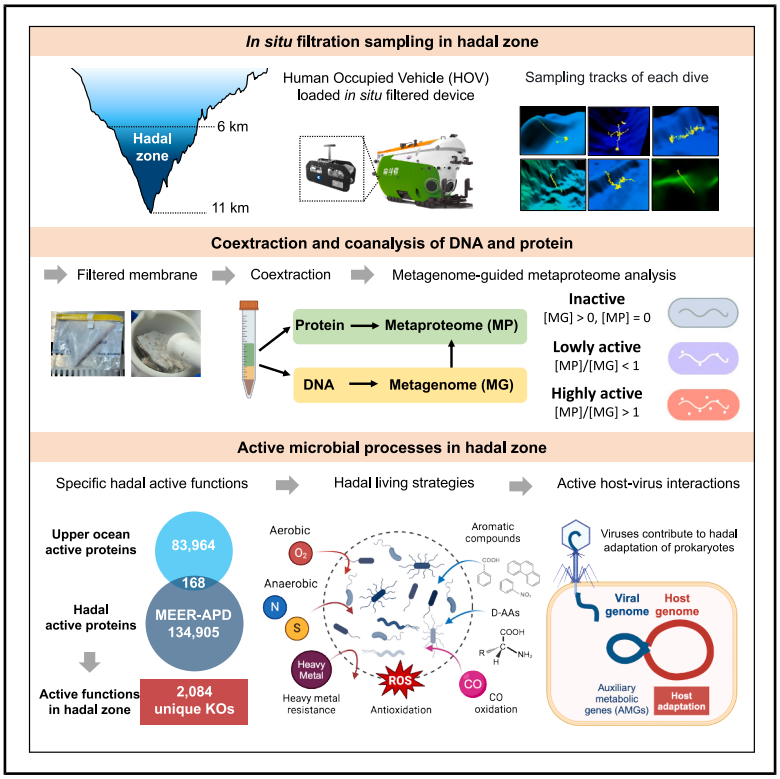
Article

Cell Host & Microbe

Unveiling active microbial processes in Earth's deepest seawater

Graphical abstract



Highlights

- A DNA-protein co-extraction and co-analysis workflow for low-biomass hadal seawater
- Constructed MEER-APD with 135,073 non-redundant hadal proteins, >95% hadal-specific
- Active hadal microbiome expands carbon/energy utilization, reshaping element cycles
- Active hadal viruses show dual strategies beyond Piggyback-the-Winner dynamics

Authors

Wei-Jia Zhang, Aoran Hu, Ziheng Wu, ..., Liang Meng, Xiang Xiao, Weishu Zhao

Correspondence

mengliang1@genomics.cn (L.M.), zjxiao2018@sjtu.edu.cn (X.X.), zwsh88@sjtu.edu.cn (W.Z.)

In brief

In situ filtration of hadal seawaters during 12 dives (6–11 km depth) enabled Zhang et al. to perform metagenome-guided metaproteomic analysis of hadal microbiomes. The deepest oceanic ecosystems show distinct active microbial processes compared with the upper ocean, suggesting the hadal zone contributes to biogeochemical cycles more than previously appreciated.

Zhang et al., 2026, Cell Host & Microbe *34*, 1540–1559
August 12, 2026
https://doi.org/10.1016/j.chom.2026.07.001

CellPress

CellPress
OPEN ACCESS

Cell Host & Microbe

Article

Unveiling active microbial processes in Earth's deepest seawater

Wei-Jia Zhang,[1,8,11] Aoran Hu,[2,3,11] Ziheng Wu,[2,3,11] Liangting Liu,[2,11] Chen Li,[4] Yujue Wang,[4] Zhanfei Wei,[6,8,9] Rui Lu,[6,8,9] Jun Li,[1] Yang He,[3] Taoliang Zhang,[2] Shanshan Liu,[7,8] Jing Wang,[5] Liang Meng,[6,8,9,*] Xiang Xiao,[2,3,5,10,12,*] and Weishu Zhao[2,3,5,*]

[1]Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya 572000, China
[2]International Center for Deep Life Investigation (IC-DLI), Hainan Research Institute, Shanghai Jiao Tong University, Sanya 572025, China
[3]Institution of Deep Life and Resource Investigation (IDLRI), State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China
[4]Center for High Performance Computing, Shanghai Jiao Tong University, Shanghai 200240, China
[5]Shanghai Changxing Ocean Laboratory, Shanghai Jiao Tong University, Shanghai 201913, China
[6]BGI Research, Sanya 572025, China
[7]BGI, Shenzhen 518083, China
[8]Institution of Deep-sea Life Sciences, IDSSE-BGI, Hainan Deep-sea Technology Laboratory, Sanya 572000, China
[9]Hainan Technology Innovation Center for Marine Biological Resources Utilization (Preparatory Period), BGI Research, Sanya 572025, China
[10]Shanghai Academy of Natural Sciences (SANS), Shanghai 200000, China
[11]These authors contributed equally
[12]Lead contact
*Correspondence: mengliang1@genomics.cn (L.M.), zjxiao2018@sjtu.edu.cn (X.X.), zwsh88@sjtu.edu.cn (W.Z.)
https://doi.org/10.1016/j.chom.2026.07.001

SUMMARY

Microorganisms dominate life in the hadal zone, yet extreme sampling difficulty and low biomass have precluded characterization of their *in situ* activities. Here, we analyze microbiome samples collected from hadal seawaters via *in situ* filtration during 12 human-occupied vehicle dives. DNA-protein co-extraction and metagenome-guided metaproteomic analysis identify 135,073 non-redundant active proteins, with over 95% being hadal-specific. Metaproteomic quantification distinguishes highly active and less active taxa that differ in biogeographic origins and genomic traits. Hadal microorganisms operate a metabolic regime fundamentally distinct from the upper ocean, preferentially utilizing refractory organic matter (aromatics, halogenated compounds, and D-amino acids) and expanded electron acceptors (thiosulfate and heavy metals), collectively shaping hadal element cycling. Active viruses extend beyond "Piggyback-the-Winner" dynamics, enhancing host adaptation through auxiliary metabolic genes. These findings provide proteome-level evidence of hadal microbial activities and reveal biogeochemical cycling distinct from that of the upper ocean, highlighting the underappreciated significance of hadal microbiomes within global ocean ecosystems.

INTRODUCTION

The hadal zone, encompassing oceanic regions from 6 km below sea level (b.s.l.) to the ultimate water depth of 11 km at the Challenger Deep in the Mariana Trench (MT), represents one of Earth's most extreme and least explored environments.[1–5] These hadal environments, characterized by 60–115 MPa ultra-high hydrostatic pressure, darkness, near-freezing temperatures, and nutrient limitations, pose extreme challenges to most life forms on Earth.[5] Microorganisms predominate in such extreme environments,[1] and both cultivation-dependent and -independent studies have shown that hadal microbial communities differ markedly from those in the overlying water column.[6,7] Comparative genomic analyses revealed a higher abundance of metal ion transporters in isolates from the MT, suggesting that hadal microorganisms are specifically adapted to the stress of high concentrations of heavy metals within the trench.[8] In addition, 16S rRNA gene-based surveys have identified microbial taxa with the potential to degrade refractory organic matter (ROM) in hadal seawaters.[9] More recently, systematic investigations of metagenomes from hadal sediment samples have observed distinctive metabolic potentials in hadal prokaryotes, distinguishing them from those in non-hadal ocean regions.[1,10,11] The hadal sediment microbiomes were characterized by capacities for aromatic compound degradation and antioxidant defense against reactive oxygen species (ROS), possibly reflecting selective adaptations to the hadal extreme conditions.[1] However, whether these metagenome-predicted metabolic potentials are actively expressed *in situ* remains unverified.

Transcription-based analyses have provided partial insights into active hadal prokaryotes. Comparison of 16S rDNA and rRNA libraries from the MT revealed significant

community-level disparities: rDNA libraries were dominated by Thaumarchaeota (ammonia oxidating archaea, now classified as class-level lineage Nitrososphaeria in metagenomic taxonomy), Marinimicrobia, and Bacteroidetes, while rRNA libraries were dominated by Gemmatimonadetes and Chloroflexi.[12] Another study found that although Thaumarchaeota dominated both 16S rDNA and rRNA libraries in the benthic boundary layer, their RNA/DNA ratio was much lower than in shallower waters.[13] To date, only two metatranscriptomic analyses of MT samples have been reported,[14,15] revealing CO_2 fixation, nitrification, sulfur oxidation, and arsenate reduction were transcripted. However, their limited sediment-core coverage leaves hadal prokaryotic activities poorly resolved, creating a significant knowledge gap for understanding global ocean ecosystems.

Active hadal viruses are even less understood, despite reaching up to 10^5 virus-like particles per mL in hadal seawater.[16] Metagenomic studies showed distinct hadal viral communities dominated by Caudoviricetes, including Myoviridae, Siphoviridae, and Podoviridae.[17–19] Viral production estimates and metagenomes suggested enhanced lysogenic viral production and enrichment of integrase genes in hadal seawater,[18] whereas only ~5% of MT sediment viruses were predicted to be lysogenic,[19] indicating poorly resolved viral life strategies. Furthermore, while viruses are known to reprogram host metabolism and facilitate host adaptation through auxiliary metabolic genes (AMGs),[20,21] direct evidence for their *in situ* expression in hadal environments remains lacking.

Metaproteomics offers a more robust methodology for verifying microbiome activities,[22–25] including both prokaryotes and viruses. However, its application in hadal samples has been severely hindered by two key challenges. First, metaproteomic analysis demands substantial biomass input,[26,27] which is difficult to obtain from hadal environments, especially for hadal seawaters. Conventional seawater sampling using Niskin bottles (with capacities of tens of liters) could collect sufficient microbial biomass for metaproteomic analysis from surface seawaters,[26,28,29] but such volumes are highly inadequate for hadal seawaters, where cell densities (10^3–10^4 cells per mL[16]) are one to two orders of magnitude lower.[30,31] Second, accurate spectral identification depends on comprehensive, well-curated gene databases, which remain scarce for poorly characterized hadal ecosystems.[27,32,33]

To address these challenges, we employed an *in situ* filtration apparatus[34] aboard a human-occupied vehicle (HOV) for microbial collection in the hadal zone. This system is capable of filtering high-volume (>100 L per sample) near-benthic seawater during each HOV dive, thus providing sufficient microbial biomass for subsequent analyses. Based on these samples, we developed a framework integrating DNA-protein co-extraction and metagenome-guided metaproteomic analysis for efficient protein identification. Despite the low microbial biomass of hadal seawater, our approach identified more expressed microbial proteins than previous surface seawater studies,[26,28,29,35,36] demonstrating the effectiveness of this methodology in obtaining protein expression data from low-biomass water samples. By using this framework, a total of 135,073 non-redundant expressed proteins were identified from 12 *in situ* filtered hadal seawater samples, establishing the Active Protein Dataset of the Mariana Trench Environment and Ecology Research Project (MEER-APD). This dataset provides systematic proteome-level evidence of active microbial processes in Earth's deepest oceanic regions. We further investigated the functional profiles associated with varying microbial expression levels, the biogeographic pattern of the hadal taxa, and active viral dynamics and virus-host interactions. These findings indicate that microbial-mediated biogeochemical cycles in the hadal zone differ fundamentally from those in the upper ocean, highlighting their previously underappreciated contribution to global ocean processes.

RESULTS

Developing a workflow for analyzing low-biomass hadal seawater

To overcome the challenges in studying low-biomass hadal seawater samples, we implemented newly developed equipment, the *in situ* microbial filtration and fixation apparatus (ISMIFF)[37] on a HOV to collect microbial samples from near-benthic seawater (1–3 m above the seafloor). Each ~4-h filtration covered 3–5 km HOV traverses, avoiding site-specific heterogeneity. We obtained 12 *in situ* filtration samples at water depths of ~6–11 km b.s.l., spanning the MT (n = 7), the Yap Trench (YT; n = 1), and the west Philippine Basin (PB; n = 4) (Figures 1A and 1B; Table S1; STAR Methods).

Each filtration sample concentrated microbiomes from over 100 L of seawater, providing sufficient microbial biomass for the development of the integrated DNA-protein co-extraction approach (Figure 1C). DNA-protein co-extraction yielded approximately 0.05–0.2 μg protein per L of hadal seawater, one to two orders of magnitude lower than that from surface seawater (Figures 2A and 2B; Table S1),[38] consistent with reduced active microbial cell abundance across pelagic depth zones.[30,31]

To optimize protein identification and minimize false discoveries associated with oversized databases,[39] we established a metagenome-guided metaproteomic analysis pipeline (Figure 1C; STAR Methods). Deep metagenomic sequencing (100 Gbp per sample) yielded 17,070,127 contigs and 44,121,717 predicted protein-coding sequences, including both full-length and partial genes (Table S1). Metagenomic binning recovered 552 species-level representative metagenome-assembled genomes (MAGs), accounting for 75.3% of mapped reads and containing 1,538,828 predicted genes (Table S2; STAR Methods).

Metaproteomic analyses generated 310,825–420,289 mass spectra per sample (Table S1). Searching against metagenome-derived datasets, we identified 629,338 peptides/149,370 proteins using the contig-based dataset and 453,341 peptides/38,055 proteins using the MAG-based dataset, collectively yielding 135,073 non-redundant active proteins, 14-fold more than UniProt-based searches[40] (Table S1). Despite the lower biomass in hadal environments, our workflow identified more proteins from hadal seawater than previously reported from biomass-rich surface seawaters (Figures 2B and 2C),[31,41] demonstrating its efficacy for low-biomass extreme environments.

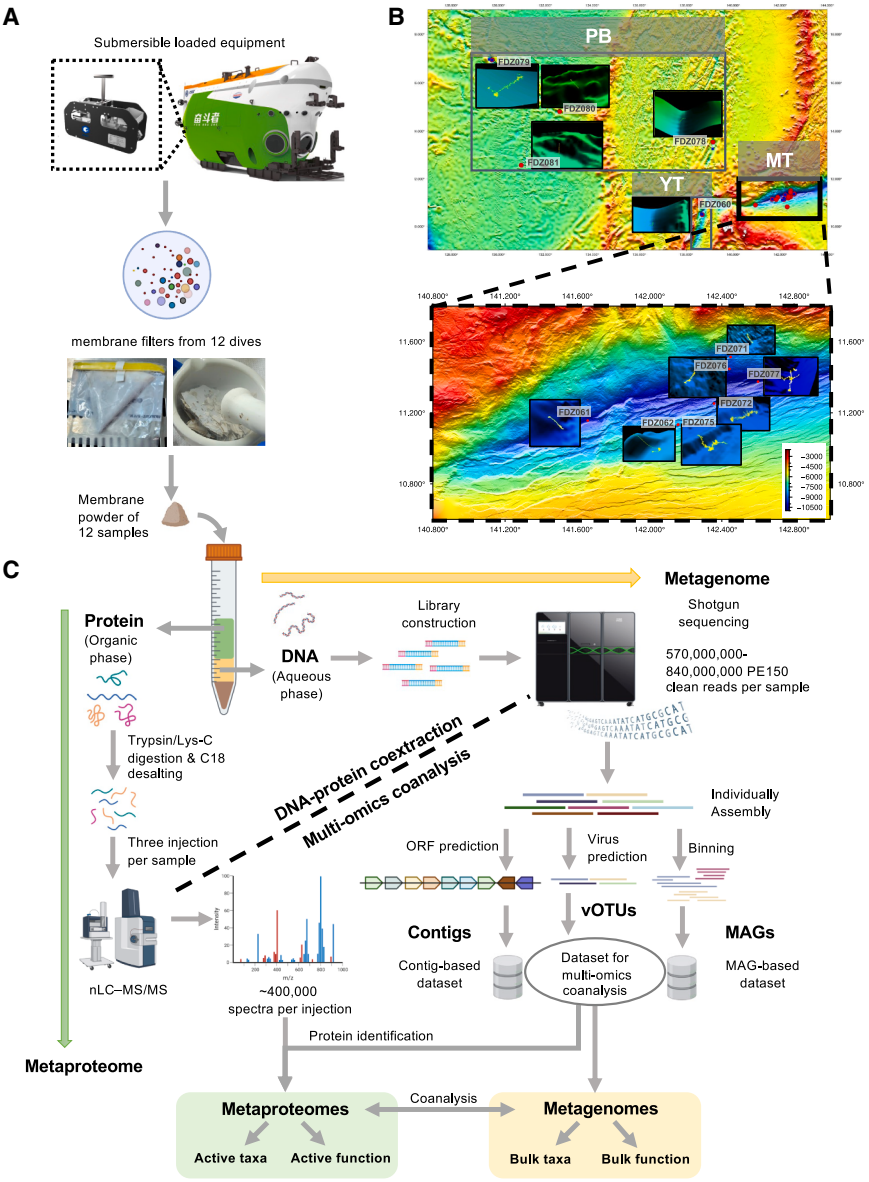


Figure 1. Developing a workflow for analyzing low-biomass hadal seawater

(A) *In situ* microbial filtration and fixation apparatus (ISMIFF) loaded on HOV and sample of filtrated membrane.

