

Supplementary Materials for

Diverse quorum sensing systems regulate microbial communication and biogeochemical processes in deep-sea cold seeps

Jiaxue Peng^{1,2#}, Xinyue Liu^{1#}, Jieni Wang¹, Nan Meng², Runlin Cai³,
Yongyi Peng^{1,4}, Yingchun Han¹, Jing Liao¹, Chengcheng Li^{5,6},
Maxim Rubin-Blum^{7,8}, Qiao Ma^{2*}, Xiyang Dong^{1,9*}

¹ Key Laboratory of Marine Genetic Resources, Third Institute of Oceanography, Ministry of Natural Resources, Xiamen, 361005, China

² College of Environmental Science and Engineering, Dalian Maritime University, Dalian, 116026, China

³ China Guangdong Provincial Key Laboratory of Marine Biology, Shantou University, Shantou, 515063, China

⁴ Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia

⁵ Qingdao New Energy Shandong Laboratory, Key Laboratory of Biofuels, Shandong Provincial Key Laboratory of Energy Genetics, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, 266101, China

⁶ Shandong Energy Institute, Qingdao, 266101, China

⁷ Israel Oceanographic and Limnological Research, National Institute of Oceanography, Haifa, 3108000, Israel

⁸ Charney School of Marine Sciences, University of Haifa, Haifa, 3498838, Israel

⁹ Lead contact

These authors contributed equally to this work.

* Corresponding authors:

Xiyang Dong (dongxiyang@tio.org.cn); Qiao Ma (xiaoma0556@dlmu.edu.cn)

Supporting Methods

The identification of quorum sensing genes from the non-redundant gene catalog with 147 million genes and 3,813 species-level representative metagenome-assembled genomes (MAGs) is based on the quorum sensing-related proteins (QSP) database [1], combined with multiple screening approaches. Detailed information is provided below (**Supplementary Table 1**):

Identification of quorum sensing genes associated with the AI-2 system

For the 3,813 species-level representative MAGs, phylogenetic trees of the amino acid sequences of LsrC, LsrD, LsrG, LsrR, LuxS, CahR, LuxP, and LsrB were constructed to validate their phylogenetic clades against reference sequences (**Supplementary Fig. 1**). Sequences were aligned using MAFFT (v7.505, default settings) [2] and trimmed using TrimAL (v1.4.1, default settings) [3]. Maximum-likelihood trees were constructed using FastTree (v2.1.11, default settings) [4], and all trees were visualized and beautified using iTOL (v6) [5]. Candidate LuxP and LuxS sequences that clustered poorly with reference sequences further validated using AlphaFold2 [6]. These structures were compared with the structural similarity clustering results obtained from Foldseek (<https://cluster.foldseek.com/>) [7], and assessments were performed using the Template Modeling Score (TM-score) [8] and Root Mean Square Deviation (RMSD) [9]. Sequences were considered valid if they met the criteria of TM-score > 0.5 [10] and RMSD < 1 (**Supplementary Fig. 1h-1i**). The comparative results were visualized using PyMOL [11]. Additionally, LsrR sequences were extracted using HMMER (v3.3.2) with the profile “PF04198” [12] and LsrK sequences using profiles “PF00370” and “PF02782” [13]. For CahR, sequences were extracted using HMMER (v3.3.2) with the profile “PF02743” [14]. Candidate CahR and LsrA sequences were further analyzed for domain structure using the Pfam database (<https://www.ebi.ac.uk/interpro/>), confirming the ATP-binding cassette domain in LsrA [15].

For the non-redundant gene catalog with 147 million genes, the CahR, LsrA and LsrK

were screened using the same method as the genome-based analysis. LuxS, LuxP, LsrB, LsrC, LsrD, LsrG, LsrR and LuxQ were identified based on their phylogenetic clades with reference genes (**Supplementary Fig. 2**). The software and parameters used for constructing the phylogenetic tree are the same as those used in the genome-based analysis.

Identification of quorum sensing genes associated with the AHLs system

For the 3,813 species-level representative MAGs, phylogenetic trees of HdtS, LuxM, LuxI, LuxN and AHL-acylase were constructed to validate their phylogenetic clades against reference sequences (**Supplementary Fig. 3**). Sequences were aligned using MAFFT (v7.505, default settings) [2] and trimmed using TrimAL (v1.4.1, default settings) [3]. Maximum-likelihood trees were constructed using FastTree (v2.1.11, default settings) [4], and all trees were visualized and beautified using iTOL (v6) [5]. For HdtS, sequences containing conserved motifs NHQS and PEGTR were retained [16]. For LuxR, sequences containing both the autoinducer-binding domain (PF03472) and the GerE domain (PF00196) were extracted [17]. Sequences of LuxI were identified using HMMER (v3.3.2) with the profile “PF00765” [17]. For AHL-acylase, sequences containing the key amino acid Ser217 were retained [18], and sequences containing the conserved motif HXHXDH-60aa-H were selected as AHL-lactonase [19]. Conserved sites were visualized using MAFFT (v7.505, default settings) [2] and Jalview [20].

For the non-redundant gene catalog with 147 million genes, LuxI, LuxM, LuxN, HdtS, AHL-acylase and AHL-lactonase were selected based on their phylogenetic clades against reference sequences. The LuxR was screened using the same method as in genome-based analysis, then constructing a phylogenetic tree with reference sequences (**Supplementary Fig. 4**). The software and parameters used for constructing the phylogenetic tree are the same as those used for the genome-based analysis.

Identification of quorum sensing genes associated with the DSF system

For the 3,813 species-level representative MAGs, RpfF sequences were identified using HMMER (v3.3.2) with profiles “PF00378” and “PF16113” [21], and sequences containing conserved sites Glu114 and Glu164 were selected [22]. Putative RpfC sequences were predicted using the SMART database [23], retaining those sequences with transmembrane domains, HisKA, HATPase, REC, and HPT domains were retained [24] (**Supplementary Fig. 5a**). For RpfG, sequences containing the HD domain and conserved GYP motif were retained using SMART [25]. Conserved sites were visualized using MAFFT (v7.505, default settings) [2] and Jalview [20].

For the non-redundant gene catalog with 147 million genes, RpfF and RpfG were screened using the same method as in genome-based analysis (**Supplementary Fig. 5a**), RpfC and RpfB were selected based on validating their phylogenetic clades against reference sequences (**Supplementary Fig. 5b-c**). Sequences were aligned using MAFFT (v7.505, default settings) [2] and trimmed using TrimAL (v1.4.1, default settings) [3]. Maximum-likelihood trees were constructed using FastTree (v2.1.11, default settings) [4], and all trees were visualized and beautified using iTOL (v6) [5].

Identification of quorum sensing genes associated with the c-di-GMP system

For the 3,813 species-level representative MAGs, DGC sequences were identified using HMMER (v3.3.2) with the profile “PF00990”, and sequences were retained if they contained conserved sites D327, N335, and D344, as well as the inhibitory I-site motif RxxD (**Supplementary Fig. 6**). PDE sequences were identified using HMMER (v3.3.2) with the profile “PF00563” or by detecting the conserved HD-GYP motif [26]. A phylogenetic tree was constructed to further validate these PDE sequences (**Fig. 6a**). Sequences were aligned using MAFFT (v7.505, default settings) [2] and trimmed using TrimAL (v1.4.1, default settings) [3]. Maximum-likelihood trees were constructed using FastTree (v2.1.11, default settings) [4], and all trees were visualized and beautified using iTOL (v6) [5]. Sequences of Clp were predicted using SMART,

retaining those sequences with the cAMP-binding domain and conserved residues D70, R166, and D170 [27]. Conserved sites were visualized using MAFFT (v7.505, default settings) [2], Jalview [20] and WebLogo3 (<https://weblogo.threeplusone.com/>).

For the non-redundant gene catalog with 147 million genes, sequences of DGC and PDE were screened using the same method as in genome-based analysis, while PDE and Clp were selected based on their phylogenetic clades against reference sequences (**Supplementary Fig. 7**). The software and parameters used for constructing the phylogenetic tree are the same as those used in the genome-based analysis.

Identification of quorum sensing genes associated with the PQS system

For the 3,813 species-level representative MAGs, phylogenetic trees of PhnA, PhnB, PqsE and PqsH were constructed to validate their phylogenetic clades against reference sequences (**Supplementary Fig. 8a-d**). Sequences were aligned using MAFFT (v7.505, default settings) [2] and trimmed using TrimAL (v1.4.1, default settings) [3]. Maximum-likelihood trees were constructed using FastTree (v2.1.11, default settings) [4], and all trees were visualized and beautified using iTOL (v6) [5]. For PqsA, sequences with the SG-TG-PK motif were retained [28], and PqsD sequences containing Cys112 were selected [29] (**Supplementary Fig. 8e**). Conserved sites were visualized using MAFFT (v7.505, default settings) [2], Jalview [20] and WebLogo3 (<https://weblogo.threeplusone.com/>). Deep learning approach AlphaFold2 [6] was applied to predict the three-dimensional structures of PqsL for validation (**Supplementary Fig. 8f**) [6]. These structures were compared with the structural similarity clustering results obtained from Foldseek (<https://cluster.foldseek.com/>) [7], and assessed using Template Modeling Score (TM-score) [8] and Root Mean Square Deviation (RMSD) [9]. Sequences were considered valid if they met the criteria of TM-score > 0.5 [10] and RMSD < 1. The comparative results were visualized using PyMOL [11].

