this differentiation remains to be elucidated, our findings implicate several genes, in particular, a distinct Cdc48-like ATPase that was mutated in all non-differentiating mutants. The observation that the morphogenetic haloarchaea are closely related to each other and, in the phylogenetic tree of *Halobacteria*, form a distinct clade within the family *Halobacteriaceae* suggests a relatively recent origin of this complex phenotype. It remains to be shown whether complex cellular differentiation, which evolved independently in several groups of bacteria, also independently emerged in other archaea.

### Taxonomic description

As mentioned above, phylogenetic analysis of 16S rRNA and concatenated conserved genes indicated that strain YIM 93972 formed a clade with the closest related species *Halocatena pleomorpha* within family *Halobacteriaceae*. The average nucleotide identity based on BLAST (ANiB) value between YIM 93972 and *H. pleomorpha*



**Fig. 6 | Role of peptide permeases in morphological differentiation: complementation of a *bldKA-KE* deletion mutant of *Streptomyces coelicolor* with the homologous operon of YIM 93972.** **a** Comparison of the *bldK* operon of *S. coelicolor* M145 with *ORF\_2669-ORF\_2673* of YIM 93972; arrows indicate the size and directions of genes; percentage of amino acid sequence identity is indicated for each gene. **b** Comparison of mycelia phenotypes among M145/pIB139,  $\Delta$ *bldKA-KE*/pIB139,  $\Delta$ *bldKA-KE*/pIB139,  $\Delta$ *bldKA-KE*/pIB139-*bldKA-bldKE*, and  $\Delta$ *bldKA-KE*/pIB139-*ORF\_2669-ORF\_2673* using scanning EM (bar, 5  $\mu$ m). Deletion of *bldKA-KE* in M145 abrogated the development of aerial mycelia and spores; this phenotype was restored to near wild-type levels by complementation with *bldKA-KE* or *ORF\_2669-ORF\_2673* ( $n = 3$ ).

SPP-AMP-1<sup>T</sup> was 70.59%, which was far below the cut-off values <75 % for genus demarcation in the class *Halobacteriia*<sup>42</sup>. Further, the low AAI (68.4%) between YIM 93972 and *H. pleomorpha* SPP-AMP-1<sup>T</sup> confirms the novelty of YIM 93972 at the genus level<sup>42</sup>. The substrate mycelium of YIM 93972 formed sporangia, and the aerial mycelium formed spore chains at maturity and the spores were cylindrical with wrinkled surfaces. The respiratory quinones of strain YIM 93972 are MK-8 and MK-8(H<sub>2</sub>) (Supplementary Fig. 14). By virtue of having a different isomer for MK-8 (H<sub>2</sub>), the respiratory quinones of YIM 93972 differ from those of *H. pleomorpha* SPP-AMP-1<sup>T</sup>, which contains only menaquinone MK-8. The major polar lipids of the strain were chromatographically identified as phosphatidylglycerol, phosphatidylglycerolphosphate methyl ester and five unidentified glycolipids, while *H. pleomorpha* SPP-AMP-1<sup>T</sup> has phosphatidylglycerol, phosphatidylglycerolphosphate methyl ester, glycosyl mannosyl glucosyl diether and sulphated glycosyl mannosyl glucosyl diether. Besides the phylogenetic and morphological differences, the isolate is differentiated from the closely related genera by its chemotaxonomic markers (Supplementary Data 1). Based on the above results, YIM 93972<sup>T</sup> merits representation as a new species in a new genus within the family *Halobacteriaceae*, for which the name *Actinoarchaeum halophilum* gen. nov., sp. nov. is herewith proposed.

### Description of *Actinoarchaeum* gen. nov

*Actinoarchaeum* (Ac.tino.ar.chae'um. Gr. n. *aktis* -inos, a ray; N.L. neut. n. *archaeum* (from Gr. adj. *archaios* -*ē* -on, ancient), archaeon; N.L. neut. n. *Actinoarchaeum* ray archaeon, referring to the radial arrangement of filaments).

Aerobic, extremely halophilic, *Streptomyces*-like colonies that produce brown substrate mycelia forms terminal sporangia with white aerial mycelia and white spore on modified ISP 4 medium. The major polar lipids are phosphatidylglycerol, phosphatidylglycerol phosphate

methyl ester and five unidentified glycolipids. The predominant menaquinones are MK-8 and MK-8(H<sub>2</sub>). The G+C content of the genomic DNA is about 56.3 mol%. The type species is *Actinoarchaeum halophilum*. Recommended three-letter abbreviation: *Aah*.

### Description of *Actinoarchaeum halophilum* sp. nov

*Actinoarchaeum halophilum* (ha.lo'phi.lum. Gr. n. *hals* halos, salt; N.L. adj. *philus* -*a* -um, from Gr. adj. *philos* -*ē* -on, friend, loving; N.L. neut. adj. *halophilum*, salt-loving).

Morphological, chemotaxonomic, and general characteristics are as given above for the genus. Cells require 2.1–6.0 M NaCl, pH 6.0–9.0, 25–50 °C and Mg<sup>2+</sup> (0.01–0.7 M) for growth. Optimal growth occurs at 3.8–4.2 M NaCl, pH 7.0–7.5, 40–45 °C. Oxidase-weakly positive and catalase-negative. Positive for nitrate reduction, hydrolysis of Tweens (20, 40, 60, 80), casein and gelatin; while negative for productions of indole and H<sub>2</sub>S. Acetate, citrate, dextrin, fructose, fumarate, glucose, glycerol, malate, mannitol, mannose, pyruvate, rhamnose, succinate, sucrose and trehalose are utilized as sole carbon sources, but galactose, lactate, lactose, xylitol and xylose are not. Acid is not produced from any of sole carbon sources stated above. The strain contains MK-8 and MK-8(H<sub>2</sub>). Polar lipids include PG, PGP-Me and 5GLs. The genomic DNA G+C content is 56.3%. The type strain is YIM 93972<sup>T</sup> (=DSM 46868<sup>T</sup> = CGMCC 1.17467<sup>T</sup>), isolated from a soil sample from Salt Lake in Xinjiang Uygur Autonomous Region of China.