(B) Sampling sites in the hadal regions. PB, the Philippine Basin; YT, the Yap Trench; MT, the Mariana Trench. The map below shows the enlarged view of the sampling area within the Mariana Trench. Red dots indicate sampling sites. Yellow lines indicate submersible dive trajectories.

(C) Schematic framework of the co-extraction and co-analysis of the metagenomes and metaproteomes. Detailed procedures are in the STAR Methods section.

See also Table S1.

High specificity of active microbial functions in hadal seawater

We constructed the MEER-APD using all 135,073 non-redundant proteins identified from hadal metaproteomes. For comparison, we also collected publicly available metaproteomes (89,762 proteins, 84,132 non-redundant proteins) from five previous studies of upper-ocean seawaters (0–3,215 m water depth)[26,28,29,35,36] (Table S1; STAR Methods). Using the commonly accepted UniRef100 criterion (100% identity) for protein differentiation,[42] clustering analysis revealed that 99.88% (134,905 proteins) of MEER-APD were exclusively expressed in the hadal seawater, with only ~0.1% (168 proteins) shared between the hadal zone and the upper ocean (Figure 2D). This remarkable specificity was maintained when applying the UniRef90 criterion (90% identity) (STAR Methods), with 95.47% of MEER-APD proteins (128,957) remaining uniquely expressed in the hadal environment.

Functional annotation of MEER-APD proteins yielded 4,656 Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthologs (KOs), of which approximately 50% (2,084 KOs) were exclusively detected in hadal metaproteomes and absent from the upper ocean (Figure 2D; Table S3). Enrichment analysis of KO functional categories revealed distinct functional preferences between microorganisms inhabiting hadal seawater and the upper ocean (Figure 2E). MEER-APD showed significant enrichment in environmental information processing, particularly two-component systems and bacterial motility proteins, while upper-ocean metaproteomes were enriched in protein synthesis and photosynthesis-related pathways (e.g., the Calvin cycle and oxidative phosphorylation). Further enrichment analysis of the 2,084 MEER-APD-specific KOs revealed significant enrichment in KEGG pathways related to the secretion system, tRNA biogenesis, benzoate degradation, biofilm formation, replication and repair, and peptidoglycan biosynthesis and degradation pathways (Figure 2E). These findings demonstrate that active microbial functions in the hadal seawater are markedly distinct from those in the upper ocean.

Expression level quantifies two distinct active statuses of prokaryotic taxa

Protein expression displayed strong taxonomic inequality (Gini index 0.72–0.93).[43] At the metagenomic level, Pseudomonadales (mean relative abundance 13.2%) and Nitrososphaerales (ammonia-oxidizing archaea, AOA; mean relative abundance 12.8%) dominated the community, followed by Burkholderiales (mean relative abundance 6.83%) (Figure S1A). However, metaproteomic profiles diverged markedly: Pseudomonadales and Burkholderiales together accounted for up to 70% of detected proteins per sample, whereas Nitrososphaerales contributed

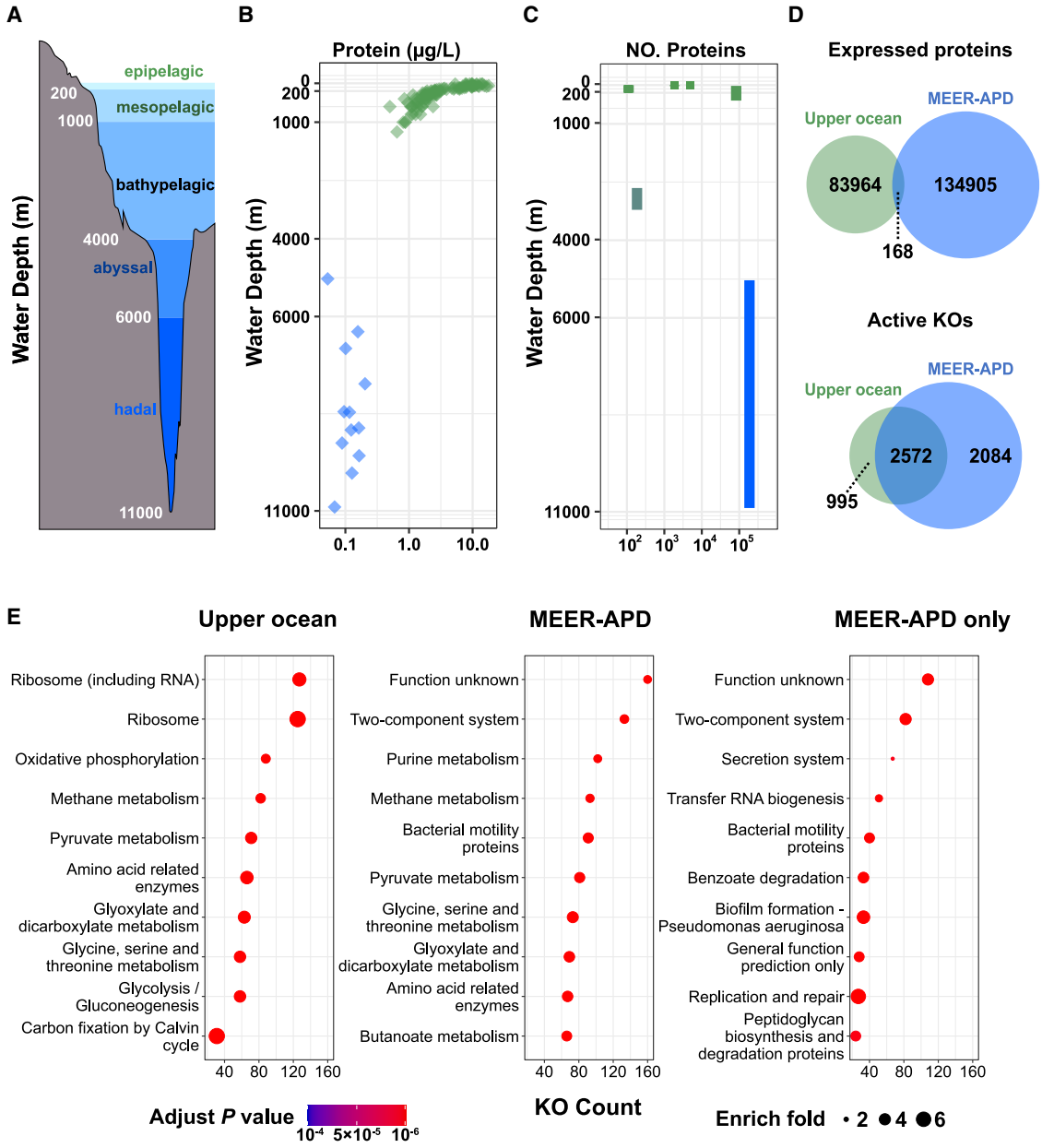


Figure 2. High specificity of active microbial functions in hadal seawater

(A) Schematic of oceanic depth zones.

(B) Depth distribution of peptide concentrations in MEER-APD compared with the upper ocean. Green and blue indicate upper ocean and hadal zone samples, respectively. Upper ocean data were from Saunders et al.[38]

(C) Depth distribution of identified protein counts in MEER-APD compared with the upper ocean. Color scheme as in (B). The upper ocean data were obtained from the studies conducted by Giljan et al.,[26] Brum et al.,[28] Hawley et al.,[29] McCain et al.,[35] and McIlvin and Saito.[36]

(D) Venn diagrams showing proteins (top) and KOs (bottom) specific to or shared between the upper ocean and MEER-APD. Numbers indicate counts in each category.

(E) Enriched KEGG metabolic pathways in the upper ocean (left) and MEER-APD (middle), and pathways uniquely enriched in MEER-APD (right). Only the top ten enriched terms are shown.

See also Tables S1 and S3.

only 2.69% on average, revealing a 4-fold disparity between genomic presence and protein-level activity (Figure S1B).

To quantify this disparity, we noted that metagenomic ([MG]) and metaproteomic ([MP]) relative abundances exhibited a robust linear correlation, described by

$$[MP] = a \cdot [MG], \qquad \text{(Equation 1)}$$

where a is the slope, reflecting the expression ratio of a taxon in hadal seawater. Saturation analyses confirmed adequate capture of the active prokaryotic community (Figure S1E), and

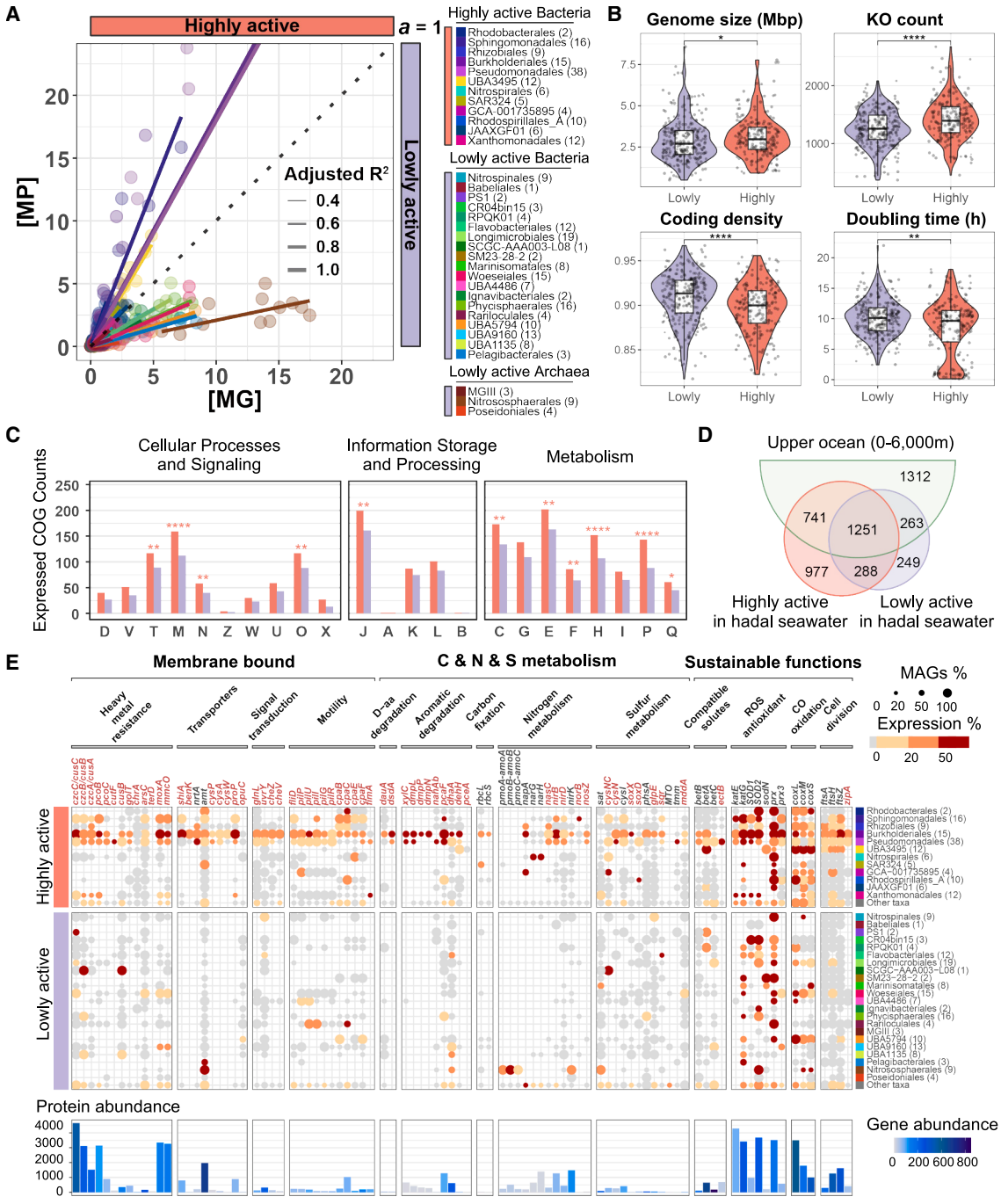


Figure 3. Expression level quantifies two distinct active statuses of prokaryotic taxa

(A) Linear relationships between the relative abundances in the [MG] and [MP] at the order level. Points represent the summed abundance of all MAGs within each order per sample, color-coded by order. Lines show linear regressions, with a dashed baseline ([MP] = [MG]). Line thickness reflects R^2. Bracket numbers indicate total MAG counts.

(B) Genomic feature comparisons between lowly active and highly active genomes. Violin plots show value distributions, with embedded boxplots showing median and interquartile range. Individual genomes are shown as points. Statistical significance was assessed using two-sided Wilcoxon rank-sum tests (highly active, $n = 199$; lowly active, $n = 287$). Significance: $^{****}p \leq 0.0001$, $^{***}p \leq 0.001$, $^{**}p \leq 0.01$, $^{*}p \leq 0.05$.

(C) Expressed COG counts for lowly active (purple) and highly active (orange) genomes across COG functional categories. Codes for COG categories: D, cell-cycle control, cell division, chromosome partitioning; V, defense mechanisms; T, signal transduction mechanisms; M, cell wall/membrane/envelope biogenesis; N, cell motility; Z, cytoskeleton; W, extracellular structures; U, intracellular trafficking, secretion, and vesicular transport; O, posttranslational modification, protein turnover, chaperones; X, mobilome: prophages, transposons; J, translation, ribosomal structure, and biogenesis; A, RNA processing and modification; K, transcription; L, replication, recombination, and repair; B, chromatin structure and dynamics; C, energy production and conversion; G, carbohydrate transport

(legend continued on next page)

only orders (133) and MAGs (527) with non-zero proteomic abundance in at least three samples were included (STAR Methods). Over 90% of orders (118/133) exhibited robust linear relationships ($R^2 \geq 0.5$, $p \leq 0.01$) and were classified as either highly active ($a > 1$; contributing more proteins than expected from genomic abundance) or lowly active ($a < 1$; contributing fewer) (Figure 3A; Table S4).

Among the 34 orders exceeding 1% metagenomic abundance, a striking archaeal-bacterial divergence emerged. All three metagenomically abundant archaeal orders—Nitrososphaerales, MGIII, and Poseidoniales—were lowly active (a = 0.12–0.35), suggesting large populations but low per-cell metabolic output. Bacterial taxa displayed broader variability. Among the most metagenomically abundant bacterial orders, Pseudomonadales (a = 1.80) and Burkholderiales (a = 1.85) were highly active, whereas others—such as Marinisomatales (a = 0.45), Pelagibacterales (a = 0.29), and Nitrospinales (a = 0.94)—were lowly active. Furthermore, high activity was observed in taxa with low metagenomic abundance. Notably, Rhodobacterales exhibited the highest expression level overall (a = 2.82), despite its low mean metagenomic abundance (0.54%) (Figure 3A; Table S4).

Genomic and functional traits of the highly and lowly active taxa

Comparison of genomic features revealed distinct signatures between the highly and lowly active groups (Figure 3B; Table S2; STAR Methods). Highly active taxa exhibited larger genomes ($p \leq 0.05$) and encoded significantly more KOs ($p \leq 0.0001$), indicating expanded genomic repertoires and broader metabolic versatility. Conversely, lowly active taxa displayed higher coding density ($p \leq 0.0001$) and longer predicted doubling times ($p \leq 0.01$), features consistent with genomic streamlining (Figure 3B). These patterns partially parallel the "copiotroph-oligotroph" distinction[44,45]; however, no significant differences were observed in codon usage bias, GC content, oxygen tolerance, or temperature preference (Figure S2; Table S2), indicating that the highly/lowly active classification does not strictly correspond to the classical copiotroph-oligotroph dichotomy.[44,45] All the above statistical comparisons used two-sided Wilcoxon rank-sum tests.