For the non-redundant gene catalog with 147 million genes, PqsD and PqsL were screened using the same method as for genome-based analysis, while PhnA, PhnB, PqsA, PqsH and PqsE were selected based on their phylogenetic clades against reference sequences (**Supplementary Fig. 9**). The software and parameters used for constructing the phylogenetic tree are the same as those used for the genome-based analysis.

Identification of quorum sensing genes associated with the AHKs system

For the 3,813 species-level representative MAGs, phylogenetic trees of CqsA and CqsS were constructed to validate their phylogenetic clades against reference sequences (**Supplementary Fig. 10a-b**). Sequences were aligned using MAFFT (v7.505, default settings) [2] and trimmed using TrimAL (v1.4.1, default settings) [3]. Maximum-likelihood trees were constructed using FastTree (v2.1.11, default settings) [4], and all trees were visualized and beautified using iTOL (v6) [5].

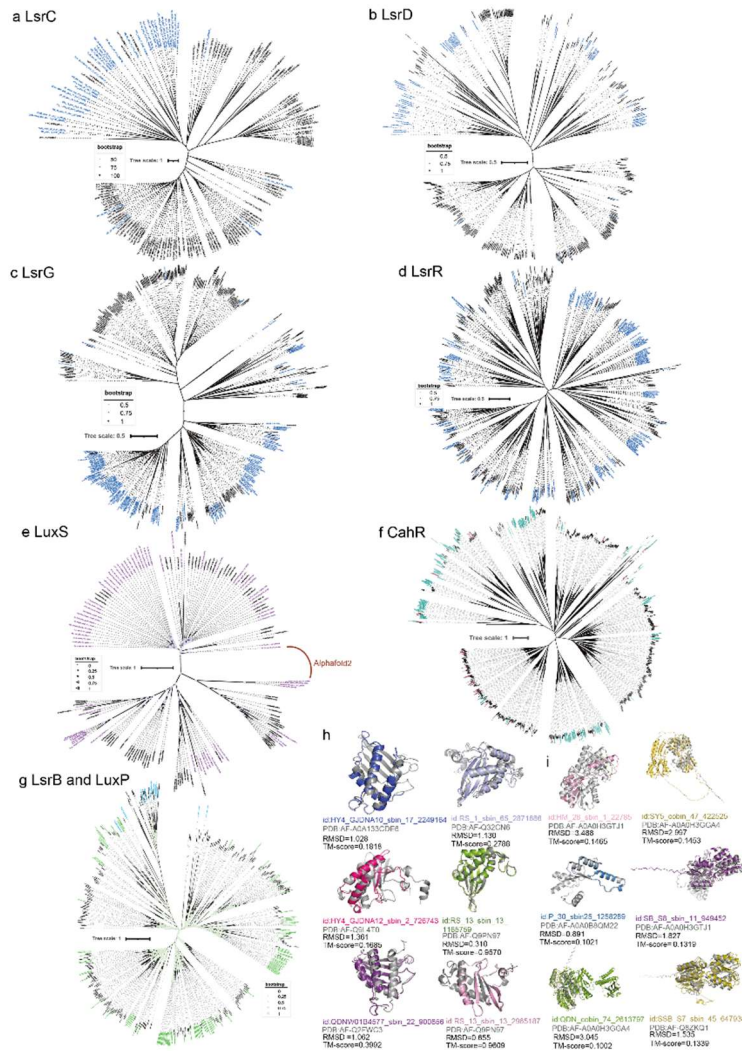
For the non-redundant gene catalog with 147 million genes, phylogenetic trees of CqsA and CqsS were constructed to validate their phylogenetic clades against reference sequences (**Supplementary Fig. 11a-b**). The software and parameters used for constructing the phylogenetic tree are the same as those used for the genome-based analysis.

References

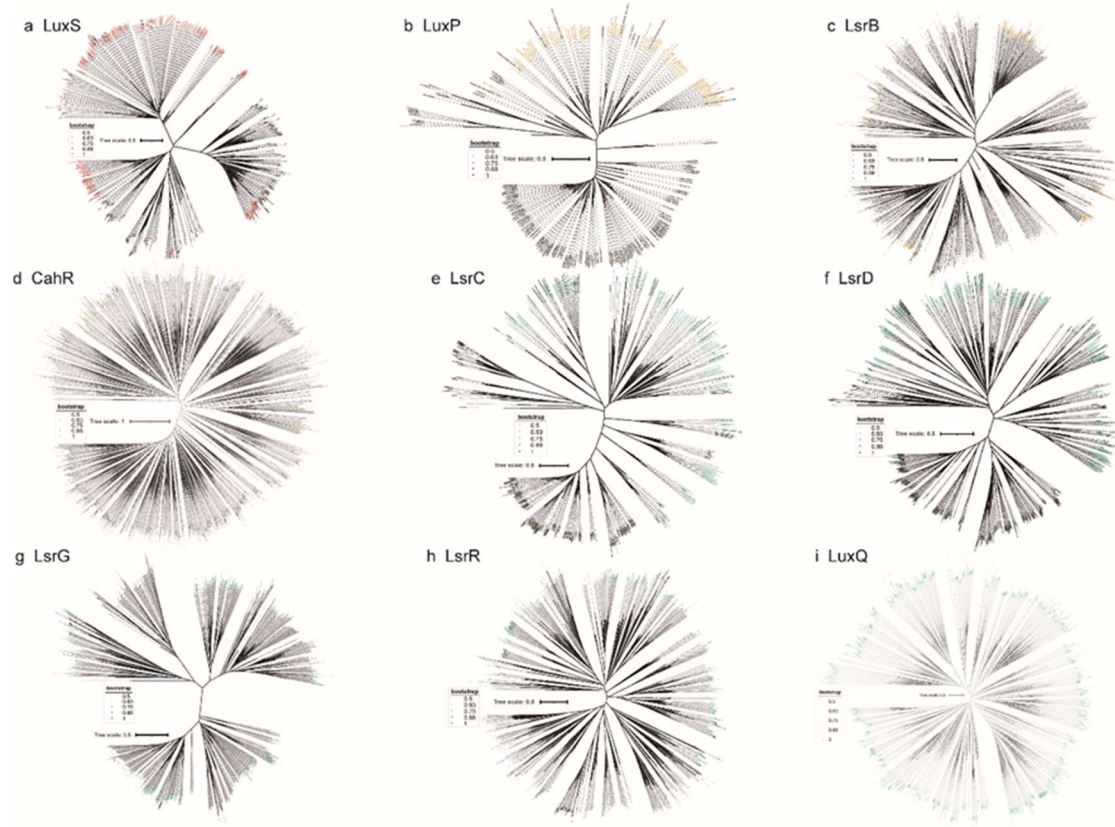
1. Dai C, Qu Y, Wu W, Li S, Chen Z, Lian S, et al.: QSP: An open sequence database for quorum sensing related gene analysis with an automatic annotation pipeline. *Water research* 2023;235:119814.
2. Nakamura T, Yamada KD, Tomii K, Katoh K: Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* 2018;34(14):2490-2492.
3. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T: trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25(15):1972-1973.
4. Price MN, Dehal PS, Arkin AP: FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular biology and evolution* 2009;26(7):1641-1650.
5. Letunic I, Bork P: Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic acids research* 2021;49(W1):W293-W296.
6. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al.: Highly accurate protein structure prediction with AlphaFold. *nature* 2021;596(7873):583-589.
7. Van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CL, et al.: Fast and accurate protein structure search with Foldseek. *Nature biotechnology* 2024;42(2):243-246.
8. Zhang Y, Skolnick J: Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics* 2007;68(4):1020.
9. Arnittali M, Rissanou AN, Harmandaris V: Structure of biomolecules through molecular dynamics simulations. *Procedia Computer Science* 2019;156:69-78.
10. Xu J, Zhang Y: How significant is a protein structure similarity with TM-score=0.5? *Bioinformatics* 2010;26(7):889-895.