## Methods

### Isolation, identification, morphological observation

The sediment samples were collected from Qijiaoqing Salt Lake in Xinjiang Uygur Autonomous Region, China. Isolation was done using the standard dilution-plating technique on Modified Gause (MG) medium<sup>11</sup> containing (g/L): soluble starch, 5; lotus root starch, 5; KNO<sub>3</sub>, 1; MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.5; K<sub>2</sub>HPO<sub>4</sub>, 0.5; NaCl, 200; agar, 20; 1 mL trace solution (2% FeSO<sub>4</sub>·7H<sub>2</sub>O; 1% MnCl<sub>2</sub>·4H<sub>2</sub>O; 1% ZnSO<sub>4</sub>·7H<sub>2</sub>O; 1% CuSO<sub>4</sub>·5H<sub>2</sub>O; pH adjusted to 7.2). Plates were incubated at 37 °C for at least 6 weeks. All isolates with actinobacteria-like filamentous colonies were collected. In the identification of these isolates based on 16S rRNA gene, one isolate designated YIM 93972 was paid more attention, because its 16S rRNA gene could not be amplified using universal primers for the bacterial 16S rRNA gene, but the primers for archaeal 16S rRNA gene worked. Cell morphology of strain YIM 93972 was further examined using a scanning electron microscopy (Quanta 2000, FEI, Hillsboro, OR, USA) at different concentrations of NaCl. Cells from a 20-day-old culture of strain YIM 93972 were fixed for 30 min in OsO<sub>4</sub> vapour on ice, and filtered onto 2  $\mu$ m isopore membrane filters (Millipore, Tokyo, Japan) modified by the method of Schubert et al.<sup>43</sup>. The fixed cultures were dehydrated in a series of 50–100% ethanol mixtures, critical-point-dried in CO<sub>2</sub> and sputter-coated by 10 nm of gold-palladium. Cells were imaged using a XL30 ESEM-TMP (Philips-FEI, Eindhoven, Holland) at 3 kV. For testing the temperature tolerance of spores, spores were separated from substrate hyphae by virtue of a nitrocellulose transfer membrane (see below for details). Spore and mechanically broken substrate hyphae were used for preparing suspension (2 OD) and bathed at 60, 70, 80 and 90 °C for 15 min (control, 37 °C). Then, 100  $\mu$ L suspension was spread on solid ISP 4 medium with 20% NaCl, and cultured for 30 days. Images were collected every 10 days.

For obtaining more strains like YIM 93972, the same isolation was performed on the soil samples collected from the Aiding, Large south, Dabancheng East, and Uzun Brac Salt Lakes in Xinjiang Uygur Autonomous region. Finally, five haloarchaea strains were isolated and confirmed with morphological differentiation (cultured on MG plates containing 20% NaCl at 37 °C for 3–4 week). Among them, YIM A00010, YIM A00011, and YIM A00014 were isolated from the Aiding Salt Lake samples; YIM A00012 and YIM A00013 were isolated from the Uzun Brac Salt Lake samples. For further understanding the diversity of YIM



93972 and its relatives in these four salt lakes, the 16S rRNA genes were amplified by using a pair of YIM 93972-specific primers (Forward primer: 5' -GGGCGTCCAGCGGAAACC-3' ; Reverse primer: 5' -CCAT-CAGCCTGACTGTCAT-3'), directly from the total DNA of 15 soil samples. The PCR products were isolated from 2% agarose gel and purified using the AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA, USA) according to manufacturer's instructions and quantified using Quantus™ Fluorometer (Promega, Madison, WA, USA). Purified amplicons were pooled in equimolar amounts and paired-end sequenced on an Illumina MiSeq PE300 platform (Illumina, San Diego, CA, USA) according to the standard protocols by Majorbio Bio-Pharm Technology Co. Ltd. (Shanghai, China)<sup>44</sup>.

### Genomic sequencing

The total genomic DNA from strain YIM 93972 was extracted using the DNeasy Power Soil Kit (QIAGEN, Hilden, Germany). The quality of DNA was assessed by 1% agarose gel electrophoresis, and the purity were measured using NanoDrop™ 2000 spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). Qualified genomic DNA (5 µg) was subjected to SMRT sequencing using a Pacific Biosciences RSII sequencer (Pacific Biosciences, Menlo Park, CA, USA). The Hierarchical Genome Assembly Process (HGAP, v2.3.0) pipeline was used to generate high quality de novo assembly of the genome with default parameters<sup>45</sup>. The genomic DNAs of other five isolates were extracted as well as, and the quality was assessed using the same procedure. Then, the genomic DNAs were split into two fetches and sequenced using Oxford Nanopore (PromethION). The raw datasets were filtered, the subreads were assembled by Canu (v1.5)<sup>46</sup>, and further corrected by Pilon (v1.22)<sup>47</sup>. For further calibration the errors of the assembled genome using second generation sequencing datasets by Illumina NovaSeq 6000, and high-quality reads were obtained to generate one or more contigs without gaps. The rRNAs and tRNAs of YIM 93972 were predicted using RNAmmer (v1.2)<sup>48</sup> and tRNAscan-SE (v1.3.1)<sup>49</sup>, respectively. The gene predication and annotation were performed using Prokka (v1.14.6). The circular maps were generated by DNA-Plotter (v1.11)<sup>50</sup>. The detail assembled and annotated results are shown in Supplementary Data 2a-f.

### Comparative genome and phylogenome

The complete genomes of 122 *Halobacteria* isolates were downloaded from NCBI Genomes site. Additionally, contig level genome assemblies of a pleomorphic haloarchaeon *Halocatena pleomorpha* SPP-AMP-1<sup>T</sup> (GCF\_003862495.1) and a closely related non-morphogenetic haloarchaeon *Halomarina orientis* JCM 16495<sup>T</sup> (GCF\_009791395.1), were downloaded from the same source. Protein sequences, encoded by these 124 publicly available *Halobacteria* genomes, as well as those encoded in the six genomes, obtained in this study, were analyzed to establish the orthology relationships between their genes (Supplementary\_data\_file\_1). Initially, all protein sequences were clustered using MMseqs2 (v14-7e284)<sup>51</sup> with sequence similarity threshold of 0.5; then the clusters were further refined via several iterations of the following procedure:

- cluster alignments, obtained using MUSCLE (v5)<sup>52</sup> were compared to each other using HHSEARCH (v3.3.0)<sup>53</sup> clusters, displaying full-length similarity were merged together;
- approximate ML phylogenetic trees were reconstructed using FastTree (v 2.1.11)<sup>54</sup> for the merged cluster alignments<sup>52</sup> and rooted by mid-point; trees were parsed into subtrees, maximizing the ratio of taxonomic coverage (number of the distinct genome assemblies in the subtree) to the paralogy index (average number of sequences per assembly);

This procedure produced a set of 14,870 clusters (excluding singleton sequences) of orthologous genes (haloCOGs) that subsequently served for comparative evolutionary genomics analysis of

*Halobacteria* genes. Functional annotation of haloCOGs was obtained by comparing haloCOG alignments to CDD<sup>55</sup> and arCOG<sup>56</sup> sequence profiles using HHSEARCH<sup>53</sup>.