Functional analyses based on Clusters of Orthologous Genes (COG) annotations further distinguished these two groups. Highly active species expressed more functional genes across all COG categories (Table S2), with Fisher's exact test revealing significantly elevated diversity in multiple categories spanning cellular processes, information processing, and metabolism (Figure 3C). This broad, non-selective enrichment across all COG categories contrasts with the category-specific enrichment expected under the copiotroph-oligotroph framework,[44,45] reinforcing the interpretation that the a-value reflects realized metabolic activity rather than trophic strategy. Moreover, highly active MAGs retained 2.3-fold more uniquely expressed KOs than lowly active ones (1,254 versus 537), indicating substantially broader functional engagement in the hadal environment (Figure 3D).

KEGG-based comparison also highlighted distinct adaptive strategies between these two groups (Figures 3E and S3; Table S4). Core functional genes related to antioxidation and CO oxidation were universally expressed in both groups, reflecting shared responses to high pressure and nutrient limitation in hadal environments.[46–48] Beyond these shared functions, highly active taxa exhibited broader and more versatile metabolic profiles. For example, Pseudomonadales and Burkholderiales expressed abundant proteins involved in heavy metal resistance, ROM utilization, chemotaxis, motility, and cell division, while Sphingomonadales deployed diverse substrate transport systems facilitating resource acquisition under nutrient limitation. Lowly active taxa, by contrast, showed more specialized functional profiles (Figure 3E). A notable example is the Nitrososphaerales (AOA), which expressed ammonia oxidation genes (*amoA*) but lacked detectable expression of key enzymes in the commonly coupled 3HP–4HB carbon fixation pathway,[49] suggesting a specialized survival strategy centered on a restricted metabolic function in the hadal zone.

Biogeographic origins and divergent traits of hadal taxa

By integrating hadal seawater MAGs with genomes from the Ocean Microbiomics Database (OMD, spanning epipelagic to abyssal zones)[50] and hadal sediment MAGs from the MEER project,[1] active hadal species were classified into four biogeographic patterns: (1) hadal-specific species (n = 486), exclusively found in the hadal zone; (2) widespread species, occurring across both hadal and non-hadal environments (n = 11); (3) "hadal-to-non-hadal" species, originating in the hadal zone and extending to non-hadal environments (n = 4); and (4) "non-hadal-to-hadal" species, derived from non-hadal environments and immigrating into the hadal zone (n = 21) (Figures 4A and S4; Table S2; STAR Methods).

These four biogeographic patterns exhibited distinct expression levels (a-value) and genomic characteristics (Figure 4B; Table S2). Hadal-specific species, representing the dominant fraction, spanned the broadest range (a = 0.06–52.08), reflecting substantial metabolic heterogeneity within endemic lineages. Widespread species showed the highest median activity (a = 1.91), followed by hadal-to-non-hadal species (a = 1.25). Non-hadal-to-hadal species displayed the lowest activity

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 metabolite biosynthesis, transport, and catabolism. COG detection rate differences assessed by two-tailed Fisher's exact test with Benjamini-Hochberg correction. Significance levels as in (B). No significance symbol is shown for functional categories without a significant difference.

(D) Venn diagrams show the specific and shared KOs between the lowly active and highly active genomes in MEER-APD and the upper ocean. Numbers indicate KO counts in each category.

(E) Expression profiles of active key genes across orders, expression levels, and samples. Genes expressed exclusively in hadal samples are marked in red. Upper: metagenome and metaproteome abundances across orders, where circle size reflects the proportion of MAGs encoding the gene and color indicates the proportion expressing it. Color scheme is consistent with (A). Lower: total hadal metaproteomic abundance (bar height) and metagenomic abundance (bar color). Bracket numbers denote total MAG counts per order.

See also Figures S1–S3 and Tables S2, S3, S4, and S5.

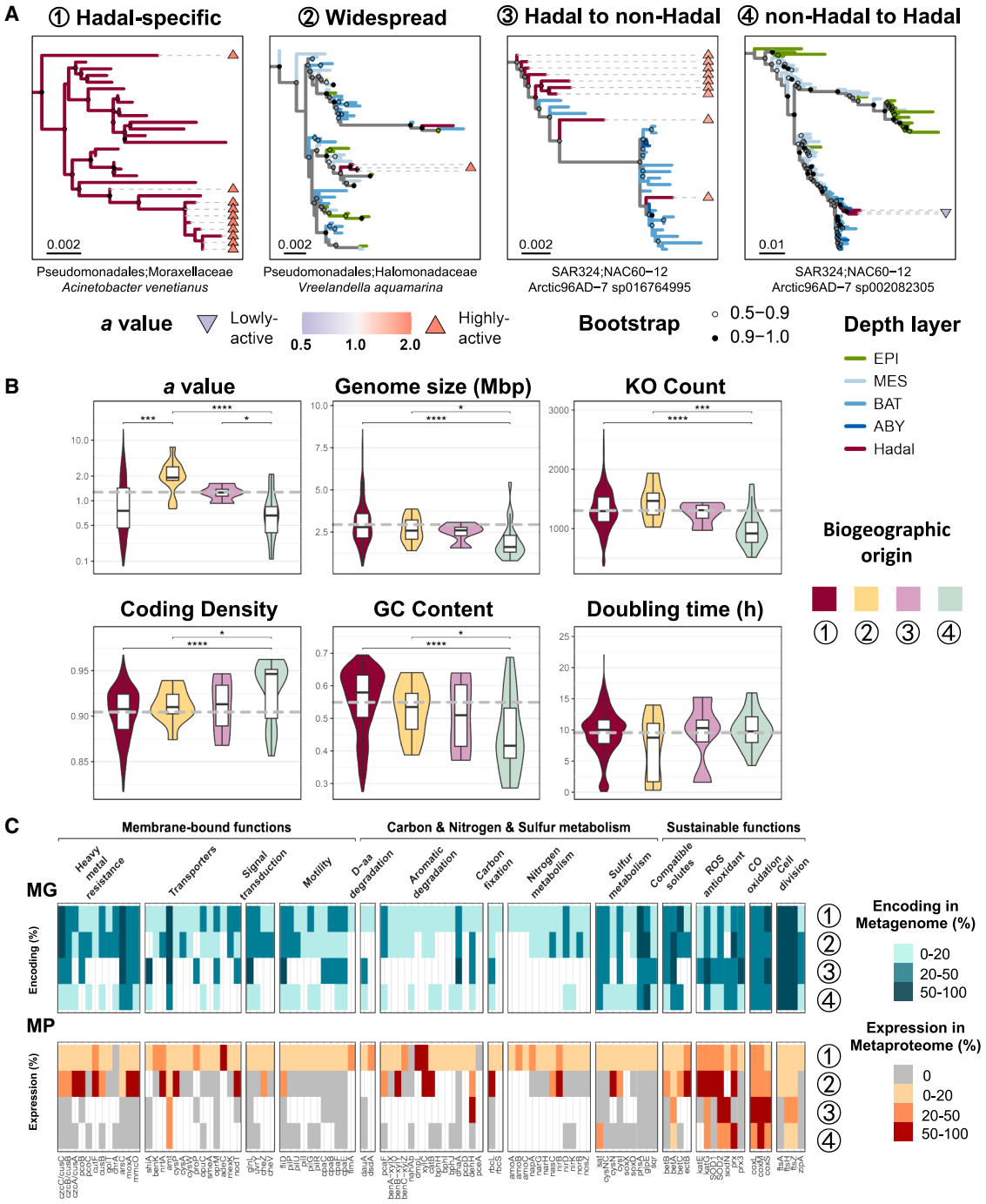


Figure 4. Biogeographic origins and divergent traits of hadal taxa

(A) Representative species for each divergence pattern. Branches are colored by depth layer when all descendant genomes share the same depth, or gray when mixed. Hadal seawater MAGs from this study are marked by dashed tip lines and triangles: upward triangles, highly active; downward triangles, lowly active; fill color indicates *a*-value. Open and filled circles denote bootstrap support values. Depth abbreviations: EPI, epipelagic; MES, mesopelagic; BAT, bathypelagic; ABY, abyssopelagic; Hadal, hadal zone.

(B) Genomic feature comparisons across divergence patterns. Dashed gray horizontal lines indicate overall mean values across all groups. Statistical significance assessed by two-sided Wilcoxon rank-sum tests. Only significant comparisons are shown. Significance: $^{****}p \leq 0.0001$, $^{***}p \leq 0.001$, $^{**}p \leq 0.01$, $^{*}p \leq 0.05$. Legend numeric labels correspond to divergence patterns in (A): non-hadal to hadal, $n = 21$; hadal to non-hadal, $n = 4$; widespread, $n = 11$; hadal-specific, $n = 486$.

(C) Heatmaps showing the prevalence of key metabolic genes in metagenomes (top) and their expression detected in metaproteomes (bottom) across divergence pattern groups. Numeric labels in the legend correspond to the divergence patterns defined in (A).

See also Figure S4 and Tables S2 and S5.

CellPress
OPEN ACCESS

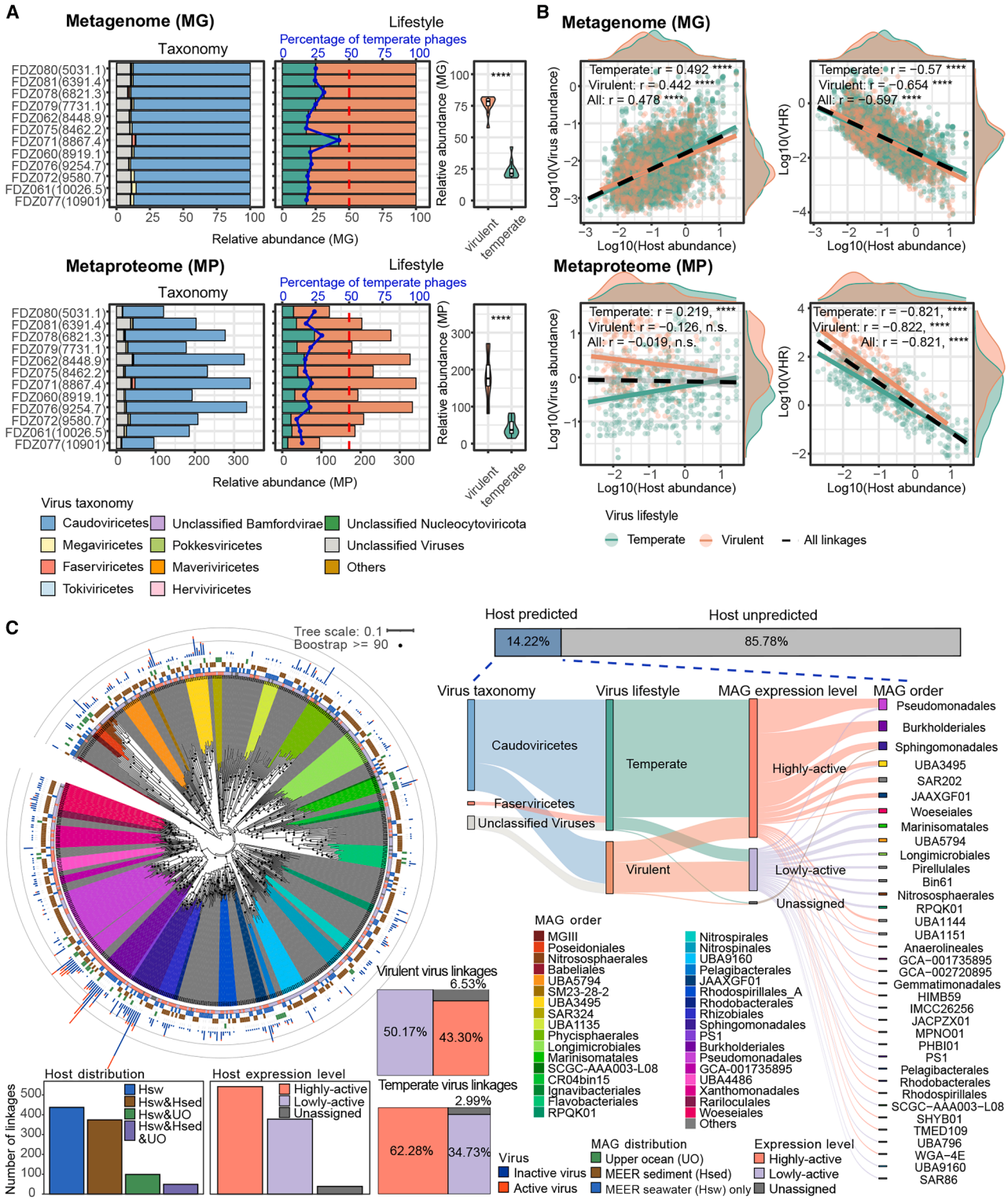


Figure 5. Active viral community in hadal seawater

(A) Viral community composition and lifestyle strategies in metagenomes (MG, top) and metaproteomes (MP, bottom). Stacked bar charts show relative abundances of viral taxonomic groups (left) and lifestyle categories (center). Samples are ordered by depth, with depths (m) in parentheses. In lifestyle charts, the red dashed line marks the 50% temperate virus threshold. The blue solid line tracks temperate virus proportion fluctuations. Violin plots (right) compare relative abundances of virulent and temperate viruses (n = 12). Significance assessed by two-sided Wilcoxon rank-sum tests (****$p \leq 0.0001$).

(legend continued on next page)

(median a = 0.67) and differed significantly from all other groups in genomic features (Wilcoxon rank-sum test, $p \leq 0.05$), possessing smaller genomes (median 1.61 Mb versus 2.58–2.79 Mb), fewer annotated KOs (918 versus 1,296–1,467), higher coding density (94.64% versus 90.78%–91.32%), and lower GC content (41.59% versus 50.96%–57.94%). These features are consistent with the genome streamlining strategy observed in pelagic strains adapted to oligotrophic conditions,[51] suggesting that these non-hadal-to-hadal immigrant taxa retain genomic signatures of their shallower-water origins and that their streamlined genomes may limit metabolic output in the hadal environment. By contrast, no significant genomic differences were detected among the other three groups, though widespread species encoded the most KOs on average (1,444), consistent with broad environmental tolerance, and hadal-specific species exhibited the highest GC content, possibly reflecting long-term adaptation to hadal selective pressures.

Protein expression profiles further delineated the ecological role of each biogeographic group (Figure 4C; Table S5). Hadal-specific species dominated most active metabolic functions and served as the primary contributors to biogeochemical cycling in the hadal zone. Widespread species showed a higher proportion of genomes expressing genes involved in heavy metal resistance, ROS detoxification, aromatic compound degradation, and compatible solute biosynthesis, consistent with versatile metabolic strategies enabling persistence across water-depth gradients. Hadal-to-non-hadal species preferentially expressed genes for oxidative stress defense and CO oxidation, potentially reflecting retained adaptive traits in the hadal zone. Non-hadal-to-hadal species, by contrast, exhibited limited expression of specialized pathways and instead prioritized functions supporting sustainability, suggesting that these upper-ocean-derived taxa emphasize physiological maintenance rather than metabolic breadth in the hadal zone.

Active viral community in hadal seawater

Beyond prokaryotes, *in situ* filtration also captured viruses associated with microorganisms and particles.[18] From metagenomic contigs, we identified a total of 28,718 species-level viral operational taxonomic units (vOTUs; Table S6; STAR Methods), from which 372,955 viral genes (including both full-length genes and partial fragments) were predicted. Of these, a total of 773 viral proteins from 682 vOTUs were detected in hadal metaproteomes (Table S6). The length distribution of active vOTUs closely mirrored that of the total vOTU dataset, with comparable median viral contig lengths, confirming that metaproteomic detection was not biased toward specific viral genome sizes (Figure S5B). Both metagenomic and metaproteomic analyses indicated that the viral community was dominated by Caudoviricetes (Figure 5A; Table S6). Metaproteomes confirmed the expression of a broad functional repertoire in the hadal virome, including virion structural proteins (capsid, tail, terminase, and connector), DNA/RNA processing enzymes, and lifecycle-associated genes for both lysogeny (integrases, recombinases, and repressors) and lysis (endolysins and release factors), indicating that both replication strategies are actively operating in hadal seawater (Table S6).