11. DeLano WL: Pymol: An open-source molecular graphics tool. *CCP4 News/Protein Crystallogr* 2002;40(1):82-92.
12. Xue T, Zhao L, Sun H, Zhou X, Sun B: LsrR-binding site recognition and regulatory characteristics in *Escherichia coli* AI-2 quorum sensing. *Cell research* 2009;19(11):1258-1268.
13. Medarametla P, Gatta V, Kajander T, Laitinen T, Tammela P, Poso A: Structure-Based Virtual Screening of LsrK Kinase Inhibitors to Target Quorum Sensing. *ChemMedChem* 2018;13(22):2400-2407.
14. Liu X, Liu Q, Sun S, Sun H, Wang Y, Shen X, et al.: Exploring AI-2-mediated interspecies communications within rumen microbial communities. *Microbiome* 2022;10(1):167.
15. Pereira CS, de Regt AK, Brito PH, Miller ST, Xavier KB: Identification of functional LsrB-like autoinducer-2 receptors. *Journal of bacteriology* 2009;191(22):6975-6987.
16. West J, TOMPKINS CK, BALANTAC N, Nudelman E, MEENGs B, WHITE T, et al.: Cloning and expression of two human lysophosphatidic acid acyltransferase cDNAs that enhance cytokine-induced signaling responses in cells. *DNA and Cell Biology* 1997;16(6):691-701.
17. Tan CH, Yeo YP, Hafiz M, Ng NKJ, Subramoni S, Taj S, et al.: Functional metagenomic analysis of quorum sensing signaling in a nitrifying community. *npj Biofilms and Microbiomes* 2021;7(1):79.
18. Yasutake Y, Kusada H, Ebuchi T, Hanada S, Kamagata Y, Tamura T, et al.: Bifunctional quorum-quenching and antibiotic-acylase MacQ forms a 170-kDa capsule-shaped molecule containing spacer polypeptides. *Scientific Reports* 2017;7(1):8946.
19. Tang K, Su Y, Brackman G, Cui F, Zhang Y, Shi X, et al.: MomL, a novel marine-derived *N*-acyl homoserine lactonase from *Muricauda olearia*. *Applied and Environmental Microbiology* 2015;81(2):774-782.
20. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ: Jalview Version 2—a multiple sequence alignment editor and analysis workbench.

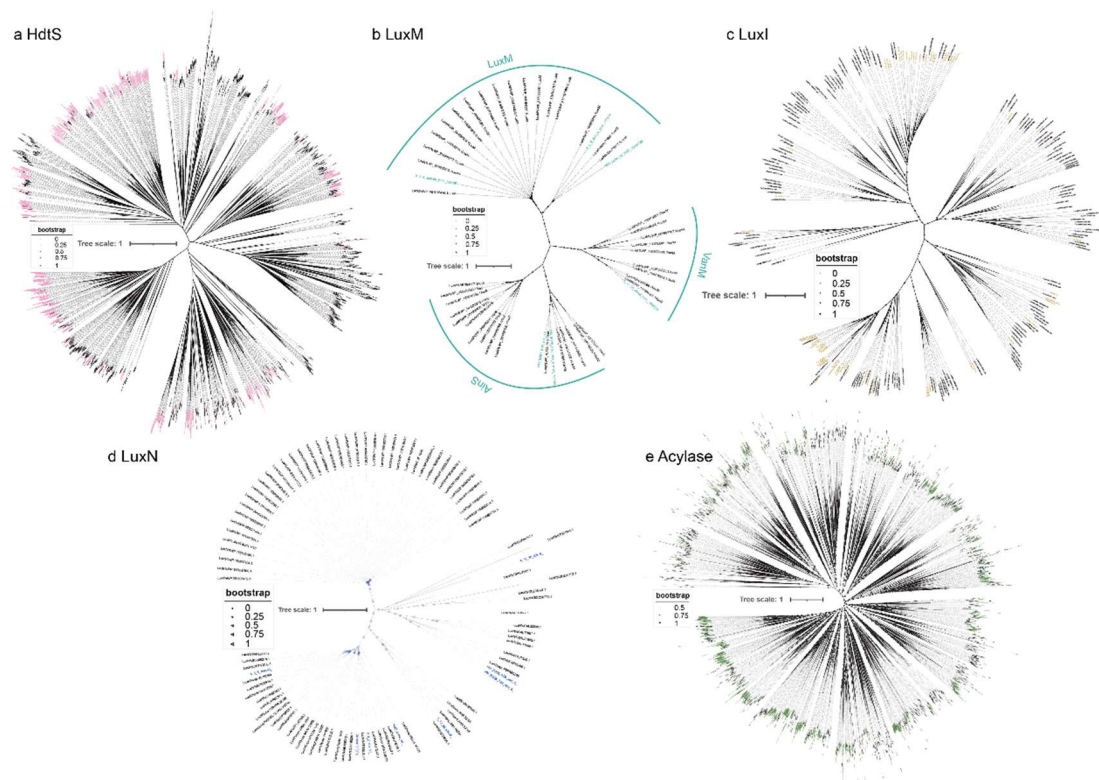
- Bioinformatics* 2009;25(9):1189-1191.
21. Ryan RP, An S-q, Allan JH, McCarthy Y, Dow JM: The DSF family of cell–cell signals: an expanding class of bacterial virulence regulators. *PLoS pathogens* 2015;11(7):e1004986.
 22. Cheng Z, He Y-W, Lim SC, Qamra R, Walsh MA, Zhang L-H, et al.: Structural basis of the sensor-synthase interaction in autoinduction of the quorum sensing signal DSF biosynthesis. *Structure* 2010;18(9):1199-1209.
 23. Letunic I, Doerks T, Bork P: SMART: recent updates, new developments and status in 2015. *Nucleic acids research* 2015;43(D1):D257-D260.
 24. He Y-W, Wang C, Zhou L, Song H, Dow JM, Zhang L-H: Dual signaling functions of the hybrid sensor kinase RpfC of *Xanthomonas campestris* involve either phosphorelay or receiver domain-protein interaction. *Journal of Biological Chemistry* 2006;281(44):33414-33421.
 25. Guo Y, Zhao Y, Tang X, Na T, Pan J, Zhao H, et al.: Deciphering bacterial social traits via diffusible signal factor (DSF)-mediated public goods in an anammox community. *Water research* 2021;191:116802.
 26. Tschowri N, Schumacher MA, Schlimpert S, babu Chinnam N, Findlay KC, Brennan RG, et al.: Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces* development. *Cell* 2014;158(5):1136-1147.
 27. Xu G, Han S, Huo C, Chin K-H, Chou S-H, Gomelsky M, et al.: Signaling specificity in the c-di-GMP-dependent network regulating antibiotic synthesis in *Lysobacter*. *Nucleic acids research* 2018;46(18):9276-9288.
 28. Coleman JP, Hudson LL, McKnight SL, Farrow III JM, Calfee MW, Lindsey CA, et al.: *Pseudomonas aeruginosa* PqsA is an anthranilate-coenzyme A ligase. *Journal of Bacteriology* 2008;190(4):1247-1255.
 29. Storz MP, Maurer CK, Zimmer C, Wagner N, Brengel C, de Jong JC, et al.: Validation of PqsD as an anti-biofilm target in *Pseudomonas aeruginosa* by development of small-molecule inhibitors. *Journal of the American Chemical Society* 2012;134(39):16143-16146.



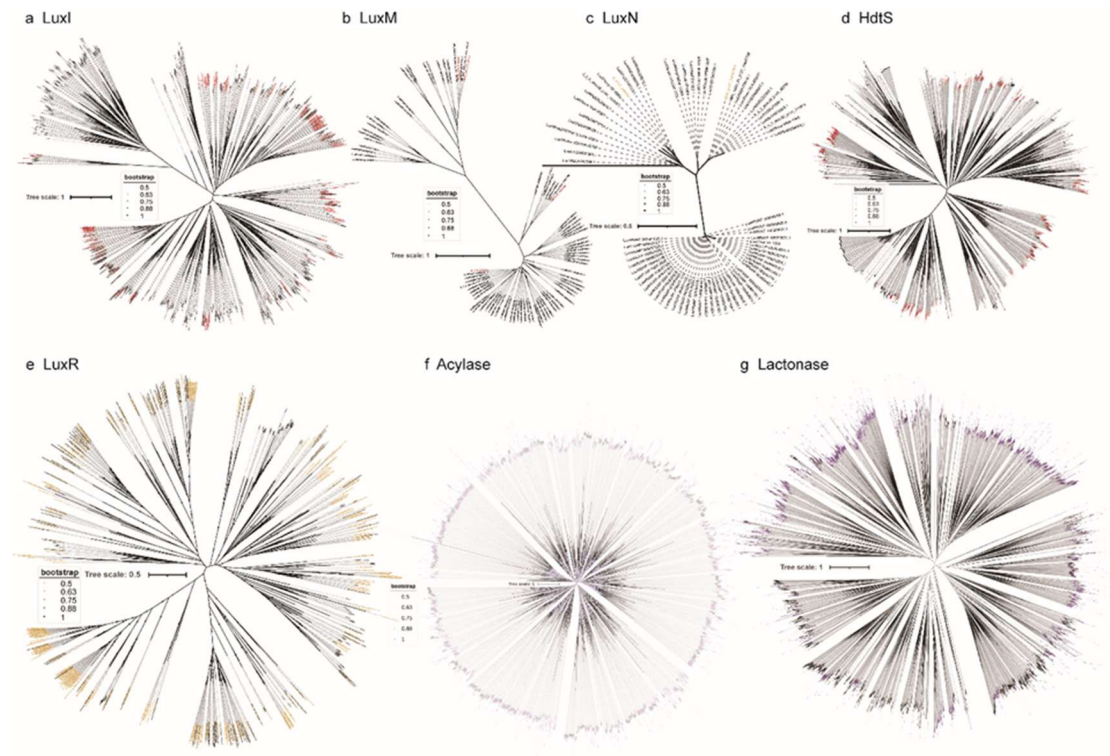
Supplementary Figure 1. Quorum sensing genes related to AI-2 system identified from cold seep microbiomes. (a-g) Maximum-likelihood phylogenetic trees of LsrC, LsrD, LsrG, LsrR, LuxS, CahR, LuxP and LsrB. Colored labels represent target genes from cold seep microbiomes, while black labels indicate reference genes. The scale bar represents the mean number of substitutions per site. The three-dimensional structures of LuxS (**h**) and LuxP (**i**) were aligned with the structures of their corresponding reference proteins for comparison. Colored are structures of target sequences that cluster poorly with reference genes in the phylogenetic tree, gray are the structures of reference proteins. TM-score and RMSD values were used to quantify the degree of structural similarity between the target and reference proteins.