268 haloCOGs with full complement of 130 genomes and at most 4 additional paralogs were used to determine the genome-level phylogeny of *Halobacteria* (Supplementary Data 3, Supplementary\_data\_file\_2). When paralogs were present, the index ortholog was selected based on the BLOSUM62<sup>57</sup> alignment score between the paralogs and the alignment consensus. Columns in the alignments of these 268 haloCOGs were filtered<sup>58</sup> for maximum fraction of gaps (0.667) and minimum homogeneity (0.05). The concatenated alignment contained 71,586 amino-acid sites. Phylogenetic tree was reconstructed using IQ-Tree (v2.2.0)<sup>59</sup> under the LG+F+R10 model, selected by the built-in model finder and rooted according to Rinke et al. 2021 (Supplementary\_data\_file\_2)<sup>60</sup>.

The history of gene gains and losses (Supplementary Data 4b, c) in *Halobacteria* was reconstructed from the haloCOG phyletic patterns and the *Halobacteria* phylogenetic tree using GLOOME (v201305)<sup>61</sup>. Gains or losses of a gene on a particular edge of a phylogenetic tree were inferred from changes in the posterior probability of this gene presence between the ancestral and the descendant genomes at the respective ends of this edge; a change in probability exceeding 0.5 in magnitude was interpreted as a likely gain or loss of the gene (Supplementary Data 4 and 4c).

### Morphologically defective mutants screening and genome resequencing

To obtain the poorly differentiated and undifferentiated mutants, a nitroso-guanidine (NTG) based mutagenesis experiment was applied to the wild type of YIM 93972. Spore suspensions were prepared by adding sterile glass beads (5 mm diameter; 1 g per 10 mL medium) in ISP 4 medium containing (g/L): (soluble starch, 10; K<sub>2</sub>PO<sub>4</sub>, 1; MgSO<sub>4</sub>, 1; (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 2; NaCl, 250) at 37 °C for 3 days, and then adjusted to 10<sup>8</sup> spores/mL. NTG treatment was carried out by incubating spore suspension with varied concentrations of NTG from 0.2 to 1 mg/mL at 30 °C for 30 min on an incubator shook at 90 g. Samples of 10 mL were harvested at 2400 g for 5 min, and spores were washed three times with 20 mL of 20% sterile NaCl. The spore lethality rates were measured by the LIVE/DEAD™ BacLight™ bacterial viability kit (ThermoFisher Scientific, Waltham, MA, USA) as described previously<sup>62</sup>. The NTG treated spores with ~85% lethality rates were diluted to 10<sup>2</sup> and 10<sup>3</sup> per milliliter, and then 100 µL aliquots were equally spread on ISP 4 medium plates containing 25% NaCl and then incubated at 37 °C for 28 days. The mutants poorly differentiated (transitional) and that undifferentiated (bald) were cultured and passaged four times for further multi-omics study (Supplementary Fig. 10a). The morphological phenotype of many mutants recovered in the next sub-cultivation. So, after 4 generations, only 5 transitional mutants and 3 bald mutants remained with stable morphological mutant phenotypes. To further confirm the morphological mutation of strain YIM 93972, we also checked the cell morphology by scanning electron microscopy as described above. Compared with wild type colonies, two transitional colonies had branched substrate mycelia and very rare aerial mycelia. However, the three bald colonies only had branched substrate mycelia (Supplementary Fig. 10b).

### Comparative transcriptome

For collecting the aerial and substrate hyphae of wild type, viable spores 10<sup>7</sup> per mL of strain YIM 93972 were collected after full sporulation of the aerial hyphae (28 days of solid medium incubation) and spread on a layer of 0.20 µm BioTrace NT Nitrocellulose Transfer Membrane (Pall China, Beijing, China) covered on a medium ISP 4 containing 20% NaCl and 2% agar in 85 mm diameter Petri dish. The plates were cultured at 37 °C for 21 days. Totally, 96 culture plates were prepared, which were split into 6 groups as biological replicates.

The aerial hyphae grown on nitrocellulose membrane, and the substrate hyphae penetrated through nitrocellulose membrane were recovered by scraping with a plain spatula, respectively. For two transitional and three bald mutants, the equal amounts of mechanically broken hyphae were spread on the same plates as described above. Totally 192 culture plates were prepared for each mutant, which were split into 6 groups as biological replicates. The substrate hyphae were recovered at 28 days. Totally, 6 pooled hyphae were collected for each strain, half of them were used for transcriptomic analysis, and another half were used for proteomic analysis.

The total RNA was extracted and analyzed as described previously<sup>63</sup>. Briefly, total RNA was extracted with Trizol method. The TruSeq™ Stranded Total RNA Library Prep Kit (Illumina, San Diego, CA, USA) was used to prepare the cDNA libraries, followed by paired-end 100 bp sequencing on an Illumina HiSeq 2000 system (Illumina, San Diego, CA, USA). The RNA-Seq reads per sample were mapped to the reference using Bowtie2 (v 2.4.2)<sup>64</sup>. RSEM<sup>65</sup> (v1.3.3) was used to estimate TPM. SNP analysis was performed by SAM tools<sup>66</sup> (v1.12). The detailed quality control was showed in Supplementary Data 6 and Fig. 3b. Genes with ratios greater than 1.5-fold and *p*-value smaller than 0.05 (Student's *t*-test) were considered as regulated and used for further bioinformatics analysis (Fig. 3d and Supplementary Data 9).

### Comparative proteome

For proteomic analysis, the aerial hyphae and substrate hyphae were suspended in lysis buffer [9 M Urea, 10 mM Tris-HCl (pH 8.0), 30 mM NaCl, 5 mM iodoacetamide (IAA), 5 mM Na<sub>4</sub>P<sub>2</sub>O<sub>7</sub>, 100 mM Na<sub>2</sub>HPO<sub>4</sub> (pH 8.0), 1 mM NaF, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 1 mM sodium glycerophosphate, 1% phosphatase inhibitor cocktail 2, 1% phosphatase inhibitor cocktail 3, EDTA-free protease inhibitor cocktail (1 tablet/10 mL lysis buffer)] and disrupted by a Soniprep sonicator (Scientz, Ningbo, Zhejiang, China) for 10 min (2 s on and 4 s off, amplitude 30%) as described<sup>67</sup>. The lysates were centrifuged at 16,200 *g* for 10 min at 4 °C to remove debris. The quality and concentration of extracted total cell lysates (TCL) were detected by a 10% SDS-PAGE (Supplementary Fig. 11a). The technical replicates were designed in wild, transitional, and bald groups, respectively.