To further dissect the ecological roles of the two viral lifestyles, we investigated virus-host dynamics by correlating viral abundance with host abundance (Figure 5B). At the metagenomic level, both temperate and virulent viruses exhibited positive correlations with host abundance (r = 0.492 and 0.442, respectively), while their corresponding virus-to-host ratios (VHRs)[52] were negatively correlated with host abundance (r = −0.570 and −0.654, respectively). At the metaproteomic level, however, a clear divergence emerged: temperate viruses maintained a positive correlation with host protein abundance (r = 0.219), whereas virulent virus abundance showed a non-significant negative trend (r = −0.126). At both levels, temperate viruses exhibited shallower VHR declines than virulent viruses with increasing host abundance and preferentially targeted highly active hosts (62.28% of linkages; Figure 5C). These patterns showed general consistency with "Piggyback-the-Winner" (PtW) predictions in which lysogeny is favored at high host densities.[53]

However, several observations for virulent viruses extend beyond the PtW framework. First, virulent viruses accounted for a significantly higher proportion of both genomic and protein-level abundance than temperate viruses (Wilcoxon rank-sum tests, $p \leq 0.0001$) (Figure 5A; Table S6), deviating from the PtW expectation that lytic viral abundance should be less dominant at high host densities.[54,55] Second, virulent viruses showed no clear preference across host expression levels (a-value), infecting both lowly active taxa (e.g., Nitrososphaerales and Pelagibacterales), which are infrequently targeted by temperate viruses, and highly active taxa (e.g., Pseudomonadales and Burkholderiales) (Figure 5C; Table S2). This broad, non-selective infection pattern, particularly their association with highly active hosts, also departs from the PtW prediction that lytic infections are disfavored in highly active host populations where lysogeny tends to dominate.[54,55]

Collectively, these results reveal a dual viral strategy in the hadal zone. Temperate viruses conform to PtW predictions by preferentially associating with highly active hosts through lysogeny. However, virulent viruses diverge from PtW expectations: rather than being marginalized at high host densities, they remain numerically dominant and maintain non-selective associations across hosts spanning all expression levels. This

(B) Virus-host dynamics based on MG (top) and MP (bottom) data. Scatterplots show relationships between viral abundance or virus-host ratio (VHR) and host abundance for temperate (green) and virulent (orange) lineages. All values were $\log_{10}$-transformed. Shaded areas represent 95% confidence intervals. Each point represents a virus-host pair per sample. MG pairs: temperate, n = 4,721; virulent, n = 2,307; all, n = 7,028. MP pairs: temperate, n = 442; virulent, n = 225; all, n = 667. Significance of Pearson correlation coefficients: $^{****}p \leq 0.0001$, $^{***}p \leq 0.001$, $^{**}p \leq 0.01$, $^{*}p \leq 0.05$; n.s., $p > 0.05$.
(C) Virus-host linkage and infection patterns. Left: phylogenetic tree of all reconstructed prokaryotic MAGs, with background colors denoting order-level taxonomy. Rings indicate (inner to outer): host expression level, MAG distribution, and infecting virus count. Bar charts below summarize the distribution of virus-host linkages classified by host type. Treemaps show proportions of temperate versus virulent linkages. Right: Sankey diagram of active virus-host interactions from viral taxonomy and lifestyle to host expression levels and orders. Flow width proportional to linkage count.
See also Figure S5 and Tables S2 and S6.

complementary partitioning suggests that the two viral lifestyles occupy distinct ecological niches in the hadal ecosystem, jointly shaping microbial community structure and function in the deepest ocean.

Active microbial and viral metabolic profiles in hadal seawater

To construct a comprehensive functional profile of the hadal zone, we analyzed highly abundant and hadal-specific functional genes from the MEER-APD together with expressed viral genes. These genes were classified into six major functional categories: heavy metal resistance, carbon utilization, nitrogen metabolism, sulfur metabolism, sustainable functions, and viral structural and lifecycle functions (Figure 6; Tables S3 and S5).

Heavy metal resistance

Heavy metal resistance dominated the most abundant hadal-specific KOs (~20% of unique peptide abundance among the top 500; Figure 6A; Table S3). Six of the eight most abundant MEER-APD-specific KOs encoded heavy metal efflux systems, led by the *czcABC* efflux system for $Co^{2+}/Zn^{2+}/Cd^{2+}$ transport,[56] followed by *cusB* for Cu^{+}/Ag^{+},[57,58] *golTS* for Au^{+},[59] and *chrA* for CrO_4^{2-}.[60,61] Beyond the efflux systems, proteins involved in heavy metal oxidation and resistance were also prominent among the most abundant MEER-APD-specific KOs, including copper oxidation (*mmcO*)[62] and resistance (*pcoBC* and *cutF*)[63,64] and manganese (Mn^{2+}) oxidation (*moxA*).[65] Additionally, tellurium resistance (*terD* for TeO_3^{2-})[66] and arsenate detoxification (*arsC* for AsO_4^{3-})[67] were exclusively expressed in the hadal zone.

We quantified the concentrations of the corresponding metals (i.e., Cu, Mn, Zn, As, Hg, and Se) in the hadal seawater samples collected during this study and compared them with published upper-ocean data (0–1,000 m; Table S7; STAR Methods). All measured elements except arsenic were significantly elevated in hadal waters (Wilcoxon rank-sum test, $p \leq 0.0001$, Figure S6A), revealing active resistance to a broad spectrum of previously unrecognized heavy metals (e.g., Mn, Au, Te, and Ag) that extends far beyond previously assumed.[68] Notably, the active expression of the arsenic detoxification gene (*arsC*) indicates that even background arsenic levels in hadal seawater impose persistent selective pressure on the hadal microbiome.[69,70]

Carbon utilization

MEER-APD revealed ROM as a key carbon source in the hadal zone (Figure 6B; Table S3). Hadal-specific expression of aromatic compound transporters (*shiA* and *benK*) and downstream degradation enzymes, including *pcaF*, *xylXYZ*, and *xylC* for monoaromatics, and *phdF*, *nahAb*, and *pcpD* for polycyclic and halogenated aromatics, demonstrated active funneling of diverse aromatic substrates through benzoate and catechol intermediates into the tricarboxylic acid (TCA) cycle.[71] Dehalogenases (*dhaA*, *dehH*, and *pceA*) and D-amino acid utilization enzymes (*dauA* and *dsdA*) were also exclusively expressed in hadal samples (Table S3). These findings support the hypothesis that ROMs that are commonly considered biologically recalcitrant[14] serve as bioavailable substrates under hadal nutrient limitation, consistent with the documented enrichment of aromatic compounds, halogenated hydrocarbons, and D-amino acids in deep-sea environments.[72–74] By contrast, enzymes for other carbon degradation, including those for alkanes, lignin, chitin, and L-sugars, were barely detectable despite their genomic presence.[15,68,75,76]

Additionally, CO oxidation genes (*coxLMS*) were expressed across all samples and diverse lineages (Figure 3E; Table S4), indicating a widespread alternative energy source.[47,48] Conversely, for carbon fixation, only Rubisco (*rbcSL*) was detected. Although Rubisco is known to be highly abundant in cellular proteomes,[77] MaxLFQ-based quantification showed that its intensity in hadal samples fell within the general proteome dynamic range rather than dominating the signal (Figure S7A), indicating that the absence of other carbon fixation pathways (Wood-Ljungdahl, reductive tricarboxylic acid (rTCA) cycle, or 3HP-4HB) is unlikely a detection artifact. Potential electron donors identified in Rubisco-expressing MAGs included CO (SAR324) and organic matter/metal ions (Burkholderiales) (Figure S7B; Table S4), suggesting versatile mixotrophic strategies supporting the Calvin-Benson-Bassham (CBB) cycle.

Viral AMGs further augmented host carbon metabolism, with expressed genes participating in aminobenzoate degradation (*glnA*), CO oxidation (*coxS*), pyruvate metabolism (*pps*, *ldhA*, and *butA*), TCA cycle-associated organic acid assimilation (*acs*, *prpE*, and *lcd*), and the glyoxylate shunt (*aceA*). These viral contributions suggest that hadal viruses actively facilitate host adaptation to nutrient-limited conditions by enhancing refractory carbon utilization and energy acquisition (Table S6).

Nitrogen metabolism

MEER-APD confirmed active internal nitrogen cycling in hadal seawaters (Figure 6C; Table S3), consistent with metagenomic predictions.[76,78–80] Proteins involved in the transport and sensing of nitrogen compounds (e.g., nitrate, nitrite, and ammonia) were widely detected, confirming the importance of nitrogen transformation for hadal microbial communities (Figure 6C). Both dissimilatory (*napA* and *narGH*) and assimilatory (*nasC*) nitrate reductases were detected, indicating nitrate as a major terminal electron acceptor. Downstream denitrification enzymes (*nirK*, *norB*, and *nosZ*) were also identified; however, the higher abundance of *nirK* relative to *norB* and *nosZ* suggests limited nitrogen loss via N_2O or N_2 production (Figure 6C). Instead, nitrite reduction to ammonium via *nirBD* was prominent, supporting nitrogen retention through dissimilatory nitrite reduction to ammonium (DNRA). The *narGH* complex, functioning as a reversible nitrite oxidoreductase, likely mediates nitrite re-oxidation to nitrate. Coupled with widespread ammonia oxidation gene (*amoABC*) expression, this establishes a complete nitrification capacity and an efficient internal nitrogen recycling loop. No bacterial MAGs harboring *amo* genes were recovered, indicating that ammonia-oxidizing archaea (AOA) rather than ammonia-oxidizing bacteria (AOB) dominate ammonia oxidation in hadal seawaters (Figure S6B). Virally expressed *aguB* further supplements the host nitrogen pool by releasing ammonium from N-carbamoylputrescine hydrolysis (Table S6). Overall, nitrogen cycling in the hadal zone is characterized by tightly coupled nitrification, DNRA, and internal recycling, with limited N_2 loss through complete denitrification.

Sulfur metabolism

Hadal sulfur metabolism centered on inorganic redox transformations (Figure 6D; Table S3). Transport proteins (*cysAPW*) indicated sulfate and thiosulfate as the primary inorganic sulfur sources. Sulfate reduction proceeded via dissimilatory (*sat*)

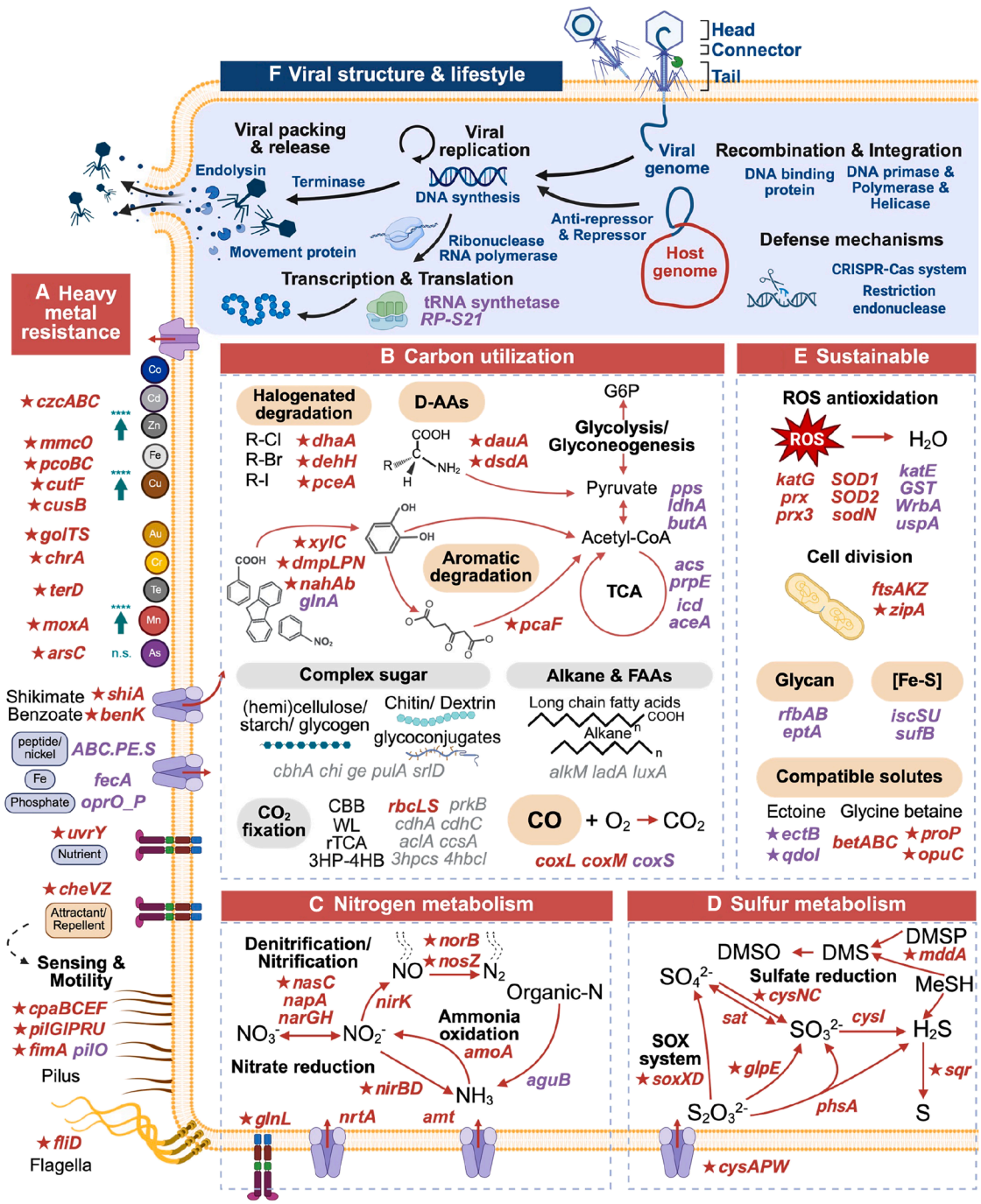


Figure 6. Active microbial and viral metabolic profiles in hadal seawater

(A–F) Functional categories: (A) heavy metal resistance, (B) carbon utilization, (C) nitrogen metabolism, (D) sulfur metabolism, (E) sustainable functions, and (F) viral structure and lifestyle. All genes were identified by metaproteomic analyses. Genes highlighted in red italics represent highly abundant microbial genes in the MEER-APD or genes with specific functional relevance. Genes shown in blue correspond to viral structural proteins or genes involved in viral replication and life cycle. Purple genes indicate virus-encoded AMGs that overlap with genes identified in the MEER-APD (only genes discussed in the main text are shown). Gray text denotes genes not detected in the MEER-APD but supported by metagenomic annotations or genes that have been frequently reported in hadal zone-related studies. Genes expressed exclusively in hadal samples are marked with an asterisk. Red solid lines indicate substance transformation or transport. In (A), upward arrows indicate higher metal ion concentrations in hadal versus surface waters. Significance: $^{****}p \leq 0.0001$, $^{***}p \leq 0.001$, $^{**}p \leq 0.01$, $^{*}p \leq 0.05$; n.s., $p > 0.05$ (two-sided Wilcoxon rank-sum tests).

Created in BioRender. Zhao, W. (2026) https://BioRender.com/yl87tdf.

See also Figures S6 and S7 and Tables S3, S6, and S7.

and assimilatory (*cysNC*, *cysN*, and *cysI*) routes. Thiosulfate metabolism was particularly versatile: sulfur oxidation (SOX) system proteins (*soxXD*) mediated oxidation to sulfate, thiosulfate reductase (*phsA*) enabled disproportionation to sulfite and sulfide, and sulfurtransferase (*glpE*) participated in cyanide detoxification.[81] Sulfide:quinone oxidoreductase (*sqr*) indicated active H_2S oxidation coupled to the quinone pool. Most thiosulfate-related enzymes were co-expressed in Burkholderiales (Figure 3E; Table S4). We also examined organic sulfur cycling, which dominates upper ocean sulfur metabolism through remineralization of sulfur-containing dissolved organic matter.[82] In the hadal seawaters, however, dimethylsulfoniopropionate (DMSP) cleavage enzyme *DddL* and downstream enzymes for methanethiol transmethylation and DMS oxidation were detected only at negligible abundance, and no DMSP biosynthesis enzymes were observed.