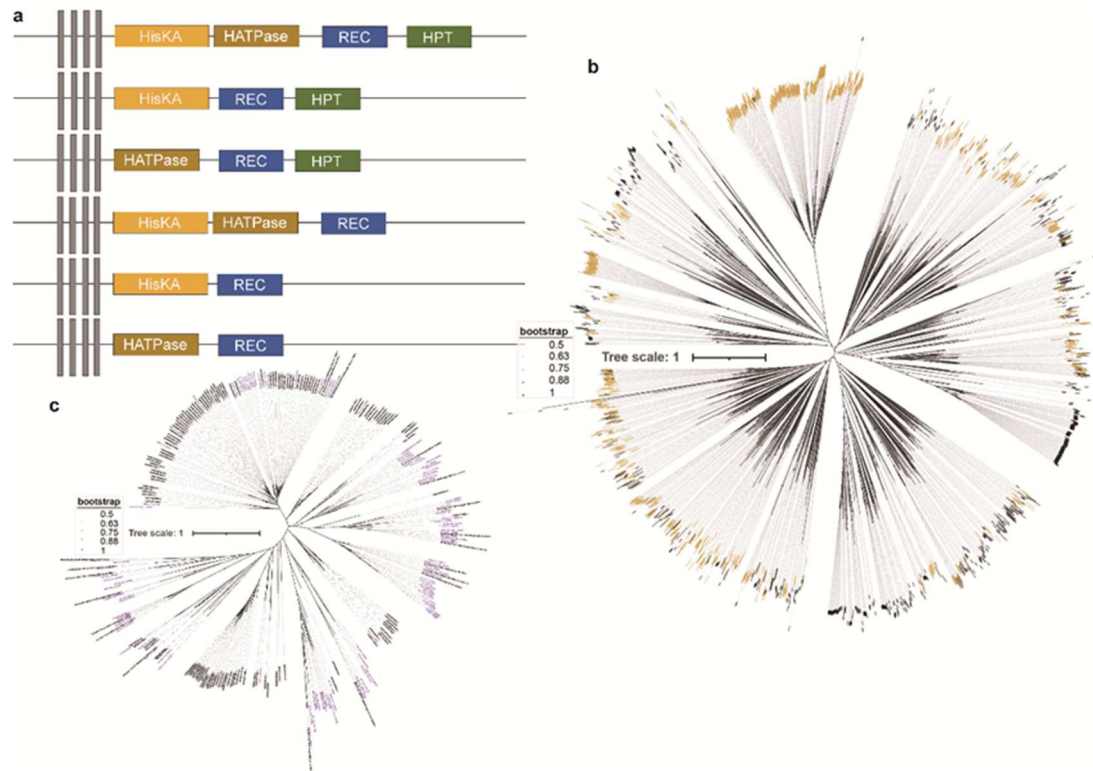
Supplementary Figure 2. Maximum-likelihood phylogenetic trees of quorum sensing genes related to AI-2 system identified from the non-redundant gene catalog. (a-i) Phylogenetic trees of LuxS, LuxP, LsrB, CahR, LsrC, LsrD, LsrG, LsrR and LuxQ. Colored labels represent target genes, while black labels indicate reference genes. The scale bar denotes the mean number of substitutions per site.



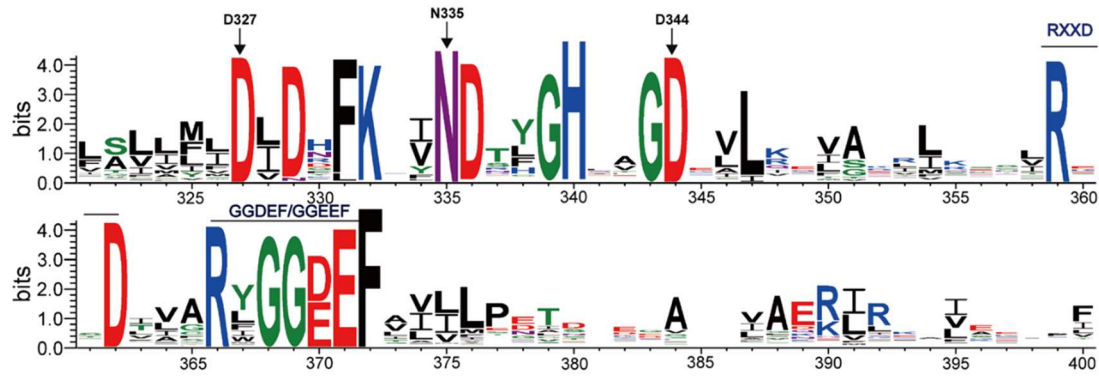
Supplementary Figure 3. Maximum-likelihood phylogenetic trees for quorum sensing genes related to AHLs system identified from cold seep microbiomes. (a-e) Phylogenetic trees of HdtS, LuxM, LuxI, LuxN, and AHL-acylase. Colored labels represent target genes from cold seep microbiomes, while black labels indicate reference genes. The scale bar indicates the mean number of substitutions per site.



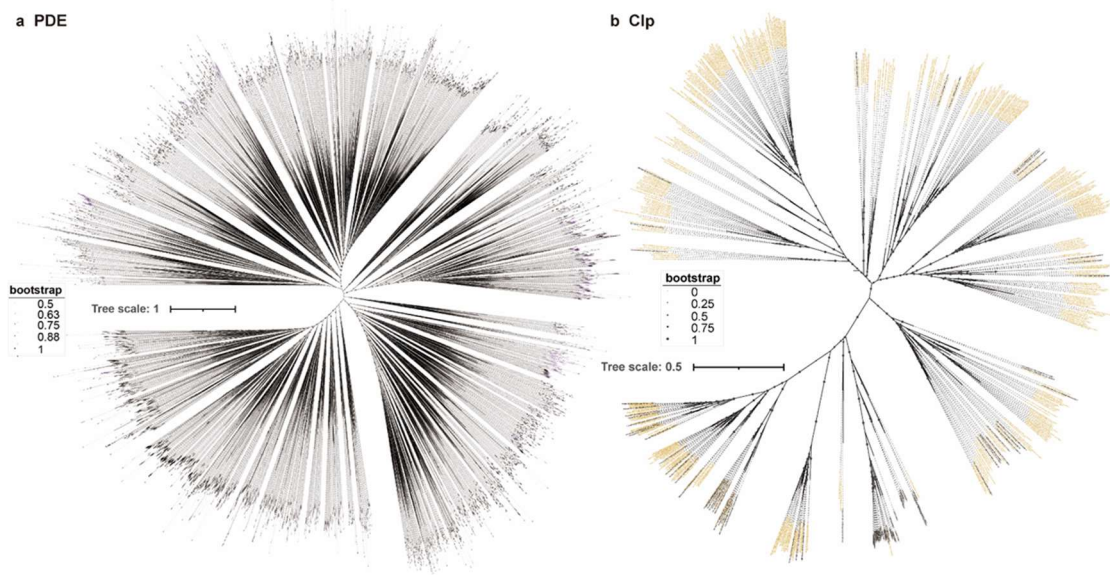
Supplementary Figure 4. Maximum-likelihood phylogenetic trees for quorum sensing genes related to AHLs system identified from the non-redundant gene catalog. (a-g) Maximum-likelihood phylogenetic trees for LuxI, LuxM, LuxN, HdtS, LuxR, AHL-acylase and AHL-lactonase. Colored labels represent target genes, while black labels indicate reference genes. The scale bar denotes the mean number of substitutions per site.