Same amount of pooled proteins (120 µg) from each sample was reduced with 5 mM of dithiothreitol (DTT) at 45 °C for 30 min, followed by alkylation with 10 mM of iodoacetamide (IAA) at room temperature for 30 min. The alkylated samples were pre-cleaned with a 10% SDS-PAGE (0.7 cm)<sup>68</sup>, and in-gel digested with 12.5 ng/µL trypsin<sup>69</sup> at 37 °C for 14 h. Equal amount of peptides from each differentiated condition were used for 10-plex TMT labeling (ThermoFisher Scientific, Waltham, MA, USA). TMT labeling was carried out as follows: wild substrate hyphae (W-SH) technical replicates were labeled with 126 and 127<sup>N</sup>; wild type aerial hyphae (W-AH) with 127<sup>C</sup>; transitional substrate hyphae (T-SH) technical replicates with 128<sup>N</sup> and 128<sup>C</sup>, T-SH biological replicate with 129<sup>N</sup>; bald substrate hyphae (B-SH) technical replicates with 130<sup>N</sup> and 130<sup>C</sup>, B-SH biological replicates with 129<sup>C</sup> and 131 following the manufacturer's protocol (Fig. 4a). The labeled peptides were mixed and dried with a vacuum dryer (LABCONCO CentriVap, Kansas City, MO, USA).

The mixed TMT labeled peptides were fractionated by a Durashell C<sub>18</sub> high pH reverse phase (RP) column (150 Å, 5 µm, 4.6 × 250 mm<sup>2</sup>, Bonna-Agela Technologies Inc., Newark, DE, USA) on a Rigol L-3120 HPLC system (Beijing, China) as described previously<sup>67</sup>. Briefly, the solvent gradients comprised buffer A (98% double distilled H<sub>2</sub>O and 2% ACN, pH 10, adjusted by ammonium hydroxide) and buffer B (2% double distilled H<sub>2</sub>O and 98% ACN, pH 10). The mixed samples were dissolved in buffer A. After loading, the peptides were separated with a 60 min linear gradient (0% B for 5 min, 0–3% B for 3 min, 3–22% B for 37 min, 22–32% B for 10 min, 32–90% B for 1 min, 90% B for 2 min, and 100% B for 2 min). The LC flow rate was set at 0.7 mL/min. The column was maintained at 45 °C. Eluent was collected every 1 min

(Supplementary Fig. 11b). The 60 fractions were dried and concatenated to 10 fractions (Supplementary Fig. 11c). The combined peptides were subjected to LC–MS/MS analysis.

The fractions were analyzed on a Q Exactive HF mass spectrometer (ThermoFisher Scientific, Waltham, MA, USA) after Easy-Nano LC 1200 separation (ThermoFisher Scientific, Waltham, MA, USA). Briefly, samples were loaded onto a self-packed capillary column (75 µm i.d. × 50 cm, 1.9 µm C<sub>18</sub>) and eluted with a 135 min linear gradient (4–8% B for 13 min, 8–25% B for 86 min, 20–50% B for 21 min, 50–90% B for 3 min, 90% B for 12 min). Full MS scans were performed with *m/z* ranges of 375–1,400 at a resolution of 1.2 × 10<sup>5</sup>; the maximum injection time (MIT) was 80 ms, and the automatic gain control (AGC) was set to 3.0 × 10<sup>6</sup>. For the MS/MS scans, the 15 most intense peptide ions with charge states of 2 to 6 were subjected to fragmentation via higher energy collision-induced dissociation (HCD) (AGC: 1 × 10<sup>5</sup>, MIT: 100 ms, Resolution: 6 × 10<sup>4</sup>). Dynamic exclusion was set as 30 s.

All of the raw files were searched with MaxQuant (v1.5.6.0) against the protein database from strain YIM 93972 (3744 entries) along with 245 common contaminant protein sequences (<http://www.maxquant.org>). Fully tryptic peptides with up to two miss cleavage sites were allowed. Oxidation of methionine was set as dynamic modification, whereas carbamidomethylation of cysteine and TMT modification at peptide N-terminus and lysine were set as static modifications. Figure 4b showed the MS identification and quantification. The detail quality control was showed in Supplementary Fig. 11d, e. Nine protein expression clusters were showed in Fig. 4e. Protein changes greater than 1.5-fold and *p*-value smaller than 0.05 (Significance A<sup>70</sup>) were considered as regulated, which were further used for bioinformatics analysis (Fig. 4e and Supplementary Data 12).

### Heterologous expression and functional analysis

To characterize the putative oligopeptide transporter encoded by *ORF\_2669-ORF\_2673* from YIM 93972, a bialaphos (from 0.1 µg/mL to 1.0 µg/mL with 0.1 as interval) resistance test<sup>34, 67</sup> was performed using solid ISP 4 medium. Then, the heterologous expression of genes *ORF\_2669-ORF\_2673* in *Streptomyces coelicolor* M145<sup>71</sup> with a deletion of gene cluster *bldKA-KE* was conducted. The  $\Delta$ *bldKA-KE* mutant was constructed by the CRISPR-Cas9 genome editing method as previously described<sup>72</sup>. The strains and plasmids used were listed in Supplementary Data 15, and the primers in Supplementary Data 16. Briefly, the upstream (1139 bp) and downstream regions (1310 bp) of *bldKA-KE* were amplified using the primer pairs *bldKA-KE*-up-F/R and *bldKA-KE*-down-F/R, respectively. The single guide RNA (sgRNA) transcription cassette was obtained from the plasmid pKCCas9dO with the primers *bldKA-KE*-gRNA F/R. Then, three above fragments were assembled by an overlapping PCR with the primers *bldKA-KE*-gRNA-F and *bldKA-KE*-down-R, followed by double digestion with *Spe* I and *Hind* III. The assembled fragment was cloned into pKCCas9dO and was introduced into the wild strain M145 by conjugal transfer. The constructed strain was subsequently grown on solid MS medium without apramycin at 37 °C to remove the plasmid pKCCas9-*bldKA-KE*. The correct double crossover colonies were verified with the primers *bldKA-KE*-J-F/R, yielding the  $\Delta$ *bldKA-KE* mutant strain.