Other sustainable functions

MEER-APD captured active cellular processing and environmental response in hadal microorganisms (Figure 6E; Table S3). Cell division proteins (*ftsAHZ* and *zipA*), environmental sensing, motility (flagellar and pili components), and nutrient uptake proteins (phosphate and iron transporters) demonstrated that hadal microorganisms actively forage for limiting resources. Coping with the oxidative stress imposed by high hydrostatic pressure[46] appeared to be a universal requirement: antioxidation enzymes (*SOD1/2*, *sodN*, *katGE*, and *prx/prx3*) were ubiquitous across all metaproteomes, supplemented by viral AMGs (*GST*, *WrbA*, and *UspA*) that likely augment host ROS defense (Table S6). Additionally, Fe–S cluster assembly systems, both constitutive (ISC) and stress-induced (SUF), were co-expressed in microbial and viral proteins, further highlighting the centrality of redox homeostasis (Tables S4 and S6).[83,84] For compatible solutes for pressure adaptation, hadal microorganisms showed selective reliance on glycine betaine (biosynthesis: *betABC*; transport: *opuC*, *proP*; detected in 11/12 samples) and ectoine (*ectB*; all samples), whereas trimethylamine N-oxide (TMAO) and DMSP synthesis enzymes were rarely detected despite their reported importance in shallower deep-sea regions,[85–87] suggesting that glycine betaine and ectoine serve as the primary piezolytes at hadal depths.

Distinct microbially mediated element cycling in the deepest ocean

By integrating the active metabolic profiles described above with published upper ocean metaproteomic studies,[38,82,88–91] we identified pronounced differences in both active microbial metabolism and the resultant element cycling between the upper ocean (20–4,000 m) and the hadal zone (>6,000 m) (Figure 7). In the upper ocean, active microbial metabolism is fueled by photosynthetically derived labile substrates and dominated by aerobic respiration. However, in the hadal zone, under extreme hydrostatic pressure and chronic nutrient limitation, metaproteomic evidence reveals a fundamentally different metabolic regime characterized by diversified carbon acquisition and expanded energy-yielding pathways, which collectively reshape hadal biogeochemical cycling.

This regime is most evident in carbon cycling, where hadal microorganisms simultaneously diversify both carbon and energy sources to compensate for the scarcity of labile substrates. ROM, particularly aromatics, halogenated compounds, and D-amino acids, serves as the preferred carbon source for hadal microorganisms, replacing the photosynthetically derived organics that dominate the upper ocean. By contrast, CO_2 fixation remains minimal in the hadal zone with only Rubisco (CBB cycle) detectably expressed at the protein level, strikingly different from the upper ocean where multiple autotrophic pathways (CBB, 3HP–4HB, and rTCA) are concurrently active.[38,82,89] Together, these patterns redefine the hadal zone as an active transformation zone for complex organic carbon rather than a site of significant autotrophic production.

For energy acquisition, hadal microorganisms exploit a remarkably expanded repertoire of electron donors and acceptors: nitrate and sulfate function as alternative terminal electron acceptors alongside oxygen, thiosulfate serves as both electron donor (via SOX-mediated oxidation) and acceptor (via reductive disproportionation), and heavy metal redox transformations (e.g., Fe^{2+}, Mn^{2+}, Cu^{+}, and AsO_3^{2-}) couple detoxification with electron transfer and energy conservation.[92,93] This diversified energy metabolism, in turn, reshapes nitrogen, sulfur, and metal cycling in ways distinct from the upper ocean. For nitrogen, the most striking difference is the convergence of multiple ammonium-generating pathways in the hadal zone, where DNRA channels nitrate back to ammonium rather than to N_2, and microbially and virally mediated hydrolysis of polyamines and other organic nitrogen compounds provides additional ammonium inputs. This convergence of ammonium sources may help explain the sustained ammonia oxidation activity observed throughout the energy-starved dark ocean, where ammonium concentrations are typically near detection limits yet ammonia oxidizers maintain continuous activity.[80] Sulfur cycling shifts from the DMSP-dominated organic transformations of the upper ocean[38,89] to inorganic redox processes centered on thiosulfate, serving dual roles in assimilation and energy generation. Metal redox cycling of Cu, Fe, and As, largely absent from upper ocean metaproteomes, emerges as a hadal-specific feature coupling detoxification with bioenergetics.

Notably, hadal viruses actively participate in these adaptive processes. Hadal viral AMGs contribute to ROM degradation and CO oxidation, in place of the photosystem augmentation characteristic of upper ocean cyanophages.[82] Additionally, virus-encoded antioxidant genes and polyamine hydrolases directly support host stress defense and ammonium recycling. Crucially, this metabolic reprogramming is mediated by both temperate and virulent viruses. While temperate viruses preferentially associate with highly active hosts—reflecting the PtW model—to enable sustained AMG expression via lysogenic integration, virulent viruses also actively contribute to elemental cycling, exemplified by the expression of *glnA* for aromatic compound degradation.

DISCUSSION

Here, we established a workflow for resolving active microbial processes in low-biomass hadal seawater. By integrating large-volume *in situ* filtration with DNA-protein co-extraction, we generated MEER-APD, a systematic catalog of actively expressed proteins in hadal microbiomes. Although hadal biomass is one to two orders of magnitude lower than in the upper ocean,[30,31] this approach identified more proteins than previous surface-water studies, thereby providing a high-resolution view

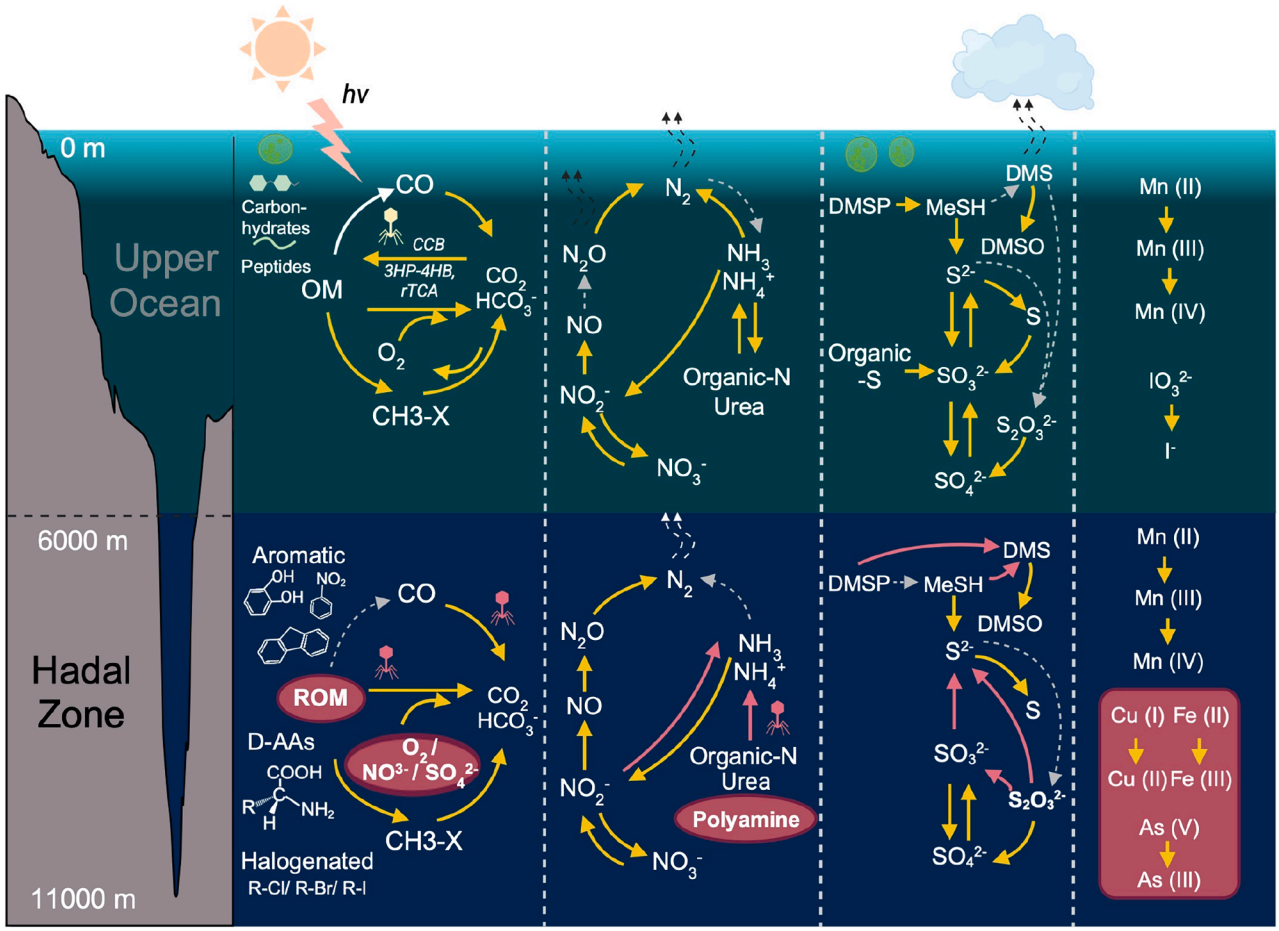


Figure 7. Distinct microbially mediated element cycling in the deepest ocean

Schematic comparison of major biogeochemical processes in the upper ocean (20–4,000 m) and hadal zone (>6,000 m). Solid arrows indicate metaproteomic-supported pathways; red arrows, hadal-predominant pathways; white arrow, non-microbially driven process; gray dashed arrows, pathways lacking metaproteomic support. Purple boxes highlight substrates actively utilized in the hadal zone. Viral icons denote processes involving viral proteins. The diagram summarizes depth-related differences in carbon, nitrogen, sulfur, and metal cycling inferred from protein expression.

Created in BioRender. Zhao, W. (2026) https://BioRender.com/ycw8bfi.

See also Tables S3 and S6.

of the distinct microbial activities in Earth's deepest ocean ecosystems. This enhanced performance demonstrates its potential for studying other challenging environments where biomass is limited and samples are precious.

MEER-APD confirmed many previous hypotheses of hadal adaptation at the proteomic level but also challenged some previous assumptions from metagenomes and cultures. For example, expression of enzymes for alkane utilization was barely detected despite high genetic potential.[15,68,75,76] Carbon fixation was limited to Rubisco expression, rather than the multiple pathways predicted by metagenomes.[75,94] Nitrogen cycling showed lower nitrogen loss potential than previously hypothesized[15] due to the low expression of nitric oxide reductase (*norB*) and nitrous oxide reductase (*nosZ*). These findings highlight the value of activity-level evidence for refining genomic predictions of hadal element cycling. Similarly, genes linked to TMAO and DMSP, previously proposed as key high-pressure piezolytes,[86,95] showed minimal expression, whereas glycine betaine- and ectoine-related genes were abundantly expressed. These results prompt us to reconsider microbial adaptation to hadal high pressure, suggesting either different piezolyte preferences among hadal microorganisms or alternative mechanisms for biomolecular stabilization under ultra-high pressure.

The systematic metaproteomic landscape of the hadal zone reveals a microbial ecosystem shaped by two overarching adaptive strategies: diversification of carbon sources and expansion of energy-yielding pathways. These strategies, while enabling survival under high hydrostatic pressure and chronic nutrient limitation, have profound implications for hadal biogeochemical cycling that could not be predicted from metagenomics alone. The *a*-value, which reflects the protein contribution in relation to metagenomic abundance, distinguished highly active and lowly active taxa. The former possess larger genomes and broader metabolic repertoires, while the latter exhibit streamlined genomes and specialized functional profiles. This pattern is generally consistent with the "streamlined versus versatile" framework recently established for hadal sediment microbiomes through the MEER metagenomic survey,[1] in which environmental selection favors either

genomically compact specialists or metabolically diverse generalists. However, our activity-based classification does not fully align with the classical copiotroph-oligotroph dichotomy,[44,45] which was developed primarily around nutrient availability in well-studied conventional environments. Highly active taxa lacked several canonical copiotrophic features[45] and preferentially associated with temperate rather than lytic viruses, contradicting the "Kill-the-Winner" dynamics expected for fast-growing copiotrophs.[44] These deviations suggest that the physical extremes of the hadal zone, particularly high hydrostatic pressure, may reshape microbial nutrient utilization in ways not captured by the copiotroph-oligotroph frameworks alone. Indeed, our metaproteomic results raise the possibility that extreme pressure expands the metabolic accessibility of carbon and energy resources that are otherwise unfavorable in shallower marine environments. If so, current ocean biogeochemical models, largely parameterized from surface and mesopelagic data,[96] may underestimate the metabolic capacity of hadal microbiomes by overlooking the influence of hydrostatic pressure on resource accessibility. Beyond individual strategies, the co-existence of highly and lowly active taxa may also enable syntrophic coupling that sustains hadal ecosystem functioning. For example, lowly active Nitrososphaerales dominate ammonia oxidation, whereas highly active Pseudomonadales, Burkholderiales, and Nitrospirales mediate DNRA, collectively maintaining bioavailable ammonium in the hadal ecosystem.

These metaproteomic insights can guide the cultivation of uncultured hadal microorganisms by delineating their physicochemical and metabolic niches. Regardless of their *a*-values, only a small fraction of these taxa includes successfully isolated strains to date, primarily from Pseudomonadales,[97–99] Rhodobacterales,[100,101] and Sphingomonadales.[102,103] Other highly active taxa, such as Burkholderiales, have never been isolated. This recalcitrance likely reflects a mismatch between conventional media and hadal nutrient preferences, especially under high pressure. Our data indicate that uncultured highly active groups such as Burkholderiales rely on substrates rarely included in enrichment media, particularly complex ROMs, heavy metals, and thiosulfate. By contrast, lowly active groups such as AOA show hadal-specific specialization relative to shallow-water counterparts.[104] Although their genomes encode pathways for carbon fixation, hadal AOA predominantly express genes for ammonia oxidation and ROS detoxification, suggesting that they function as obligate ammonia oxidizers, which therefore require distinct cultivation strategies. Collectively, these metaproteomic results provide guidance for designing targeted cultivation strategies for hadal microorganisms, especially under high-pressure cultivation conditions. Future efforts should consider specific metabolic requirements, such as supplying ROMs as carbon sources and diverse energy sources beyond oxygen, as well as exploring co-cultivation with partner species or phages.

Finally, our results indicate that viruses shape microbial adaptation in the hadal seawaters. Temperate viruses conform to PtW predictions by preferentially associating with highly active hosts, where lysogenic integration may sustain AMG expression and enhance host metabolic capacity, linking viral persistence to hadal elemental cycling. By contrast, virulent viruses remain dominant in the active virome despite decreasing VHR with increasing host protein abundance and infect hosts across broad expression levels. Their active AMG expression suggests ecological roles beyond transient lysis, consistent with persistent, non-canonical infections that enable co-existence without classical lysogeny.[105,106] The hadal zone may further favor such non-lytic persistence, as co-existence with hosts may be a more energy-efficient strategy than extensive viral replication in this nutrient-limited environment.[107]

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Xiang Xiao (zjxiao2018@sjtu.edu.cn).

Materials availability

This study did not generate any new, unique reagents.

Data and code availability

- The metagenomic data have been deposited in the National Omics Data Encyclopedia (NODE, https://www.biosino.org/node) with the project identifier OEP00004557.
- The mass spectrometry proteomics data have been deposited in the iProX repository (https://www.iprox.cn/page/home.html) with the dataset identifier PXD045060.
- All original code and related tables have been deposited at Zenodo and are publicly available at 10.5281/zenodo.20785637.
- Any additional information required to reanalyze the data reported in this work is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank the Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences (IDSSE CAS), for providing critical access to R/V Tan Suo Yi Hao as well as the HOV *Fendouzhe* and the crews and engineers for their technical support. This work was financially supported by the National Key R&D Program of China (grant nos. 2022YFC2805404 and 2022YFC2805300), the National Natural Science Foundation of China (grant nos. 42330206, 32530002, 92451301, 32570058, 42076127, 91751108, and 42106087), the Key Research and Development Program of Hainan Province (grant no. ZDKJ2021028), the Major Science and Technology Research Project of Sanya Yazhou Bay Science and Technology City (grant no. SKJC-2021-01-001), and the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDA22040600). We thank Prof. Douglas Bartlett from the Scripps Institution of Oceanography for his constructive comments and careful editing of the manuscript, Xiangyu Wang from Shanghai Jiao Tong University for assistance with DNA and protein co-extractions, and Dr. Jingli Hou from the Instrumental Analysis Center of Shanghai Jiao Tong University for contributions to optimizing enzymatic cleavage methods and mass spectrometry analysis. We acknowledge support from the Center for High Performance Computing (HPC) at the Shanghai Jiao Tong University and the Yazhou Bay Science and Technology City Advanced Computing Center (YZBSTCACC). This work was supported by the Project of the State Key Laboratory of Microbial Metabolism, Shanghai Jiao Tong University (no. SKLMMZD_02). We acknowledge support from the Ministry of Education of China, through the Extremophile Resource Division of the Engineered Microorganisms and Genetic Tools Collection Center of China (EMTCC), hosted at Shanghai Jiao Tong University. The author Xiang Xiao is a SANS Exploration Scholar. We sincerely thank the reviewers for their constructive comments, which greatly improved the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, X.X., W.Z., and W.-J.Z.; methodology, W.-J.Z., W.Z., L.L., A.H., Z. Wu, J.W., T.Z., and Y.H.; software, A.H., L.L., C.L., Y.W., and Z. Wu; investigation, W.-J.Z., W.Z., and X.X.; resources, X.X., S.L., W.-J.Z., W.Z., L.M., C.L., Y.W., Z. Wei, R.L., and J.L.; writing – original draft, W.Z., A.H., Z.