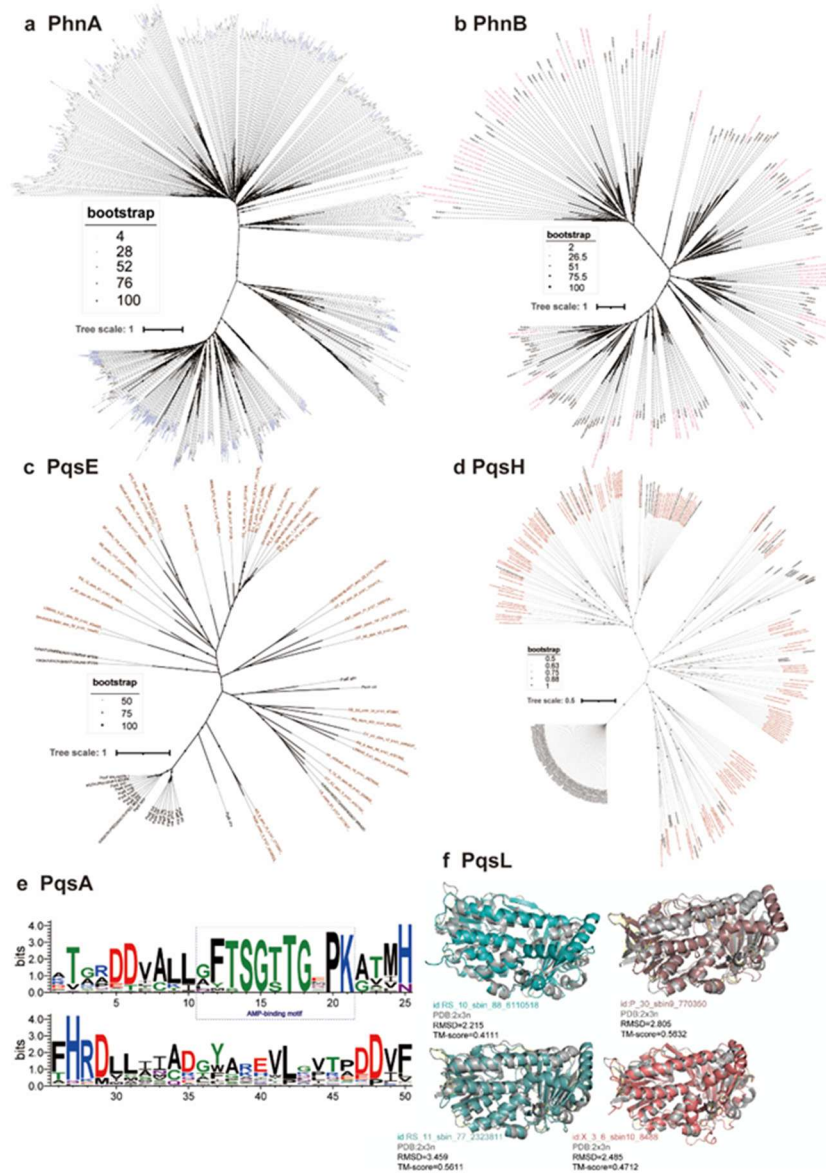
Supplementary Figure 5. Quorum sensing genes related to DSF system identified from the non-redundant gene catalog. (a) Gene structure features of *rpfC* are illustrated, showing domain arrangements such as HisKA, HATPase, REC, and HPT. (b-c) Maximum-likelihood phylogenetic trees for RpfC and RpfB. Colored labels represent target genes, while black labels indicate reference genes. The scale bar indicates the mean number of substitutions per site.



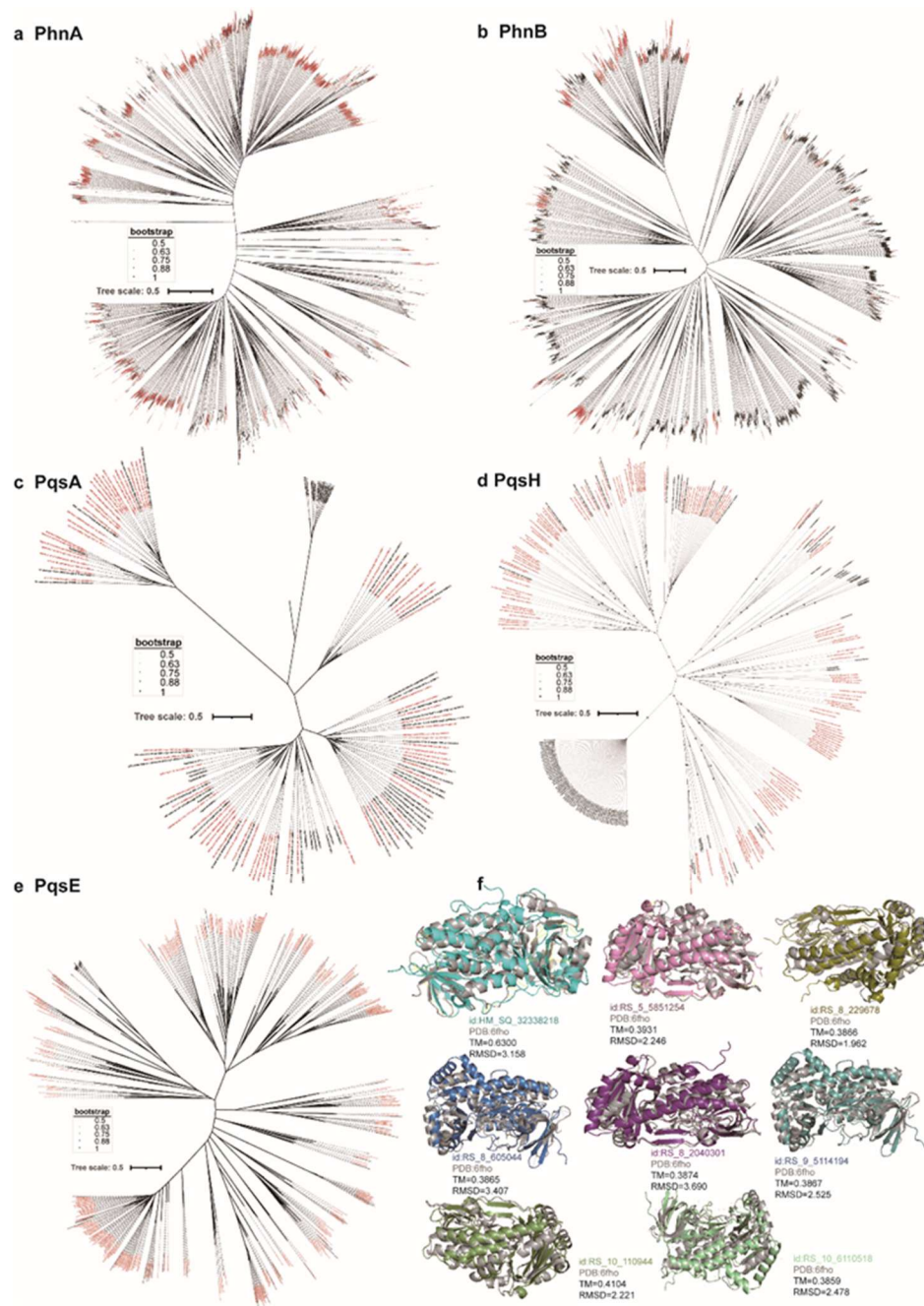
Supplementary Figure 6. The DGC domain conservation model related to c-di-GMP system identified from cold seep microbiomes. Multiple sequence alignment logo of 8,367 DGC domain sequences selected from cold seep microbiomes, highlighting 10 conserved sites of the DGC gene domain, including D327, N335, D344, RxxD, and GGDEF/GGEEF motifs. The multiple sequence alignment was performed using MAFFT and visualized with WebLogo3.



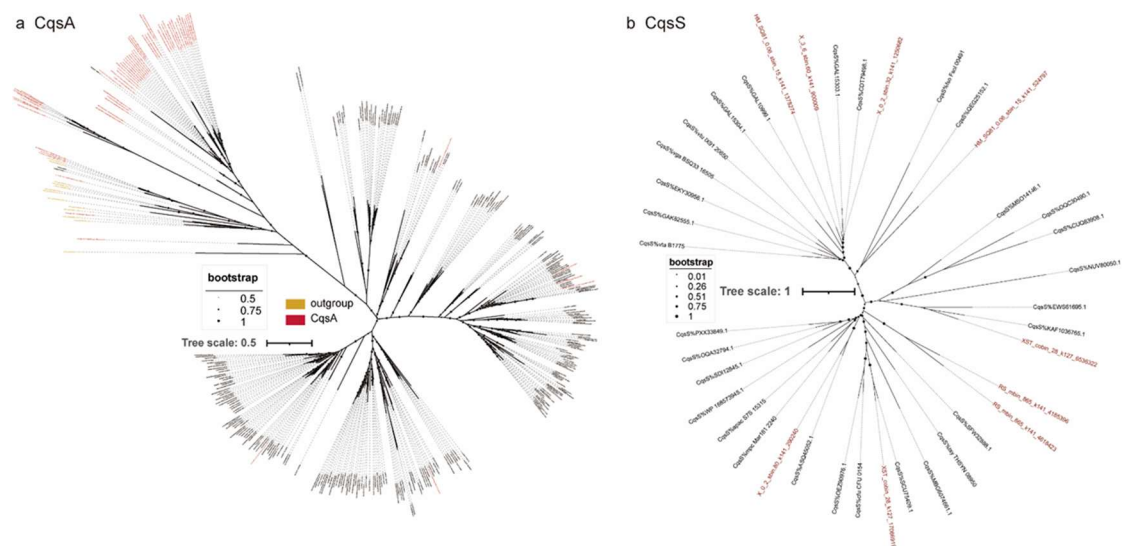
Supplementary Figure 7. Maximum-likelihood phylogenetic trees for PDE and Clp related to c-di-GMP system identified from the non-redundant gene catalog. Phylogenetic trees are shown for PDE (**a**) and Clp (**b**). Colored labels represent target genes, while black labels indicate reference genes. The scale bar shows the average number of substitutions per site.