Using M145 genomic DNA as the template, the *bldKA-KE* from M145 and *ORF\_2669-ORF\_2673* from YIM 93972 were obtained with the primer pairs *bldKA-KE*-F/R and *ORF\_2669-ORF\_2673* respectively. The amplified products were cloned in the pIB139 vector to generate the complemented plasmids pIB139-*bldKA-KE* and pIB139-*ORF\_2669-ORF\_2673*. The obtained plasmids were introduced into the  $\Delta$ *bldKA-KE* mutant, resulting in two complemented strains, the  $\Delta$ *bldKA-KE*/pIB139-*bldKA-KE* and  $\Delta$ *bldKA-KE*/pIB139-*ORF\_2669-ORF\_2673* strains. The empty vector pIB139 was transferred into the M145 strain, generating the positive control. To evaluate the role of oligopeptide ABC

transporter *ORF\_2669-ORF\_2673* for controlling mycelium differentiation as *bldKA-KE* in *S. coelicolor*, the same viable spores of the four strains, M145/pIB139,  $\Delta$ *bldKA-KE*/pIB139,  $\Delta$ *bldKA-KE*/pIB139-*bldKA-KE* and  $\Delta$ *bldKA-KE*/pIB139-*ORF\_2669-ORF\_2673* were inoculated and grown on an ISP 4 agar plate at 37 °C. About 700 visible clones were grown on each agar plate. The substrate and aerial mycelium from a 78 h old culture of the above four constructed strains were examined by a scanning electron microscopy (Nova NanoSEM 450, FEI, USA).

### Characterization of cell wall S-layer proteins of YIM 93972

The S-layer proteins of YIM 93972 were extracted as previously described<sup>73</sup>. The mid-log phase cells were harvested by centrifugation at 3500 g, and pellets were resuspended in cell lysis buffer containing 20 mM Tris-HCl (pH 6.5), 1.0 mM phenylmethylsulfonyl fluoride (PMSF), and 1x protease inhibitor cocktail (Roche, Basel, Switzerland), and disrupted by a Soniprep sonicator (2-s on and 4-s off, amplitude 25%) for 10 min. Cell debris were removed by centrifugation at 4000 g for 15 min at 4 °C. The supernatant was further ultracentrifuged for 1 h at 250,000 g and 4 °C. Then the pellet was dissolved in a buffer containing 4% SDS, 50 mM Tris (pH 8.0) and 20 mM DTT to obtain cell wall proteins. 20 µg of cell wall proteins were reduced with 5 mM of DTT at 45 °C for 30 min and alkylated with 10 mM of IAA at room temperature for 30 min. After separated with a 12% SDS-PAGE for 5 cm, the whole lane was sliced into 25 fractions based on the distinguishing bands on the Coomassie blue G-250 stained gel. The gel pieces were further excised into 1 cubic millimeter gels and digested in-gel with trypsin (12 ng/µL) at 37 °C for 14 h. The tryptic peptides were analyzed by an LTQ-Orbitrap Velos mass spectrometer after high pH RP chromatography separation. The acquired raw files were then submitted to the MaxQuant (v1.5.6.0) against the protein database from strain YIM 93972 (3744 entries) along with common contaminant protein sequences. Two S-layer proteins were detected by peptide profile matching with the annotated protein database of YIM 93972 based on peptide numbers  $\geq 2$ , including ORF\_1400 and ORF\_1704 (Supplementary Fig. 9a–d).

### Statistics and reproducibility

The strain culture, scanning EM image, transmission EM image, were performed at least three independent experiments. Statistical differences between two groups in transcriptomic and proteomic datasets were analyzed using two-tailed unpaired *t*-tests (inter-group) and Significance  $A^{70}$  (intra-group), respectively.  $P < 0.05$  was considered statistically significant. Data statistical analysis and visualization were realized by R (v4.1.2).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The genome and transcriptome data generated in this study have been deposited in GenBank under accession codes CP071306-CP071310 and GSE168067, respectively. The raw reads of amplicon sequencing are deposited in GenBank, under accession number PRJNA703326. All proteomic MS data sets can be obtained from iProX<sup>74</sup> database with identifier IPX0002724000, or ProteomeXchange with identifier PXD023481. Raw reads of genome for strains YIM\_A00010, YIM\_A00011, YIM\_A00012, YIM\_A00013, and YIM\_A00014 were deposited in GenBank under accession codes JAKCFJ0000000000, JAKCFK0000000000, JAKCFL0000000000, JAKCFM0000000000, and JAKCFN0000000000, respectively. The complete Halobacterial COG data archive (supplementary\_data\_file\_1.tgz) and the phylogenetic trees and alignments archive (supplementary\_data\_file\_2.tgz) can be downloaded from [https://ftp.ncbi.nih.gov/pub/wolf/\\_suppl/halo22/](https://ftp.ncbi.nih.gov/pub/wolf/_suppl/halo22/).