Wu, and W.-J.Z.; writing – review & editing, W.Z., W.-J.Z., X.X., Z. Wu, and A.H.; funding acquisition, X.X., S.L., L.M., W.-J.Z., and W.Z.; supervision, W.Z. and X.X.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Sample collection and processing
  - Heavy metal concentration measurements
  - Coextraction of DNA and protein
  - Enzymatic cleavage of proteins and purification of peptides
  - Mass spectrometry data acquisition
  - Sequencing data processing, metagenomics assembly, and binning
  - Gene annotation and metaproteomic database construction
  - Metaproteomic searching, peptide validation, protein inference and quantification
  - Metagenomic and metaproteomic relative abundance of contig genes/proteins
  - Metagenomic and metaproteomic relative abundance of MAGs
  - Correlations between metagenomic abundance and metaproteomic expression
  - Comparing the metaproteome from hadal seawater with other metaproteomic datasets
  - Phylogenetic inference of evolutionary dynamics and biogeographic patterns
  - Viral sequence identification, clustering, taxonomic assignment, and lifestyle prediction
  - Prediction of virus–host linkage
  - vOTU abundance calculation
  - Viral gene annotation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chom.2026.07.001.

Received: April 29, 2026
Revised: May 26, 2026
Accepted: July 1, 2026
Published: July 28, 2026

REFERENCES

1. Xiao, X., Zhao, W., Song, Z., Qi, Q., Wang, B., Zhu, J., Lin, J., Wang, J., Hu, A., Huang, S., et al. (2025). Microbial ecosystems and ecological driving forces in the deepest ocean sediments. Cell *188*, 1363–1377.e9. https://doi.org/10.1016/j.cell.2024.12.036.
2. Stewart, H.A., and Jamieson, A.J. (2018). Habitat heterogeneity of hadal trenches: Considerations and implications for future studies. Prog. Oceanogr. *161*, 47–65. https://doi.org/10.1016/j.pocean.2018.01.007.
3. Jamieson, A.J., Fujii, T., Mayor, D.J., Solan, M., and Priede, I.G. (2010). Hadal trenches: the ecology of the deepest places on Earth. Trends Ecol. Evol. *25*, 190–197. https://doi.org/10.1016/j.tree.2009.09.009.
4. Liu, R., Wei, X., Song, W., Wang, L., Cao, J., Wu, J., Thomas, T., Jin, T., Wang, Z., Wei, W., et al. (2022). Novel Chloroflexi genomes from the deepest ocean reveal metabolic strategies for the adaptation to deep-sea habitats. Microbiome *10*, 75. https://doi.org/10.1186/s40168-022-01263-6.
5. Liu, R., Wang, L., Wei, Y., and Fang, J. (2018). The hadal biosphere: Recent insights and new directions. Deep Sea Res. II *155*, 11–18. https://doi.org/10.1016/j.dsr2.2017.04.015.
6. Peoples, L.M., and Bartlett, D.H. (2017). Ecogenomics of Deep-Ocean Microbial Bathytypes. In Microbial Ecology of Extreme Environments, C. Chénard and F.M. Lauro, eds. (Springer), pp. 7–50. https://doi.org/10.1007/978-3-319-51686-8_2.
7. Peoples, L.M., Norenberg, M., Price, D., McGoldrick, M., Novotny, M., Bochdansky, A., and Bartlett, D.H. (2019). A full-ocean-depth rated modular lander and pressure-retaining sampler capable of collecting hadal-endemic microbes under in situ conditions. Deep Sea Res. I *143*, 50–57. https://doi.org/10.1016/j.dsr.2018.11.010.
8. Zhang, W.-J., Zhang, C., Zhou, S., Li, X.-G., Mangenot, S., Fouteau, S., Guerin, T., Qi, X.-Q., Yang, J., Bartlett, D.H., et al. (2021). Comparative genomic analysis of obligately piezophilic Moritella yayanosii DB21MT-5 reveals bacterial adaptation to the Challenger Deep, Mariana Trench. Microb. Genomics *7*, 000591. https://doi.org/10.1099/mgen.0.000591.
9. Peoples, L.M., Grammatopoulou, E., Pombrol, M., Xu, X., Osuntokun, O., Blanton, J., Allen, E.E., Nunnally, C.C., Drazen, J.C., Mayor, D.J., et al. (2019). Microbial Community Diversity Within Sediments from Two Geographically Separated Hadal Trenches. Front. Microbiol. *10*, 347. https://doi.org/10.3389/fmicb.2019.00347.
10. Hiraoka, S., Hirai, M., Matsui, Y., Makabe, A., Minegishi, H., Tsuda, M., Juliarni, R., Rastelli, E., Danovaro, R., Corinaldesi, C., et al. (2020). Microbial community and geochemical analyses of trans-trench sediments for understanding the roles of hadal environments. ISME J. *14*, 740–756. https://doi.org/10.1038/s41396-019-0564-z.
11. Schauberger, C., Thamdrup, B., Lemonnier, C., Trouche, B., Poulain, J., Wincker, P., Arnaud-Haond, S., Glud, R.N., and Maignien, L. (2024). Metagenome-assembled genomes of deep-sea sediments: changes in microbial functional potential lag behind redox transitions. ISME Commun. *4*, ycad005. https://doi.org/10.1093/ismeco/ycad005.
12. Liu, R., Wang, Z., Wang, L., Li, Z., Fang, J., Wei, X., Wei, W., Cao, J., Wei, Y., and Xie, Z. (2020). Bulk and Active Sediment Prokaryotic Communities in the Mariana and Mussau Trenches. Front. Microbiol. *11*, 1521. https://doi.org/10.3389/fmicb.2020.01521.
13. Jing, H., Zhu, W., Liu, H., Zheng, L., and Zhang, Y. (2018). Particle-Attached and Free-Living Archaeal Communities in the Benthic Boundary Layer of the Mariana Trench. Front. Microbiol. *9*, 2821. https://doi.org/10.3389/fmicb.2018.02821.
14. Gao, Z.M., Huang, J.M., Cui, G.J., Li, W.L., Li, J., Wei, Z.F., Chen, J., Xin, Y.Z., Cai, D.S., Zhang, A.Q., et al. (2019). *In situ* meta-omic insights into the community compositions and ecological roles of hadal microbes in the Mariana Trench. Environ. Microbiol. *21*, 4092–4108. https://doi.org/10.1111/1462-2920.14759.
15. Zhou, Y.-L., Mara, P., Cui, G.-J., Edgcomb, V.P., and Wang, Y. (2022). Microbiomes in the Challenger Deep slope and bottom-axis sediments. Nat. Commun. *13*, 1515. https://doi.org/10.1038/s41467-022-29144-4.
16. Nunoura, T., Takaki, Y., Hirai, M., Shimamura, S., Makabe, A., Koide, O., Kikuchi, T., Miyazaki, J., Koba, K., Yoshida, N., et al. (2015). Hadal biosphere: Insight into the microbial ecosystem in the deepest ocean on Earth. Proc. Natl. Acad. Sci. USA *112*, E1230–E1236. https://doi.org/10.1073/pnas.1421816112.
17. Jian, H., Yi, Y., Wang, J., Hao, Y., Zhang, M., Wang, S., Meng, C., Zhang, Y., Jing, H., Wang, Y., et al. (2021). Diversity and distribution of viruses inhabiting the deepest ocean on Earth. ISME J. *15*, 3094–3110. https://doi.org/10.1038/s41396-021-00994-y.
18. Gao, C., Liang, Y., Jiang, Y., Paez-Espino, D., Han, M., Gu, C., Wang, M., Yang, Y., Liu, F., Yang, Q., et al. (2022). Virioplankton assemblages from challenger deep, the deepest place in the oceans. iScience *25*, 104680. https://doi.org/10.1016/j.isci.2022.104680.
19. Zhou, Y.-L., Mara, P., Vik, D., Edgcomb, V.P., Sullivan, M.B., and Wang, Y. (2022). Ecogenomics reveals viral communities across the Challenger

Deep oceanic trench. Commun. Biol. *5*, 1055. https://doi.org/10.1038/s42003-022-04027-y.

20. Breitbart, M., Thompson, L., Suttle, C., and Sullivan, M. (2007). Exploring the Vast Diversity of Marine Viruses. Oceanography *20*, 135–139. https://doi.org/10.5670/oceanog.2007.58.

21. Tian, F., Wainaina, J.M., Howard-Varona, C., Domínguez-Huerta, G., Bolduc, B., Gazitúa, M.C., Smith, G., Gittrich, M.R., Zablocki, O., Cronin, D.R., et al. (2024). Prokaryotic-virus-encoded auxiliary metabolic genes throughout the global oceans. Microbiome *12*, 159. https://doi.org/10.1186/s40168-024-01876-z.

22. Moran, M.A., Satinsky, B., Gifford, S.M., Luo, H., Rivers, A., Chan, L.-K., Meng, J., Durham, B.P., Shen, C., Varaljay, V.A., et al. (2013). Sizing up metatranscriptomics. ISME J. *7*, 237–243. https://doi.org/10.1038/ismej.2012.94.

23. Steiner, P.A., De Corte, D., Geijo, J., Mena, C., Yokokawa, T., Rattei, T., Herndl, G.J., and Sintes, E. (2019). Highly variable mRNA half-life time within marine bacterial taxa and functional genes. Environ. Microbiol. *21*, 3873–3884. https://doi.org/10.1111/1462-2920.14737.

24. Moore, E., Harvey, H., Faux, J., Goodlett, D., and Nunn, B. (2014). Protein recycling in Bering Sea algal incubations. Mar. Ecol. Prog. Ser. *515*, 45–59. https://doi.org/10.3354/meps10936.

25. Pelikan, C., Wasmund, K., Glombitza, C., Hausmann, B., Herbold, C.W., Flieder, M., and Loy, A. (2021). Anaerobic bacterial degradation of protein and lipid macromolecules in subarctic marine sediment. ISME J. *15*, 833–847. https://doi.org/10.1038/s41396-020-00817-6.

26. Giljan, G., Kamennaya, N.A., Otto, A., Becher, D., Ellrott, A., Meyer, V., Murton, B.J., Fuchs, B.M., Amann, R.I., and Zubkov, M.V. (2020). Bacterioplankton reveal years-long retention of Atlantic deep-ocean water by the Tropic Seamount. Sci. Rep. *10*, 4715. https://doi.org/10.1038/s41598-020-61417-0.

27. Saito, M.A., Bertrand, E.M., Duffy, M.E., Gaylord, D.A., Held, N.A., Hervey, W.J., Hettich, R.L., Jagtap, P.D., Janech, M.G., Kinkade, D.B., et al. (2019). Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing. J. Proteome Res. *18*, 1461–1476. https://doi.org/10.1021/acs.jproteome.8b00761.

28. Brum, J.R., Ignacio-Espinoza, J.C., Kim, E.-H., Trubl, G., Jones, R.M., Roux, S., VerBerkmoes, N.C., Rich, V.I., and Sullivan, M.B. (2016). Illuminating structural proteins in viral "dark matter" with metaproteomics. Proc. Natl. Acad. Sci. USA *113*, 2436–2441. https://doi.org/10.1073/pnas.1525139113.

29. Hawley, A.K., Torres-Beltrán, M., Zaikova, E., Walsh, D.A., Mueller, A., Scofield, M., Kheirandish, S., Payne, C., Pakhomova, L., Bhatia, M., et al. (2017). A compendium of multi-omic sequence information from the Saanich Inlet water column. Sci. Data *4*, 170160. https://doi.org/10.1038/sdata.2017.160.

30. Amano, C., Zhao, Z., Sintes, E., Reinthaler, T., Stefanschitz, J., Kisadur, M., Utsumi, M., and Herndl, G.J. (2022). Limited carbon cycling due to high-pressure effects on the deep-sea microbiome. Nat. Geosci. *15*, 1041–1047. https://doi.org/10.1038/s41561-022-01081-3.

31. Peoples, L.M., Donaldson, S., Osuntokun, O., Xia, Q., Nelson, A., Blanton, J., Allen, E.E., Church, M.J., and Bartlett, D.H. (2018). Vertically distinct microbial communities in the mariana and kermadec trenches. PLoS One *13*, e0195102. https://doi.org/10.1371/journal.pone.0195102.

32. Wang, D.-Z., Xie, Z.-X., and Zhang, S.-F. (2014). Marine metaproteomics: current status and future directions. J. Proteom. *97*, 27–35. https://doi.org/10.1016/j.jprot.2013.08.024.

33. Nebauer, D.J., Pearson, L.A., and Neilan, B.A. (2024). Critical steps in an environmental metaproteomics workflow. Environ. Microbiol. *26*, e16637. https://doi.org/10.1111/1462-2920.16637.

34. Du, M., Peng, X., Zhang, H., Ye, C., Dasgupta, S., Li, J., Li, J., Liu, S., Xu, H., Chen, C., et al. (2021). Geology, environment, and life in the deepest part of the world's oceans. Innovation *2*, 100109. https://doi.org/10.1016/j.xinn.2021.100109.

35. McCain, J.S.P., Tagliabue, A., Susko, E., Achterberg, E.P., Allen, A.E., and Bertrand, E.M. (2021). Cellular costs underpin micronutrient limitation in phytoplankton. Sci. Adv. *7*, eabg6501. https://doi.org/10.1126/sciadv.abg6501.

36. McIlvin, M.R., and Saito, M.A. (2021). Online Nanoflow Two-Dimension Comprehensive Active Modulation Reversed Phase-Reversed Phase Liquid Chromatography High-Resolution Mass Spectrometry for Metaproteomics of Environmental and Microbiome Samples. J. Proteome Res. *20*, 4589–4597. https://doi.org/10.1021/acs.jproteome.1c00588.

37. Wang, Y., Gao, Z.-M., Li, J., He, L.-S., Cui, G.-J., Li, W.-L., Chen, J., Xin, Y.-Z., Cai, D.-S., and Zhang, A.-Q. (2019). Hadal water sampling by in situ microbial filtration and fixation (ISMIFF) apparatus. Deep Sea Res. I *144*, 132–137. https://doi.org/10.1016/j.dsr.2019.01.009.

38. Saunders, J.K., McIlvin, M.R., Dupont, C.L., Kaul, D., Moran, D.M., Horner, T., Laperriere, S.M., Webb, E.A., Bosak, T., Santoro, A.E., et al. (2022). Microbial functional diversity across biogeochemical provinces in the central Pacific Ocean. Proc. Natl. Acad. Sci. USA *119*, e2200014119. https://doi.org/10.1073/pnas.2200014119.

39. Timmins-Schiffman, E., May, D.H., Mikan, M., Riffle, M., Frazar, C., Harvey, H.R., Noble, W.S., and Nunn, B.L. (2017). Critical decisions in metaproteomics: achieving high confidence protein annotations in a sea of unknowns. ISME J. *11*, 309–314. https://doi.org/10.1038/ismej.2016.132.

40. The UniProt Consortium (2025). UniProt: the Universal Protein Knowledgebase in 2025. Nucleic Acids Res. *53*, D609–D617. https://doi.org/10.1093/nar/gkae1010.

41. Arandia-Gorostidi, N., Parada, A.E., and Dekas, A.E. (2023). Single-cell view of deep-sea microbial activity and intracommunity heterogeneity. ISME J. *17*, 59–69. https://doi.org/10.1038/s41396-022-01324-6.

42. Suzek, B.E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C.H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics *23*, 1282–1288. https://doi.org/10.1093/bioinformatics/btm098.

43. Jiang, L., Tsoucas, D., and Yuan, G.-C. (2018). Assessing Inequality in Transcriptomic Data. Cell Syst. *6*, 149–150. https://doi.org/10.1016/j.cels.2018.02.007.