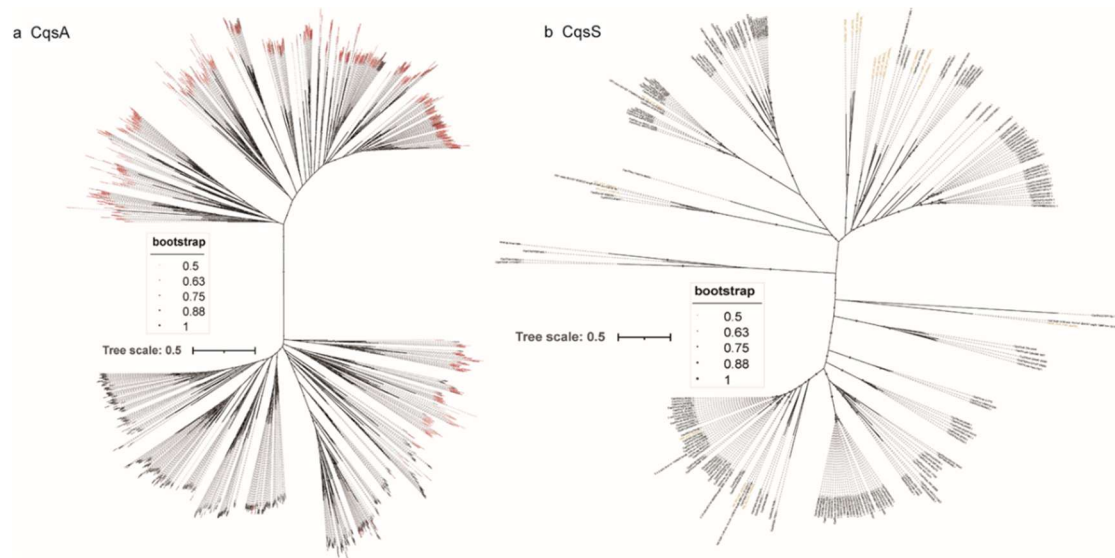
Supplementary Figure 8. Quorum sensing genes related to PQS system identified from cold seep microbiomes. (a-d) Maximum-likelihood phylogenetic trees for PhnA, PhnB, PqsE and PqsH. Colored labels represent target genes from cold seep microbiomes, while black labels indicate reference genes. The scale bar represents the mean number of substitutions per site. (e) The PqsA domain conservation model, showing key conserved sites. (f) Structural alignments of PqsL with their corresponding reference proteins. The alignment results include TM-score and RMSD values, which quantify the degree of structural similarity. Colored structures represent target protein structures, while gray structures represent reference protein structures.



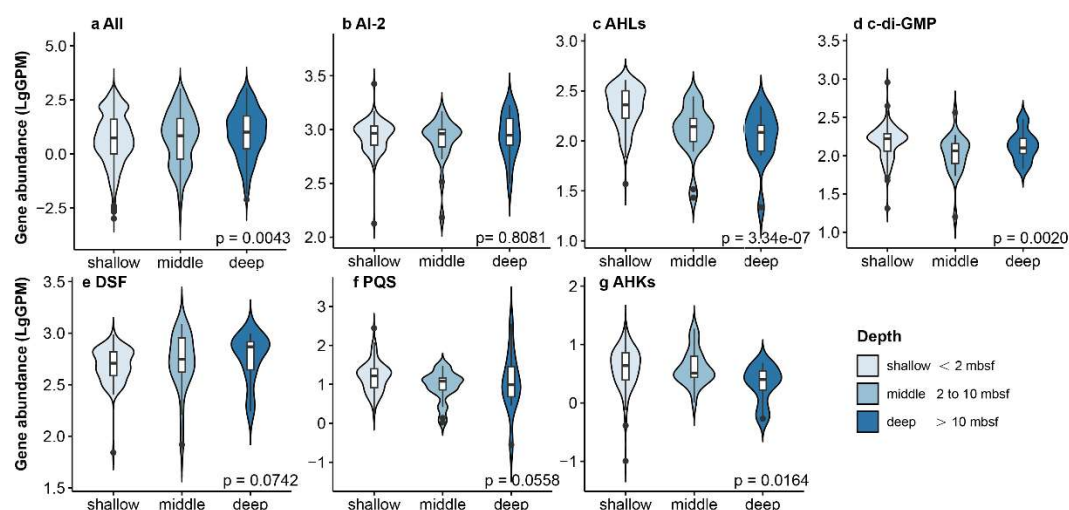
Supplementary Figure 9. Quorum sensing genes related to PQS system identified from the non-redundant gene catalog. (a-e) Maximum-likelihood phylogenetic trees for PhnA, PhnB, PqsA, PqsH and PqsE. Colored labels represent target genes, while black labels indicate reference genes. (f) Structural alignments of PqsL with their corresponding reference proteins. TM-score and RMSD values were used to quantify the degree of structural similarity.



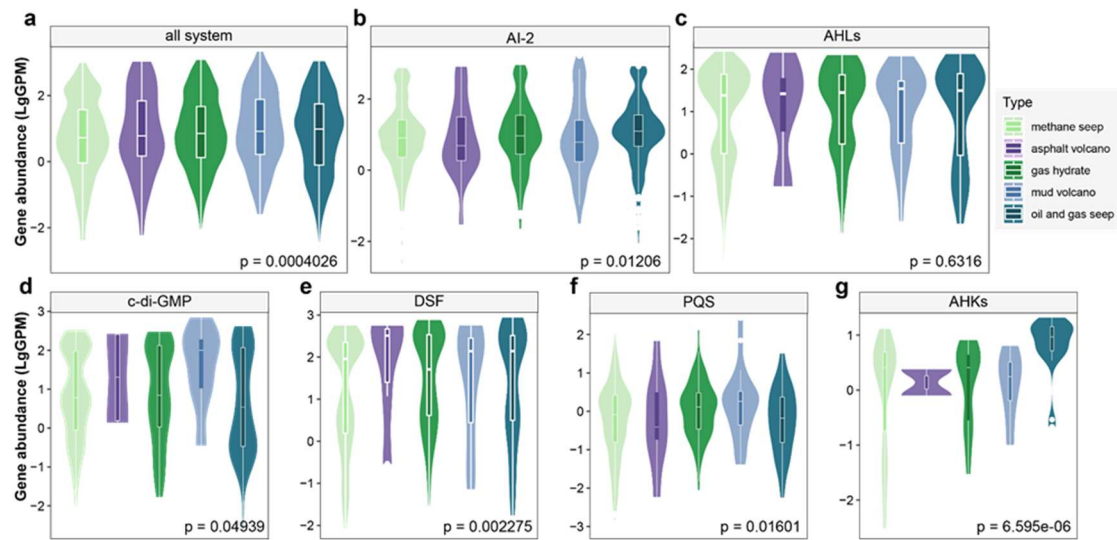
Supplementary Figure 10. Maximum-likelihood phylogenetic trees for quorum sensing genes related to the AHKs system identified from cold seep microbiomes. Phylogenetic trees are shown for CqsA (a) and CqsS (b). Colored labels represent target genes, while black labels indicate reference genes. The scale bar indicates the mean number of substitutions per site.