## References

- Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *P. Natl Acad. Sci.* **74**, 5088–5090 (1977).
- Brock, T. D., Brock, K. M., Belly, R. T. & Weiss, R. L. *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* **84**, 54–68 (1972).
- Ferry, J. G., Smith, P. H. & Wolfe, R. S. *Methanospirillum*, a New Genus of Methanogenic Bacteria, and Characterization of *Methanospirillum hungatii* sp.nov. *Int. J. Syst. Bacteriol.* **24**, 465–469 (1974).
- Zillig, W. et al. The Archaeobacterium *Thermofilum pendens* Represents, a Novel Genus of the Thermophilic, Anaerobic Sulfur Respiring Thermoproteales. *Syst. Appl. Microbiol.* **4**, 79–87 (1983).
- Janekovic, D. et al. TTV1, TTV2 and TTV3, a family of viruses of the extremely thermophilic, anaerobic, sulfur reducing archaeobacterium *Thermoproteus tenax*. *Mol. Gen. Genet.* **192**, 39–45 (1983).
- Völkl, P. et al. *Pyrobaculum aerophilum* sp. nov., a novel nitrate-reducing hyperthermophilic archaeum. *Appl. Environ. Microbiol.* **59**, 2918–2926 (1993).
- Burns, D. G. et al. *Haloquadratum walsbyi* gen. nov., sp. nov., the square haloarchaeon of Walsby, isolated from saltern crystallizers in Australia and Spain. *Int. J. Syst. Evol. Microbiol.* **57**, 387–392 (2007).
- Sowers, K. R., Baron, S. F. & Ferry, J. G. *Methanosarcina acetivorans* sp. nov., an Acetotrophic Methane-Producing Bacterium Isolated from Marine Sediments. *Appl. Environ. Microbiol.* **47**, 971–978 (1984).
- Kern, T., Fischer, M. A., Deppenmeier, U., Schmitz, R. A. & Rother, M. *Methanosarcina flavescens* sp. nov., a methanogenic archaeon isolated from a full-scale anaerobic digester. *Int. J. Syst. Evol. Microbiol.* **66**, 1533–1538 (2016).
- Claessen, D., Rozen, D. E., Kuipers, O. P., Sogaard-Andersen, L. & van Wezel, G. P. Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies. *Nat. Rev. Microbiol.* **12**, 115–124 (2014).
- Liu, B. B. et al. Description of *Halegenticoccus soli* gen. nov., sp. nov., a halophilic archaeon isolated from a soil sample of Ebi lake. *Extremophiles* **23**, 521–528 (2019).
- Nicholson, W. L. et al. Bacterial endospores and their significance in stress resistance. *Antonie Van. Leeuwenhoek.* **81**, 27–32 (2002).
- Peck, R. F., Plesa, A. M., Graham, S. M., Angelini, D. R. & Shaw, E. L. Opsin-Mediated Inhibition of Bacterioruberin Synthesis in Halophilic Archaea. *J. Bacteriol.* **199**, e00303–e00317 (2017).
- Peck, R. F., Graham, S. M. & Gregory, A. M. Species Widely Distributed in Halophilic Archaea Exhibit Opsin-Mediated Inhibition of Bacterioruberin Biosynthesis. *J. Bacteriol.* **201**, e00576–00518 (2019).
- Verma, A., Pal, Y., Kumar, P. & Krishnamurthi, S. *Halocatena pleomorpha* gen. nov. sp. nov., an extremely halophilic archaeon of family *Halobacteriaceae* isolated from saltpan soil. *Int. J. Syst. Evol. Microbiol.* **70**, 3693–3700 (2020).
- Inoue, K., Itoh, T., Ohkuma, M. & Kogure, K. *Halomarina oriensis* gen. nov., sp. nov., a halophilic archaeon isolated from a seawater aquarium. *Int. J. Syst. Evol. Microbiol.* **61**, 942–946 (2011).
- Torreblanca, M. et al. Classification of non-alkaliphilic halobacteria based on numerical taxonomy and polar lipid composition, and description of *Haloarcula* gen. nov. and *Haloferax* gen. nov. *Syst. Appl. Microbiol.* **8**, 89–99 (1986).
- Hwang, H. B. et al. *Haloplanus salinarum* sp. nov., an extremely halophilic archaeon isolated from a solar saltern. *Int. J. Syst. Evol. Microbiol.* **67**, 4456–4461 (2017).
- Duggin, I. G. et al. Cetz tubulin-like proteins control archaeal cell shape. *Nature* **519**, 362–365 (2015).