44. Lauro, F.M., McDougald, D., Thomas, T., Williams, T.J., Egan, S., Rice, S., DeMaere, M.Z., Ting, L., Ertan, H., Johnson, J., et al. (2009). The genomic basis of trophic strategy in marine bacteria. Proc. Natl. Acad. Sci. USA *106*, 15527–15533. https://doi.org/10.1073/pnas.0903507106.

45. Rey-Velasco, X., Auladell, A., Deulofeu-Capo, O., Lundin, D., Pinhassi, J., Ferrera, I., Sánchez, O., and Gasol, J.M. (2025). Decoding the genetic drivers of marine bacterial blooms through comparative genomics. Microbiome *13*, 198. https://doi.org/10.1186/s40168-025-02182-y.

46. Xiao, X., Zhang, Y., and Wang, F. (2021). Hydrostatic pressure is the universal key driver of microbial evolution in the deep ocean and beyond. Environ. Microbiol. Rep. *13*, 68–72. https://doi.org/10.1111/1758-2229.12915.

47. Acinas, S.G., Sánchez, P., Salazar, G., Cornejo-Castillo, F.M., Sebastián, M., Logares, R., Royo-Llonch, M., Paoli, L., Sunagawa, S., Hingamp, P., et al. (2021). Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. Commun. Biol. *4*, 604. https://doi.org/10.1038/s42003-021-02112-2.

48. Lappan, R., Shelley, G., Islam, Z.F., Leung, P.M., Lockwood, S., Nauer, P.A., Jirapanjawat, T., Ni, G., Chen, Y.-J., Kessler, A.J., et al. (2023). Molecular hydrogen in seawater supports growth of diverse marine bacteria. Nat. Microbiol. *8*, 581–595. https://doi.org/10.1038/s41564-023-01322-0.

49. Reji, L., Cardarelli, E.L., Boye, K., Bargar, J.R., and Francis, C.A. (2022). Diverse ecophysiological adaptations of subsurface Thaumarchaeota in floodplain sediments revealed through genome-resolved metagenomics. ISME J. *16*, 1140–1152. https://doi.org/10.1038/s41396-021-01167-7.

50. Paoli, L., Ruscheweyh, H.-J., Forneris, C.C., Hubrich, F., Kautsar, S., Bhushan, A., Lotti, A., Clayssen, Q., Salazar, G., Milanese, A., et al.

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

51. Liang, J., Liu, J., Wang, X., Sun, H., Zhang, Y., Ju, F., Thompson, F., and Zhang, X.-H. (2022). Genomic Analysis Reveals Adaptation of *Vibrio campbellii* to the Hadal Ocean. Appl. Environ. Microbiol. *88*, e0057522. https://doi.org/10.1128/aem.00575-22.
52. Coutinho, F.H., Silveira, C.B., Gregoracci, G.B., Thompson, C.C., Edwards, R.A., Brussaard, C.P.D., Dutilh, B.E., and Thompson, F.L. (2017). Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. Nat. Commun. *8*, 15955. https://doi.org/10.1038/ncomms15955.
53. Silveira, C.B., and Rohwer, F.L. (2016). Piggyback-the-Winner in host-associated microbial communities. npj Biofilms Microbiomes *2*, 16010. https://doi.org/10.1038/npjbiofilms.2016.10.
54. Knowles, B., Silveira, C.B., Bailey, B.A., Barott, K., Cantu, V.A., Cobián-Güemes, A.G., Coutinho, F.H., Dinsdale, E.A., Felts, B., Furby, K.A., et al. (2016). Lytic to temperate switching of viral communities. Nature *531*, 466–470. https://doi.org/10.1038/nature17193.
55. Chen, X., Weinbauer, M.G., Jiao, N., and Zhang, R. (2021). Revisiting marine lytic and lysogenic virus-host interactions: Kill-the-Winner and Piggyback-the-Winner. Sci. Bull. *66*, 871–874. https://doi.org/10.1016/j.scib.2020.12.014.
56. Nies, D.H. (1995). The cobalt, zinc, and cadmium efflux system CzcABC from Alcaligenes eutrophus functions as a cation-proton antiporter in Escherichia coli. J. Bacteriol. *177*, 2707–2712. https://doi.org/10.1128/jb.177.10.2707-2712.1995.
57. Bagai, I., Liu, W., Rensing, C., Blackburn, N.J., and McEvoy, M.M. (2007). Substrate-linked Conformational Change in the Periplasmic Component of a Cu(I)/Ag(I) Efflux System. J. Biol. Chem. *282*, 35695–35702. https://doi.org/10.1074/jbc.M703937200.
58. Wand, M.E., and Sutton, J.M. (2022). Efflux-mediated tolerance to cationic biocides, a cause for concern? Microbiology *168*. https://doi.org/10.1099/mic.0.001263.
59. Checa, S.K., Espariz, M., Audero, M.E.P., Botta, P.E., Spinelli, S.V., and Soncini, F.C. (2007). Bacterial sensing of and resistance to gold salts. Mol. Microbiol. *63*, 1307–1318. https://doi.org/10.1111/j.1365-2958.2007.05590.x.
60. Alvarez, A.H., Moreno-Sánchez, R., and Cervantes, C. (1999). Chromate Efflux by Means of the ChrA Chromate Resistance Protein from *Pseudomonas aeruginosa*. J. Bacteriol. *181*, 7398–7400. https://doi.org/10.1128/JB.181.23.7398-7400.1999.
61. Nies, D.H., Koch, S., Wachi, S., Peitzsch, N., and Saier, M.H. (1998). CHR, a Novel Family of Prokaryotic Proton Motive Force-Driven Transporters Probably Containing Chromate/Sulfate Antiporters. J. Bacteriol. *180*, 5799–5802. https://doi.org/10.1128/JB.180.21.5799-5802.1998.
62. Rowland, J.L., and Niederweis, M. (2013). A Multicopper Oxidase Is Required for Copper Resistance in Mycobacterium tuberculosis. J. Bacteriol. *195*, 3724–3733. https://doi.org/10.1128/JB.00546-13.
63. Brown, N.L., Barrett, S.R., Camakaris, J., Lee, B.T.O., and Rouch, D.A. (1995). Molecular genetics and transport analysis of the copper-resistance determinant (*pco*) from *Escherichia coli* plasmid pRJ1004. Mol. Microbiol. *17*, 1153–1166. https://doi.org/10.1111/j.1365-2958.1995.mmi_17061153.x.
64. Gupta, S.D., Lee, B.T., Camakaris, J., and Wu, H.C. (1995). Identification of cutC and cutF (nlpE) genes involved in copper tolerance in Escherichia coli. J. Bacteriol. *177*, 4207–4215. https://doi.org/10.1128/jb.177.15.4207-4215.1995.
65. Ridge, J.P., Lin, M., Larsen, E.I., Fegan, M., McEwan, A.G., and Sly, L.I. (2007). A multicopper oxidase is essential for manganese oxidation and laccase-like activity in *Pedomicrobium* sp. ACM 3067. Environ. Microbiol. *9*, 944–953. https://doi.org/10.1111/j.1462-2920.2006.01216.x.
66. Pan, Y.-R., Lou, Y.-C., Seven, A.B., Rizo, J., and Chen, C. (2011). NMR Structure and Calcium-Binding Properties of the Tellurite Resistance Protein TerD from Klebsiella pneumoniae. J. Mol. Biol. *405*, 1188–1201. https://doi.org/10.1016/j.jmb.2010.11.041.
67. Diorio, C., Cai, J., Marmor, J., Shinder, R., and DuBow, M.S. (1995). An Escherichia coli chromosomal ars operon homolog is functional in arsenic detoxification and is conserved in gram-negative bacteria. J. Bacteriol. *177*, 2050–2056. https://doi.org/10.1128/jb.177.8.2050-2056.1995.
68. Zhang, X., Xu, W., Liu, Y., Cai, M., Luo, Z., and Li, M. (2018). Metagenomics reveals microbial diversity and metabolic potentials of seawater and surface sediment from a hadal biosphere at the yap trench. Front. Microbiol. *9*, 2402. https://doi.org/10.3389/fmicb.2018.02402.
69. Li, Z., He, Y., Zhang, H., Qian, H., and Wang, Y. (2024). Biotransformations of arsenic in marine sediments across marginal slope to hadal zone. J. Hazard. Mater. *480*, 135955. https://doi.org/10.1016/j.jhazmat.2024.135955.
70. Yuan, C., Duan, G., and Li, F. (2025). Missing pieces in the puzzle of ecology of microbial arsenate reduction. J. Hazard. Mater. *486*, 137054. https://doi.org/10.1016/j.jhazmat.2024.137054.
71. Inoue, J., Shaw, J.P., Rekik, M., and Harayama, S. (1995). Overlapping substrate specificities of benzaldehyde dehydrogenase (the xylC gene product) and 2-hydroxymuconic semialdehyde dehydrogenase (the xylG gene product) encoded by TOL plasmid pWW0 of Pseudomonas putida. J. Bacteriol. *177*, 1196–1201. https://doi.org/10.1128/jb.177.5.1196-1201.1995.
72. Fatayer, S., Coppola, A.I., Schulz, F., Walker, B.D., Broek, T.A., Meyer, G., Druffel, E.R.M., McCarthy, M., and Gross, L. (2018). Direct Visualization of Individual Aromatic Compound Structures in Low Molecular Weight Marine Dissolved Organic Carbon. Geophys. Res. Lett. *45*, 5590–5598. https://doi.org/10.1029/2018GL077457.
73. Liu, Y., Wang, L., Liu, R., and Fang, J. (2024). Biogeochemical cycling of halogenated organic compounds in the ocean: Current progress and future directions. Deep Sea Res. I *205*, 104237. https://doi.org/10.1016/j.dsr.2024.104237.
74. Kaiser, K., and Benner, R. (2008). Major bacterial contribution to the ocean reservoir of detrital organic carbon and nitrogen. Limnol. Oceanogr. *53*, 99–112. https://doi.org/10.4319/lo.2008.53.1.0099.
75. Chen, P., Zhou, H., Huang, Y., Xie, Z., Zhang, M., Wei, Y., Li, J., Ma, Y., Luo, M., Ding, W., et al. (2021). Revealing the full biosphere structure and versatile metabolic functions in the deepest ocean sediment of the Challenger Deep. Genome Biol. *22*, 207. https://doi.org/10.1186/s13059-021-02408-w.
76. Hu, A., Zhao, W., Wang, J., Qi, Q., Xiao, X., and Jing, H. (2024). Microbial communities reveal niche partitioning across the slope and bottom zones of the challenger deep. Environ. Microbiol. Rep. *16*, e13314. https://doi.org/10.1111/1758-2229.13314.
77. Bar-On, Y.M., and Milo, R. (2019). The global mass and average rate of rubisco. Proc. Natl. Acad. Sci. USA *116*, 4738–4743. https://doi.org/10.1073/pnas.1816654116.
78. Yang, N., Lv, Y., Ji, M., Wu, S., and Zhang, Y. (2024). High hydrostatic pressure stimulates microbial nitrate reduction in hadal trench sediments under oxic conditions. Nat. Commun. *15*, 2473. https://doi.org/10.1038/s41467-024-46897-2.
79. Huang, Y., Zhang, X., Xin, Y., Tian, J., and Li, M. (2024). Distinct microbial nitrogen cycling processes in the deepest part of the ocean. mSystems *9*, e0024324. https://doi.org/10.1128/msystems.00243-24.
80. Zhang, Y., Qin, W., Hou, L., Zakem, E.J., Wan, X., Zhao, Z., Liu, L., Hunt, K.A., Jiao, N., Kao, S.-J., et al. (2020). Nitrifier adaptation to low energy flux controls inventory of reduced nitrogen in the dark ocean. Proc. Natl. Acad. Sci. USA *117*, 4823–4830. https://doi.org/10.1073/pnas.1912367117.
81. Ray, W.K., Zeng, G., Potters, M.B., Mansuri, A.M., and Larson, T.J. (2000). Characterization of a 12-kilodalton rhodanese encoded by glpE of Escherichia coli and its interaction with thioredoxin. J. Bacteriol. *182*, 2277–2284. https://doi.org/10.1128/JB.182.8.2277-2284.2000.

82. Xie, Z.-X., He, Y.-B., Zhang, S.-F., Lin, L., Wang, M.-H., and Wang, D.-Z. (2022). Metaexoproteomics Reveals Microbial Behavior in the Ocean's Interior. Front. Microbiol. *13*, 749874. https://doi.org/10.3389/fmicb.2022.749874.

83. Talla, E., Hedrich, S., Mangenot, S., Ji, B., Johnson, D.B., Barbe, V., and Bonnefoy, V. (2014). Insights into the pathways of iron- and sulfur-oxidation, and biofilm formation from the chemolithotrophic acidophile Acidithiobacillus ferrivorans CF27. Res. Microbiol. *165*, 753–760. https://doi.org/10.1016/j.resmic.2014.08.002.

84. Bai, Y., Chen, T., Happe, T., Lu, Y., and Sawyer, A. (2018). Iron–sulphur cluster biogenesis *via* the SUF pathway. Metallomics *10*, 1038–1052. https://doi.org/10.1039/C8MT00150B.

85. Yin, Q.-J., Zhang, W.-J., Qi, X.-Q., Zhang, S.-D., Jiang, T., Li, X.-G., Chen, Y., Santini, C.-L., Zhou, H., Chou, I.-M., et al. (2018). High Hydrostatic Pressure Inducible Trimethylamine N-Oxide Reductase Improves the Pressure Tolerance of Piezosensitive Bacteria Vibrio fluvialis. Front. Microbiol. *8*, 2646. https://doi.org/10.3389/fmicb.2017.02646.

86. Zheng, Y., Wang, J., Zhou, S., Zhang, Y., Liu, J., Xue, C.-X., Williams, B.T., Zhao, X., Zhao, L., Zhu, X.-Y., et al. (2020). Bacteria are important dimethylsulfoniopropionate producers in marine aphotic and high-pressure environments. Nat. Commun. *11*, 4658. https://doi.org/10.1038/s41467-020-18434-4.

87. Shaw, D.K., Sekar, J., and Ramalingam, P.V. (2022). Recent insights into oceanic dimethylsulfoniopropionate biosynthesis and catabolism. Environ. Microbiol. *24*, 2669–2700. https://doi.org/10.1111/1462-2920.16045.

88. Williams, T.J., Long, E., Evans, F., DeMaere, M.Z., Lauro, F.M., Raftery, M.J., Ducklow, H., Grzymski, J.J., Murray, A.E., and Cavicchioli, R. (2012). A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. ISME J. *6*, 1883–1900. https://doi.org/10.1038/ismej.2012.28.

89. Xie, Z.-X., He, Y.-B., Wang, M.-H., Zhang, S.-F., Kong, L.-F., Lin, L., Liu, S.-Q., and Wang, D.-Z. (2020). Dissecting microbial community structure and metabolic activities at an oceanic deep chlorophyll maximum layer by size-fractionated metaproteomics. Prog. Oceanogr. *188*, 102439. https://doi.org/10.1016/j.pocean.2020.102439.

90. Fuchsman, C.A., Palevsky, H.I., Widner, B., Duffy, M., Carlson, M.C.G., Neibauer, J.A., Mulholland, M.R., Keil, R.G., Devol, A.H., and Rocap, G. (2019). Cyanobacteria and cyanophage contributions to carbon and nitrogen cycling in an oligotrophic oxygen-deficient zone. ISME J. *13*, 2714–2726. https://doi.org/10.1038/s41396-019-0452-6.

91. Zhao, Z., Amano, C., Reinthaler, T., Baltar, F., Orellana, M.V., and Herndl, G.J. (2024). Metaproteomic analysis decodes trophic interactions of microorganisms in the dark ocean. Nat. Commun. *15*, 6411. https://doi.org/10.1038/s41467-024-50867-z.

92. Yu, H., and Leadbetter, J.R. (2020). Bacterial chemolithoautotrophy via manganese oxidation. Nature *583*, 453–458. https://doi.org/10.1038/s41586-020-2468-5.

93. He, Y., Zeng, X., Xu, F., and Shao, Z. (2022). Diversity of Mixotrophic Neutrophilic Thiosulfate- and Iron-Oxidizing Bacteria from Deep-Sea Hydrothermal Vents. Microorganisms *11*, 100. https://doi.org/10.3390/microorganisms11010100.