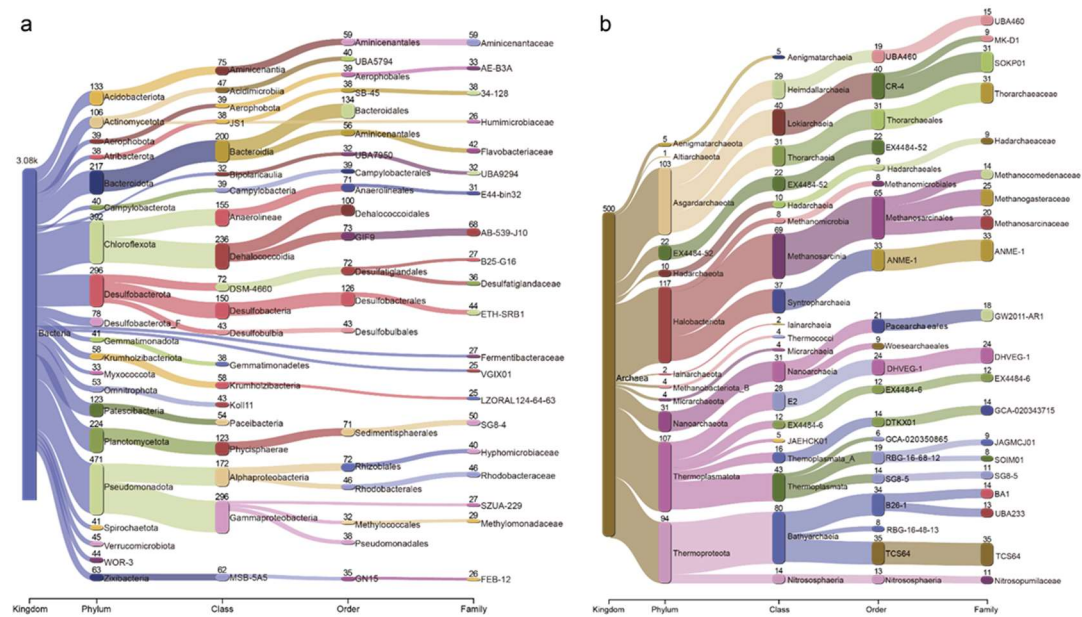
Supplementary Figure 11. Maximum-likelihood phylogenetic trees for quorum sensing genes related to AHKs system identified from the non-redundant gene catalog. Phylogenetic trees are shown for CqsA (**a**) and CqsS (**b**). Colored labels represent target genes, while black labels indicate reference genes. The scale bar indicates the mean number of substitutions per site.



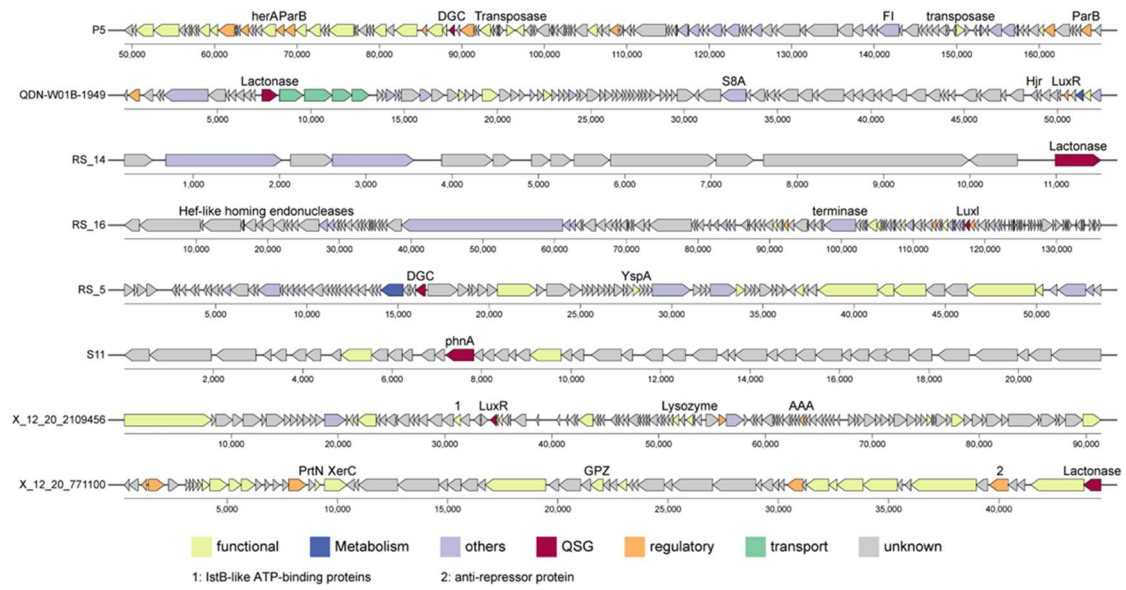
Supplementary Figure 12. Distribution of the relative abundance of six types of quorum sensing systems across cold seep samples collected from various sediment depth layers. Sediment depths are categorized into three groups: surface (< 2 meters below seafloor, mbsf), shallow (2-10 mbsf), and deep (> 10 mbsf). The vertical axis represents the lg-transformed values of gene relative abundance (lg-transformed genes per million, lgGPM) at different sediment layer depths. Kruskal-Wallis rank-sum tests were used to determine differences in the abundance of (a) all quorum sensing systems and (b-g) individual types of quorum sensing systems at different sediment layer depths, with *P* values calculated to determine the statistical significance of the observed variations. Violin plots show: the center line representing the median, box limits indicating the 25th and 75th percentiles, and whiskers extending 1.5 times the interquartile range from the 25th and 75th percentiles. Detailed statistical data are available in **Supplementary Table 9**.



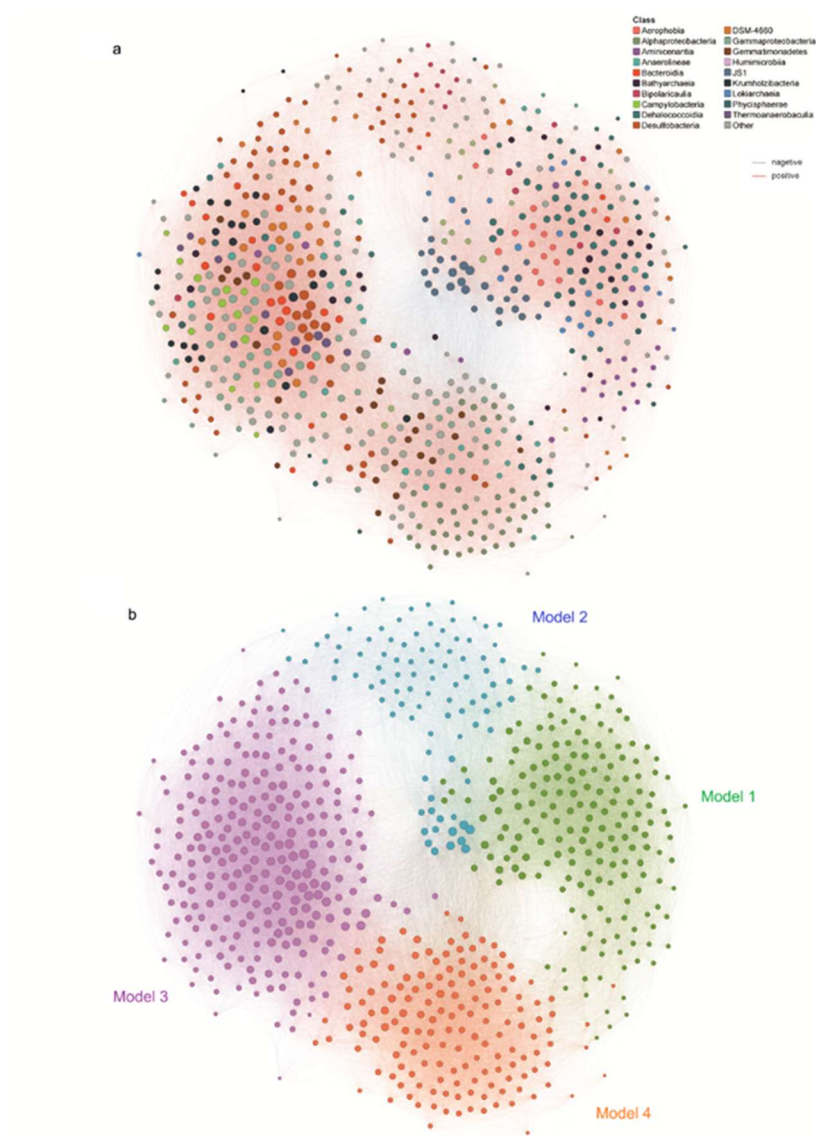
Supplementary Figure 13. Distribution of relative abundance of six quorum sensing systems across 165 cold seep sediment samples collected from five cold seep types. Kruskal-Wallis rank-sum tests were used to determine differences in the abundance of (a) all quorum sensing systems and (b-g) individual types of quorum sensing systems across different cold seep types. *P* values indicate the significance of these differences. Violin plots show: the center line representing the median, box limits indicating the 25th and 75th percentiles, and whiskers extending 1.5 times the interquartile range from the 25th and 75th percentiles. The vertical axis represents the log-transformed values of gene relative abundance (lg-transformed genes per million, lgGPM).