20. Makarova, K. S., Koonin, E. V. & Albers, S. V. Diversity and Evolution of Type IV pili Systems in Archaea. *Front. Microbiol.* **7**, 667 (2016).
21. Shi, H., Bratton, B. P., Gitai, Z. & Huang, K. C. How to Build a Bacterial Cell: MreB as the Foreman of *E. coli* Construction. *Cell* **172**, 1294–1305 (2018).
22. Moir, D., Stewart, S. E., Osmond, B. C. & Botstein, D. Cold-sensitive cell-division-cycle mutants of yeast: isolation, properties, and pseudoreversion studies. *Genetics* **100**, 547–563 (1982).
23. Barthelme, D. & Sauer, R. T. Origin and Functional Evolution of the Cdc48/p97/VCP AAA+ Protein Unfolding and Remodeling Machine. *J. Mol. Biol.* **428**, 1861–1869 (2016).
24. Guillemet, E. et al. The InhA metalloproteases of *Bacillus cereus* contribute concomitantly to virulence. *J. Bacteriol.* **192**, 286–294 (2010).
25. Sleytr, U. B., Schuster, B., Egelseer, E. M. & Pum, D. S-layers: principles and applications. *FEMS Microbiol. Rev.* **38**, 823–864 (2014).
26. Rodrigues-Oliveira, T., Belmok, A., Vasconcellos, D., Schuster, B. & Kyaw, C. M. Archaeal S-Layers: Overview and Current State of the Art. *Front. Microbiol.* **8**, 2597 (2017).
27. Eaton, D. & Ensign, J. C. *Streptomyces viridochromogenes* spore germination initiated by calcium ions. *J. Bacteriol.* **143**, 377–382 (1980).
28. McBride, M. J. & Zusman, D. R. Trehalose accumulation in vegetative cells and spores of *Myxococcus xanthus*. *J. Bacteriol.* **171**, 6383–6386 (1989).
29. Setlow, B., Atluri, S., Kitchel, R., Koziol-Dube, K. & Setlow, P. Role of dipicolinic acid in resistance and stability of spores of *Bacillus subtilis* with or without DNA-protective alpha/beta-type small acid-soluble proteins. *J. Bacteriol.* **188**, 3740–3747 (2006).
30. Matias, F., Bonatto, D., Padilla, G., Rodrigues, M. F. & Henriques, J. A. Polyhydroxyalkanoates production by actinobacteria isolated from soil. *Can. J. Microbiol.* **55**, 790–800 (2009).
31. Kaur, G., Burroughs, A. M., Iyer, L. M. & Aravind, L. Highly regulated, diversifying NTP-dependent biological conflict systems with implications for the emergence of multicellularity. *Elife* **9**, e52696 (2020).
32. Nodwell, J. R. & Losick, R. Purification of an extracellular signaling molecule involved in production of aerial mycelium by *Streptomyces coelicolor*. *J. Bacteriol.* **180**, 1334–1337 (1998).
33. Akanuma, G., Ueki, M., Ishizuka, M., Ohnishi, Y. & Horinouchi, S. Control of aerial mycelium formation by the BldK oligopeptide ABC transporter in *Streptomyces griseus*. *FEMS Microbiol. Lett.* **315**, 54–62 (2011).
34. Nodwell, J. R. & McGovern, K. & Losick, R. J. M. m. An oligopeptide permease responsible for the import of an extracellular signal governing aerial mycelium formation in *Streptomyces coelicolor*. *Mol. Microbiol.* **22**, 881–893 (1996).
35. Zacharia, V. M. et al. Genetic Network Architecture and Environmental Cues Drive Spatial Organization of Phenotypic Division of Labor in *Streptomyces coelicolor*. *mBio* **12**, e00794–00721 (2021).
36. Bush, M. J., Bibb, M. J., Chandra, G., Findlay, K. C. & Buttner, M. J. Genes required for aerial growth, cell division, and chromosome segregation are targets of WhiA before sporulation in *Streptomyces venezuelae*. *mBio* **4**, e00684–00613 (2013).
37. Flores, E., Picossi, S., Valladares, A. & Herrero, A. Transcriptional regulation of development in heterocyst-forming cyanobacteria. *Biochim. Biophys. Acta Gene Regul. Mech.* **1862**, 673–684 (2019).
38. Fimlaid, K. A. & Shen, A. Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes. *Curr. Opin. Microbiol.* **24**, 88–95 (2015).
39. Tortosa, P., Albano, M. & Dubnau, D. Characterization of *ylbF*, a new gene involved in competence development and sporulation in *Bacillus subtilis*. *Mol. Microbiol.* **35**, 1110–1119 (2000).
40. Carabetta, V. J. et al. A complex of YlbF, YmcA and YaaT regulates sporulation, competence and biofilm formation by accelerating the phosphorylation of SpoOA. *Mol. Microbiol.* **88**, 283–300 (2013).
41. Moreira, R. N. et al. BolA Is Required for the Accurate Regulation of c-di-GMP, a Central Player in Biofilm Formation. *mBio* **8**, e00443–00417 (2017).
42. Konstantinidis, K. T., Rossello-Mora, R. & Amann, R. Uncultivated microbes in need of their own taxonomy. *ISME J.* **11**, 2399–2406 (2017).
43. Schubert, B. A., Lowenstein, T. K. & Timofeeff, M. N. Microscopic identification of prokaryotes in modern and ancient halite, Saline Valley and Death Valley, California. *Astrobiology* **9**, 467–482 (2009).
44. He, Q. et al. Effects of phenol on extracellular polymeric substances and microbial communities from aerobic granular sludge treating low strength and salinity wastewater. *Sci. Total Environ.* **752**, 141785 (2021).
45. Gong, L. et al. Complete genome sequencing of the luminescent bacterium, *Vibrio qinghaiensis* sp. Q67 using PacBio technology. *Sci. Data* **5**, 170205 (2018).
46. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
47. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
48. Lagesen, K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucl. Acids Res.* **35**, 3100–3108 (2007).
49. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucl. Acids Res.* **25**, 955–964 (1997).
50. Carver, T., Thomson, N., Bleasby, A., Berriman, M. & Parkhill, J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* **25**, 119–120 (2009).
51. Steinegger, M. & Soding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
52. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinforma.* **5**, 113–132 (2004).
53. Zimmermann, L. et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
54. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
55. Marchler-Bauer, A. et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucl. Acids Res.* **45**, D200–D203 (2017).
56. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Archaeal Clusters of Orthologous Genes (arCOGs): An Update and Application for Analysis of Shared Features between Thermococcales, Methanococcales, and Methanobacteriales. *Life (Basel)* **5**, 818–840 (2015).
57. Styczynski, M. P., Jensen, K. L., Rigoutsos, I. & Stephanopoulos, G. BLOSUM62 miscalculations improve search performance. *Nat. Biotechnol.* **26**, 274–275 (2008).
58. Esterman, E. S., Wolf, Y. I., Kogay, R., Koonin, E. V. & Zhaxybayeva, O. Evolution of DNA packaging in gene transfer agents. *Virus Evol.* **7**, veab015 (2021).
59. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
60. Rinke, C. et al. A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nat. Microbiol.* **6**, 946–959 (2021).

61. Cohen, O., Ashkenazy, H., Belinky, F., Huchon, D. & Pupko, T. GLOOME: gain loss mapping engine. *Bioinformatics* **26**, 2914–2915 (2010).
62. Manteca, A., Fernandez, M. & Sanchez, J. A death round affecting a young compartmentalized mycelium precedes aerial mycelium dismantling in confluent surface cultures of *Streptomyces antibioticus*. *Microbiology* **151**, 3689–3697 (2005).
63. Yang, Y., Yin, X., Yang, H. & Xu, Y. Histone demethylase LSD2 acts as an E3 ubiquitin ligase and inhibits cancer cell growth through promoting proteasomal degradation of OGT. *Mol. Cell* **58**, 47–59 (2015).
64. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
65. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinforma.* **12**, 323–339 (2011).
66. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
67. Zhang, Y. et al. Quantitative proteomics reveals membrane protein-mediated hypersaline sensitivity and adaptation in halophilic *Nocardiopsis xinjiangensis*. *J. Proteome Res.* **15**, 68–85 (2016).
68. Xu, P., Duong, D. M. & Peng, J. Systematical optimization of reverse-phase chromatography for shotgun proteomics. *J. Proteome Res.* **8**, 3944–3950 (2009).
69. Wu, F. et al. Recombinant acetylated trypsin demonstrates superior stability and higher activity than commercial products in quantitative proteomics studies. *Rapid Commun. Mass Sp.* **30**, 1059–1066 (2016).
70. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
71. Kieser, T., Foundation, J. I., Bibb, M. J. & Buttner, M. J. *Practical Streptomyces Genetics*. (John Innes Foundation, Norwich, 2000).
72. Huang, H., Zheng, G., Jiang, W., Hu, H. & Lu, Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta Biochim. Biophys. Sin. (Shanghai)* **47**, 231–243 (2015).
73. Li, L. et al. Enhanced glycosylation of an S-layer protein enables a psychrophilic methanogenic archaeon to adapt to elevated temperatures in abundant substrates. *FEBS Lett.* **594**, 665–677 (2020).
74. Ma, J. et al. iProX: an integrated proteome resource. *Nucl Acids Res.* **47**, D1211–D1217 (2019).
75. Kumar, L. & M, E. F. Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* **2**, 5–7 (2007).
- Province (202002AA100007). K.S.M., Y.I.W. and E.V.K. are supported by intramural funds of the US Department of Health and Human Services (National Institutes of Health, National Library of Medicine).