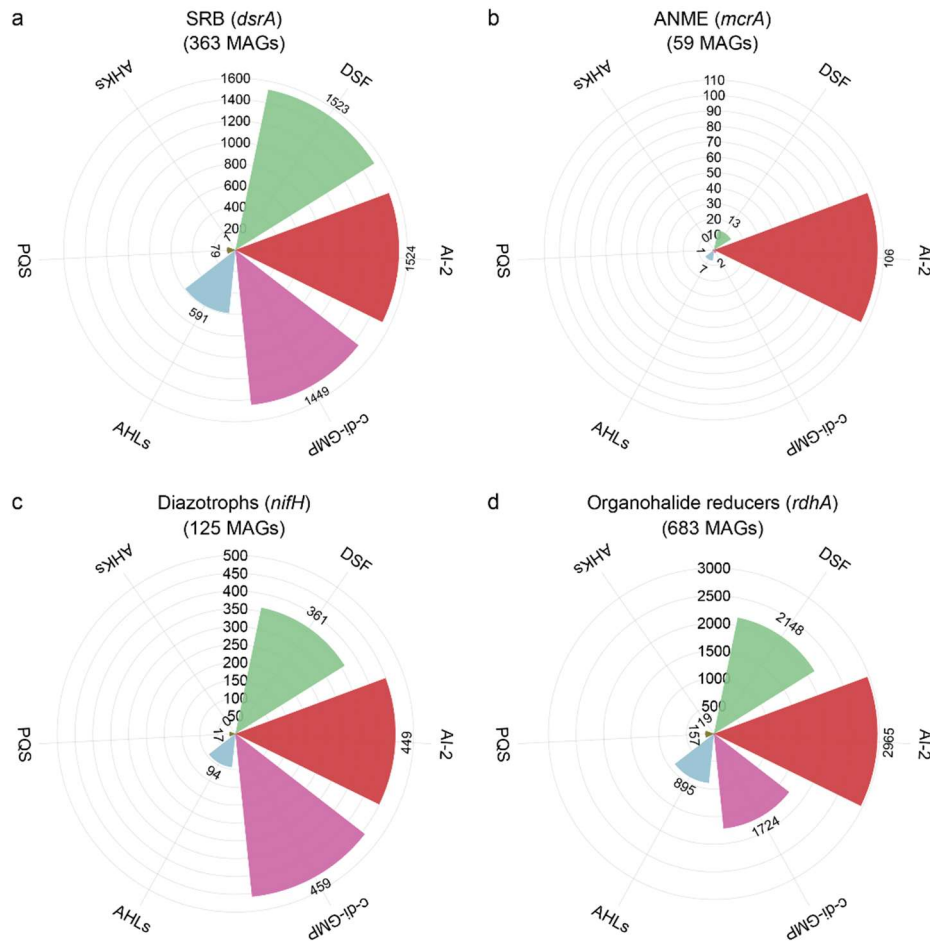
Supplementary Figure 14. Sankey plots showing cold seep microorganisms with quorum sensing genes across different taxonomic levels. (a) Archaeal MAGs and (b) bacterial MAGs. Numbers indicate the quantity of MAGs recovered for each lineage. Detailed classification information for the MAGs is provided in **Supplementary Table 10.**



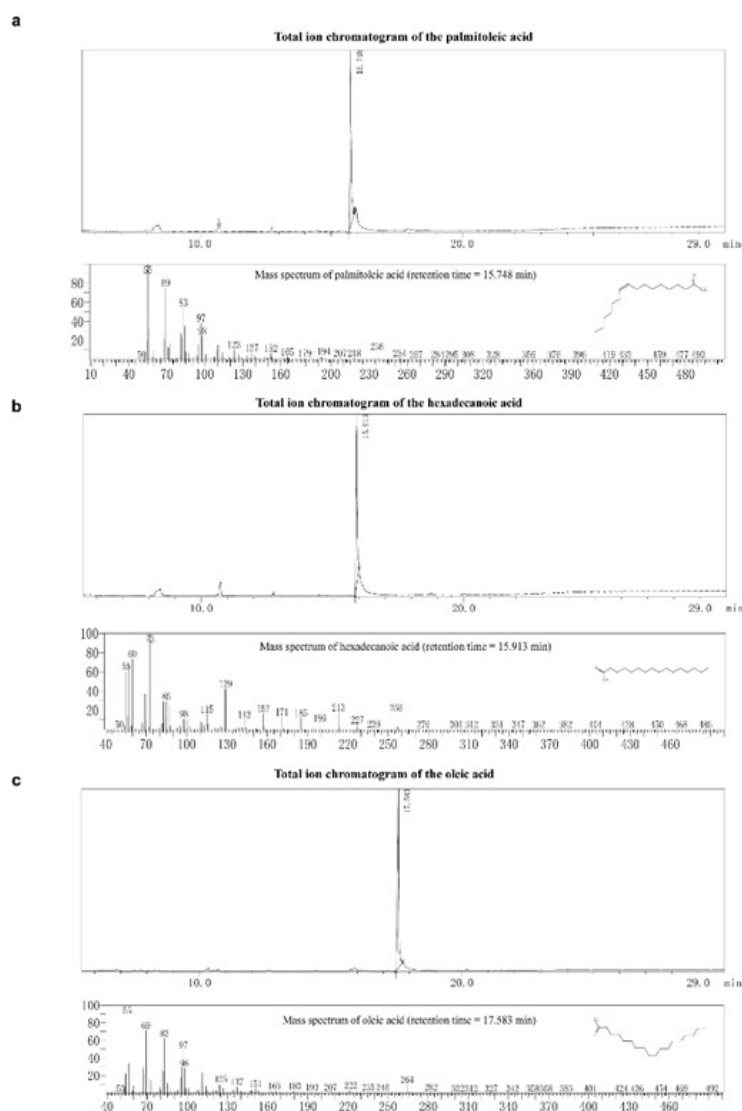
Supplementary Figure 15. Genome synteny of eight representative viral contigs containing quorum sensing genes and their upstream and downstream genes. Different colors represent various gene types: functional genes (yellow-green), metabolism-related genes (blue), other annotated genes (purple), quorum sensing genes (red), regulatory genes (yellow), transport-related genes (green), and unknown genes (gray).



Supplementary Figure 16. Co-occurrence networks for microorganisms significantly correlated with quorum sensing genes. Node size reflects the degree of connectivity. Nodes are colored by class in (a) and by model in (b). In panel (a), red edges indicate significant positive correlations, while blue edges represent significant negative correlations.



Supplementary Figure 17. Distribution of quorum sensing systems across functionally key microbial groups, including sulfate-reducing bacteria (SRB; a), methane-oxidizing archaea (ANME; b), diazotrophs (c), and organohalide reducers (d). Different colors indicate distinct quorum sensing systems, and numbers represent the identified quorum sensing genes in MAGs of each group.



Supplementary Figure 18. GC-MS analysis of reference standards for palmitoleic acid, hexadecanoic acid, and oleic acid. (a) Top: total ion chromatogram (TIC) of palmitoleic acid, with a retention time of 15.748 min. Bottom: mass spectrum of palmitoleic acid at the chromatographic signal corresponding to a retention time of 15.748 min. (b) Top: TIC of hexadecanoic acid, with a retention time of 15.913 min. Bottom: mass spectrum of hexadecanoic acid at the chromatographic signal corresponding to a retention time of 15.913 min. (c) Top: TIC of oleic acid, with a retention time of 17.583 min. Bottom: mass spectrum of oleic acid at the chromatographic signal corresponding to a retention time of 17.583 min.

Table S1. Sampling and sequencing information for metagenomic data used in this study, including geographic location of cold seeps, cold seep types, sample depths, water depths, sequencing platform, assembly and binning statistics, NCBI accession numbers, and source references.

Table S2. Sampling and sequencing information for metatranscriptomic data used in this study, including geographic location of cold seeps, cold seep types, sample depths, water depths, sequencing platform, NCBI accession numbers, and source references.

Table S3. Summary of quorum sensing protein screening methods.

Table S4. Detailed information of 299,355 quorum sensing proteins in the non-redundant gene catalog.

Table S5. Average abundance of quorum sensing systems.

Table S6. Average abundance of quorum sensing proteins.

Table S7. Pairwise significance tests for relative abundance among six quorum sensing systems using the Wilcoxon test.

Table S8. Pairwise significance tests for relative abundance among quorum sensing proteins using the Wilcoxon test.

Table S9. Pairwise significance tests for the relative abundance of quorum sensing proteins across three depth layers using the Wilcoxon test.

Table S10. Detailed information on 32,500 quorum sensing proteins in 3,576 cold seep metagenome-assembled genomes (MAGs).

Table S11. Comprehensive details of 84 quorum sensing proteins identified in viral genomes using HMMER from the QSP database.

Table S12. Pairwise correlation analysis of quorum sensing proteins based on Spearman's rank correlation analysis.

Table S13. Network characteristics of quorum sensing systems.

Table S14. Quorum sensing proteins and metagenome-assembled genomes (MAGs) with significant correlations.

Table S15. Nodes of metagenome-assembled genomes (MAGs) with higher degrees (8-14) and prediction of hub nodes in microbial co-occurrence networks.

Table S16. Statistical data for linear regression between the abundance of *dsrA*, *mcrA*, *nifH*, and *rdhA* and quorum sensing genes.

Table S17. Distribution of quorum sensing proteins in functionally key microbial groups and taxonomic classification of metagenome-assembled genomes (MAGs).

Table S18. Diversity of dCACHE_1-containing AI-2 receptors predicted using the Pfam database.

Table S19. List of identified *lux* box sequences and their downstream genes predicted in functionally key microbial groups using FIMO-MEME.