## Author contributions

P.X., S.K.T., E.V.K and G.P.Z. designed experiments. K.L. and Y.W. collected the environmental samples. S.K.T. and B.B.L. performed the strain isolation, polyphasic taxonomic analysis, culture of YIM 93972 and all the other 6 morphological differentiation haloarchaeon strains, and SSU rRNA verification with the help of Y.R.Z., E.Y.L., Y.Z.F., M.X.X., Z.Q.L., X.Z., R.L., M.Y., L.L.Y., T.W.G., H.L.C., Z.K.Z. and T.S.T. M.X.X. and Z.Q.L. performed heat stress experiment under the direction of T.S.T., P.X., Y.Z. and S.K.T. X.M.C performed the mutagenesis experiments of YIM 93972. H.J.Z. performed the genomic sequencing of the wild and mutated strains. X.Y.Z., K.S.M. and Y.I.W. performed the genome annotation and phylogenomic analysis. Y.Z. performed all of the proteomic experiments and data analysis under the help of S.H.J., T.Z., P.R.C., J.H.S., C.C., L.C. and H.Y.G. X.Y.Z. and Y.Z. performed transcriptomic analysis, and arranged all figures and tables. P.X., Z.P.Z., Y.Z. and K.S.M. performed transcriptional regulator analysis. Z.P.Z. and S.H.J. helped data analysis and uploading. Y.H.L. and G.S.Z. constructed all the recombinant plasmids and gene deletion strains for gene complementation experiments. F.C.H. helped article frame arrangement about six cellular differentia-tional halophilic archaea. P.X., Y.Z., G.S.Z., Y.H.L. and S.K.T. designed and performed spore-genetic gene cluster verification with the help of G.P.Z., E.V.K. and K.S.M. X.Y.Z., Y.Z., K.S.M., P.X., E.V.K. and G.P.Z. wrote the manuscript with the help of all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-37389-w>.

**Correspondence** and requests for materials should be addressed to Shu-Kun Tang, Eugene V. Koonin, Guo-Ping Zhao or Ping Xu.

**Peer review information** *Nature Communications* thanks William Whitman and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

## Acknowledgements

We are indebted to Dr. Wen-Jun Li for support in the early stage of this project. The authors thank Dr. Xiu-Zhu Dong, Yong-Jun Liu, Lei Song and Man Cai for strains, reagents, and discussion. We thank Dr. Lingyan Li for discussion of S-layer protein extraction. We thank Professor Aharon Oren for the etymology of new generic name and of new epithet, and Professors Leonard Krall, Felix Cheung and Xu-Na Wu for critical reading and editing. We also thank Shu-Jia Wu, Jin-Shuai Sun, Wen-Hui Wu, Jie Ma and Xue Wang for experiments and data processing. P.X., G.P.Z., S.K.T., X.Y.Z., Y.Z., B.B.L. and F.C.H. are supported by the Chinese National Basic Research Programs (2022YFA1304600 and 2020YFE0202200), the Innovation Foundation of Medicine (2022F12010, 20SQAQX34 and AWS17J008), the National Natural Science Foundation of China (32141003, 31901037, 31870824, 91839302, 32071431, 31760003, 32060003, 32070668, 92151001 and 31800001), the Beijing-Tianjin-Hebei Basic Research Cooperation Project (J200001), CAMS Innovation Fund for Medical Sciences (2019-I2M-5-017, 2019-12M-5-063 and 2022-I2M-C&T-B-082), the Foundation of State Key Lab of Proteomics (SKLP-C202002, 2021-NCPSB-001, SKLP-K201704, & SKLP-K201901) and Major Science and Technology Projects of Yunnan

<sup>1</sup>Yunnan Institute of Microbiology, Key Laboratory for Microbial Resources of the Ministry of Education, School of Life Sciences, Yunnan University, Kunming 650091, China. <sup>2</sup>State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences Beijing, Research Unit of Proteomics & Research and Development of New Drug, Research Unit of Proteomics Driven Cancer Precision Medicine, Chinese Academy of Medical Sciences, Institute of Lifeomics, Beijing 102206, China. <sup>3</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, MD 20894, USA. <sup>4</sup>Henan Key Laboratory of Industrial Microbial Resources and Fermentation Technology, College of Biological and Chemical Engineering, Nanyang Institute of Technology, Nanyang 473004, China. <sup>5</sup>College of Life Sciences, Shanghai Normal University, Shanghai 200234, China. <sup>6</sup>Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai and Shanghai Institute for Biomedical and Pharmaceutical Technologies, Shanghai 201203, China. <sup>7</sup>Hebei Province Key Lab of Research and Application on Microbial Diversity, College of Life Sciences, Hebei University, Hebei 071002, China. <sup>8</sup>Key Laboratory of Desert and Desertification, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China. <sup>9</sup>Xinjiang Institute of Microbiology, Xinjiang Academy of Agricultural Science, Urumqi 830091, China. <sup>10</sup>Biotechnology and Genetic Germplasm Resources Research Institute, Yunnan Academy of Agricultural Sciences, Kunming 650205, China. <sup>11</sup>Department of Microbiology, Key Laboratory of Combinatorial Biosynthesis and Drug Discovery of Ministry of Education, School of Pharmaceutical Sciences, Wuhan University, Wuhan 430072, China. <sup>12</sup>College of Food and Biological Engineering, Xihua University, Chengdu 610039, China. <sup>13</sup>School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China. <sup>14</sup>Key Laboratory of Medical Molecular Virology (MOE/NHC/CAMS), School of Basic Medical Sciences and Department of Microbiology and Microbial Engineering, School of Life Sciences, Fudan University, Shanghai 200032, China. <sup>15</sup>Guizhou University, School of Medicine, Guiyang 550025, China. <sup>16</sup>State Key Laboratory of Dampness Syndrome of Chinese Medicine, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 510120, China. <sup>17</sup>These authors contributed equally: Shu-Kun Tang, Xiao-Yang Zhi, Yao Zhang, Kira S. Makarova, Bing-Bing Liu. ✉e-mail: [tangshukun@ynu.edu.cn](mailto:tangshukun@ynu.edu.cn); [koonin@ncbi.nlm.nih.gov](mailto:koonin@ncbi.nlm.nih.gov); [gpzhaos@sibs.ac.cn](mailto:gpzhaos@sibs.ac.cn); [xuping@ncpsb.org.cn](mailto:xuping@ncpsb.org.cn)