94. Jing, H., Xiao, X., Zhang, Y., Li, Z., Jian, H., Luo, Y., and Han, Z. (2022). Composition and Ecological Roles of the Core Microbiome along the Abyssal-Hadal Transition Zone Sediments of the Mariana Trench. Microbiol. Spectr. *10*, e0198821. https://doi.org/10.1128/spectrum.01988-21.

95. Qin, Q.-L., Wang, Z.-B., Su, H.-N., Chen, X.-L., Miao, J., Wang, X.-J., Li, C.-Y., Zhang, X.-Y., Li, P.-Y., Wang, M., et al. (2021). Oxidation of trimethylamine to trimethylamine *N*-oxide facilitates high hydrostatic pressure tolerance in a generalist bacterial lineage. Sci. Adv. *7*, eabf9941. https://doi.org/10.1126/sciadv.abf9941.

96. Ismail, K.A., and Al-Shehhi, M.R. (2023). Applications of biogeochemical models in different marine environments: a review. Front. Environ. Sci. *11*, 1198856. https://doi.org/10.3389/fenvs.2023.1198856.

97. Tamegai, H., Li, L., Masui, N., and Kato, C. (1997). A denitrifying bacterium from the deep sea at 11,000-m depth. Extremophiles *1*, 207–211. https://doi.org/10.1007/s007920050035.

98. Kobayashi, H., Takaki, Y., Kobata, K., Takami, H., and Inoue, A. (1998). Characterization of alpha-maltotetraohydrolase produced by Pseudomonas sp. MS300 isolated from the deepest site of the mariana trench. Extremophiles *2*, 401–407. https://doi.org/10.1007/s007920050085.

99. Ahmad, W., Zheng, Y., Li, Y., Sun, W., Hu, Y., He, X., Liu, R., Xue, C.-X., and Zhang, X.-H. (2020). Marinobacter salinexigens sp. nov., a marine bacterium isolated from hadal seawater of the Mariana Trench. Int. J. Syst. Evol. Microbiol. *70*, 3794–3800. https://doi.org/10.1099/ijsem.0.004236.

100. Liu, P., Ding, W., Lai, Q., Liu, R., Wei, Y., Wang, L., Xie, Z., Cao, J., and Fang, J. (2020). Physiological and genomic features of Paraoceanicella profunda gen. nov., sp. nov., a novel piezophile isolated from deep seawater of the Mariana Trench. MicrobiologyOpen *9*, e966. https://doi.org/10.1002/mbo3.966.

101. Wang, Z., Du, R.-B., Sun, Q.-L., Sun, Y.-Y., Zhang, J., and Tang, Y.-Z. (2021). Pontivivens ytuae sp. nov., isolated from deep sea sediment of the Mariana Trench. Int. J. Syst. Evol. Microbiol. *71*. https://doi.org/10.1099/ijsem.0.004967.

102. Meng, F.-X., Li, G., Fang, C., Wu, Y.-H., Cheng, H., Chen, Y., Wang, C.-S., Shao, Z., and Xu, X.-W. (2019). Altererythrobacter aerophilus sp. nov., isolated from deep-sea water of the north-west Pacific. Int. J. Syst. Evol. Microbiol. *69*, 1689–1695. https://doi.org/10.1099/ijsem.0.003377.

103. Yang, S., Li, X., Xiao, X., Zhuang, G., and Zhang, Y. (2020). Sphingomonas profundi sp. nov., isolated from deep-sea sediment of the Mariana Trench. Int. J. Syst. Evol. Microbiol. *70*, 3809–3815. https://doi.org/10.1099/ijsem.0.004235.

104. Yang, Y., Zhang, C., Lenton, T.M., Yan, X., Zhu, M., Zhou, M., Tao, J., Phelps, T.J., and Cao, Z. (2021). The Evolution Pathway of Ammonia-Oxidizing Archaea Shaped by Major Geological Events. Mol. Biol. Evol. *38*, 3637–3648. https://doi.org/10.1093/molbev/msab129.

105. Perfilyev, A., Gæde, A., Hooton, S., Zahran, S.A., Kalatzis, P.G., Winther-Have, C.S., Ibarra Chavez, R., Wilkinson, R.C., Thanki, A.M., Liu, Z., et al. (2025). Large-scale analysis of bacterial genomes reveals thousands of lytic phages. Nat. Microbiol. *11*, 42–52. https://doi.org/10.1038/s41564-025-02203-4.

106. Dougherty, P.E., Bernard, C., Carstens, A.B., Bumunang, E., Gerovac, M., Müsken, M., Stanford, K., McAllister, T.A., Rocha, E.P.C., and Hansen, L.H. (2025). Persistent virulent phages exist across bacterial isolates. Nat. Microbiol. *11*, 31–41. https://doi.org/10.1038/s41564-025-02207-0.

107. Silveira, C.B., Luque, A., and Rohwer, F. (2021). The landscape of lysogeny across microbial community density, diversity and energetics. Environ. Microbiol. *23*, 4098–4111. https://doi.org/10.1111/1462-2920.15640.

108. Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., and Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. *50*, D785–D794. https://doi.org/10.1093/nar/gkab776.

109. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Res. *49*, D412–D419. https://doi.org/10.1093/nar/gkaa913.

110. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y., and Ishiguro-Watanabe, M. (2025). KEGG: biological systems database as a model of the real world. Nucleic Acids Res. *53*, D672–D677. https://doi.org/10.1093/nar/gkae909.

111. Menzel, P., Ng, K.L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat. Commun. *7*, 11257. https://doi.org/10.1038/ncomms11257.

112. Galperin, M.Y., Wolf, Y.I., Makarova, K.S., Vera Alvarez, R., Landsman, D., and Koonin, E.V. (2021). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res. *49*, D274–D281. https://doi.org/10.1093/nar/gkaa1018.

113. Trgovec-Greif, L., Hellinger, H.-J., Mainguy, J., Pfundner, A., Frishman, D., Kiening, M., Webster, N.S., Laffy, P.W., Feichtinger, M., and Rattei, T. (2024). VOGDB—Database of Virus Orthologous Groups. Viruses *16*, 1191. https://doi.org/10.3390/v16081191.

114. Terzian, P., Olo Ndela, E., Galiez, C., Lossouarn, J., Pérez Bucio, R.E., Mom, R., Toussaint, A., Petit, M.-A., and Enault, F. (2021). PHROG: families of prokaryotic virus proteins clustered using remote homology. NAR Genom. Bioinform. *3*, lqab067. https://doi.org/10.1093/nargab/lqab067.

115. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

116. Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics *31*, 1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

117. Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner (Lawrence Berkeley National Laboratory).

118. Kang, D.D., Li, F., Kirton, E.S., Thomas, A., Egan, R.S., An, H., and Wang, Z. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ *7*, e7359. https://doi.org/10.7717/peerj.7359.

119. Wu, Y.-W., Simmons, B.A., and Singer, S.W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics *32*, 605–607. https://doi.org/10.1093/bioinformatics/btv638.

120. Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. Nat. Methods *11*, 1144–1146. https://doi.org/10.1038/nmeth.3103.

121. Nissen, J.N., Johansen, J., Allesøe, R.L., Sønderby, C.K., Armenteros, J.J.A., Grønbech, C.H., Jensen, L.J., Nielsen, H.B., Petersen, T.N., Winther, O., et al. (2021). Improved metagenome binning and assembly using deep variational autoencoders. Nat. Biotechnol. *39*, 555–560. https://doi.org/10.1038/s41587-020-00777-4.

122. Liu, C.-C., Dong, S.-S., Chen, J.-B., Wang, C., Ning, P., Guo, Y., and Yang, T.-L. (2022). MetaDecoder: a novel method for clustering metagenomic contigs. Microbiome *10*, 46. https://doi.org/10.1186/s40168-022-01237-8.

123. Parks, D.H. (2022). UniteM (GitHub). https://github.com/dparks1134/UniteM.

124. Orakov, A., Fullam, A., Coelho, L.P., Khedkar, S., Szklarczyk, D., Mende, D.R., Schmidt, T.S.B., and Bork, P. (2021). GUNC: detection of chimerism and contamination in prokaryotic genomes. Genome Biol. *22*, 178. https://doi.org/10.1186/s13059-021-02393-0.

125. Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. *11*, 2864–2868. https://doi.org/10.1038/ismej.2017.126.

126. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics *38*, 5315–5316. https://doi.org/10.1093/bioinformatics/btac672.

127. Barnum, T.P., Crits-Christoph, A., Molla, M., Carini, P., Lee, H.H., and Ostrov, N. (2024). Predicting microbial growth conditions from amino acid composition. Preprint at bioRxiv. https://doi.org/10.1101/2024.03.22.586313.

128. Weissman, J.L., Hou, S., and Fuhrman, J.A. (2021). Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. Proc. Natl. Acad. Sci. USA *118*, e2016810118. https://doi.org/10.1073/pnas.2016810118.

129. Larralde, M. (2023). pyrodigal-gv (GitHub). https://github.com/althonos/pyrodigal-gv.

130. Queirós, P., Delogu, F., Hickl, O., May, P., and Wilmes, P. (2021). Mantis: flexible and consensus-driven genome annotation. GigaScience *10*, giab042. https://doi.org/10.1093/gigascience/giab042.

131. Larralde, M. (2022). Pyrodigal: Python bindings and interface to Prodigal, an efficient method for gene prediction in prokaryotes. JOSS *7*, 4296. https://doi.org/10.21105/joss.04296.

132. Espinoza, J.L. (2025). PyHMMSearch (GitHub). https://github.com/jolespin/pyhmmsearch.

133. Figueroa, J.L., III, Dhungel, E., Bellanger, M., Brouwer, C.R., and White, R.A., III (2024). MetaCerberus: distributed highly parallelized HMM-based processing for robust functional annotation across the tree of life. Bioinformatics *40*, btae119. https://doi.org/10.1093/bioinformatics/btae119.

134. Asnicar, F., Thomas, A.M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., Zhu, Q., Bolzan, M., Cumbo, F., May, U., et al. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat. Commun. *11*, 2500. https://doi.org/10.1038/s41467-020-16366-7.

135. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

136. Yu, F., Haynes, S.E., Teo, G.C., Avtonomov, D.M., Polasky, D.A., and Nesvizhskii, A.I. (2020). Fast Quantitative Analysis of timsTOF PASEF Data with MSFragger and IonQuant. Mol. Cell. Proteomics *19*, 1575–1585. https://doi.org/10.1074/mcp.TIR120.002048.

137. Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. Nat. Methods *14*, 513–520. https://doi.org/10.1038/nmeth.4256.

138. da Veiga Leprevost, F., Haynes, S.E., Avtonomov, D.M., Chang, H.Y., Shanmugam, A.K., Mellacheruvu, D., Kong, A.T., and Nesvizhskii, A.I. (2020). Philosopher: a versatile toolkit for shotgun proteomics data analysis. Nat. Methods *17*, 869–870. https://doi.org/10.1038/s41592-020-0912-y.

139. Yang, K.L., Yu, F., Teo, G.C., Li, K., Demichev, V., Ralser, M., and Nesvizhskii, A.I. (2023). MSBooster: improving peptide identification rates using deep learning-based features. Nat. Commun. *14*, 4539. https://doi.org/10.1038/s41467-023-40129-9.

140. The, M., MacCoss, M.J., Noble, W.S., and Käll, L. (2016). Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. J. Am. Soc. Mass Spectrom. *27*, 1719–1727. https://doi.org/10.1007/s13361-016-1460-7.

141. Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. *75*, 4646–4658. https://doi.org/10.1021/ac0341261.

142. Yu, F., Haynes, S.E., and Nesvizhskii, A.I. (2021). IonQuant enables accurate and sensitive label-free quantification with FDR-controlled match-between-runs. Mol. Cell. Proteomics *20*, 100077. https://doi.org/10.1016/J.MCPRO.2021.100077.

143. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

144. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

145. Aroney, S.T.N., Newell, R.J.P., Nissen, J.N., Camargo, A.P., Tyson, G.W., and Woodcroft, B.J. (2025). CoverM: read alignment statistics for metagenomics. Bioinformatics *41*, btaf147. https://doi.org/10.1093/bioinformatics/btaf147.

146. Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. *35*, 1026–1028. https://doi.org/10.1038/nbt.3988.

147. Jombart, T., Balloux, F., and Dray, S. (2010). *adephylo*: new tools for investigating the phylogenetic signal in biological traits. Bioinformatics *26*, 1907–1909. https://doi.org/10.1093/bioinformatics/btq292.

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

149. Camargo, A.P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P.S.G., Nayfach, S., and Kyrpides, N.C. (2024). Identification of mobile genetic elements with geNomad. Nat. Biotechnol. *42*, 1303–1312. https://doi.org/10.1038/s41587-023-01953-y.

150. Nayfach, S., Camargo, A.P., Schulz, F., Eloe-Fadrosh, E., Roux, S., and Kyrpides, N.C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat. Biotechnol. *39*, 578–585. https://doi.org/10.1038/s41587-020-00774-7.

151. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinform. *10*, 421. https://doi.org/10.1186/1471-2105-10-421.

152. Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal. Complex Syst. *1695*, 1–9.

153. Camargo, A.P. (2022). bioinformatics-snakemake-pipelines (GitHub). https://github.com/apcamargo/bioinformatics-snakemake-pipelines/tree/main/contig-ani-leiden-clustering-pipeline.

154. Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLoS Comput. Biol. *7*, e1002195. https://doi.org/10.1371/journal.pcbi.1002195.

155. Kieft, K., Zhou, Z., and Anantharaman, K. (2020). VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome *8*, 90. https://doi.org/10.1186/s40168-020-00867-0.

156. Shang, J., Tang, X., and Sun, Y. (2023). PhaTYP: predicting the lifestyle for bacteriophages using BERT. Brief. Bioinform. *24*, bbac487. https://doi.org/10.1093/bib/bbac487.

157. Shang, J., Peng, C., Liao, H., Tang, X., and Sun, Y. (2023). PhaBOX: a web server for identifying and characterizing phage contigs in metagenomic data. Bioinform. Adv. *3*, vbad101. https://doi.org/10.1093/bioadv/vbad101.

158. Ligeti, B., Szepesi-Nagy, I., Bodnár, B., Ligeti-Nagy, N., and Juhász, J. (2024). ProkBERT family: genomic language models for microbiome applications. Front. Microbiol. *14*, 1331233. https://doi.org/10.3389/fmicb.2023.1331233.

159. Roux, S., Camargo, A.P., Coutinho, F.H., Dabdoub, S.M., Dutilh, B.E., Nayfach, S., and Tritt, A. (2023). iPHoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. PLoS Biol. *21*, e3002083. https://doi.org/10.1371/journal.pbio.3002083.

160. Chan, P.P., Lin, B.Y., Mak, A.J., and Lowe, T.M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. *49*, 9077–9096. https://doi.org/10.1093/nar/gkab688.

161. Shang, J., and Sun, Y. (2022). CHERRY: a Computational metHod for accuratE pRediction of virus–pRokarYotic interactions using a graph encoder–decoder model. Brief. Bioinform. *23*, bbac182. https://doi.org/10.1093/bib/bbac182.

162. Bouras, G., Nepal, R., Houtak, G., Psaltis, A.J., Wormald, P.-J., and Vreugde, S. (2023). Pharokka: a fast scalable bacteriophage annotation tool. Bioinformatics *39*, btac776. https://doi.org/10.1093/bioinformatics/btac776.

163. Kosmopoulos, J.C., Martin, C., and Anantharaman, K. (2025). CheckAMG (GitHub). https://github.com/AnantharamanLab/CheckAMG.

164. R Core Team (2025). R: a Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

165. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS A J. Integr. Biol. *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

166. Wickham, H. (2016). ggplot2, Second Edition (Springer). https://doi.org/10.1007/978-3-319-24277-4.

167. Kassambara, A. (2025) Ggpubr: "ggplot2" Based Publication Ready Plots. https://CRAN.R-project.org/package=ggpubr.

168. Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., et al. (2025) Vegan: Community Ecology Package. http://CRAN.R-project.org/package=vegan.

169. Zeileis, A. (2014) Ineq: Measuring Inequality, Concentration, and Poverty. 10.32614/CRAN.package.ineq.

170. Sun, X., Yang, Q., and Xia, X. (2013). An Improved Implementation of Effective Number of Codons (Nc). Mol. Biol. Evol. *30*, 191–196. https://doi.org/10.1093/molbev/mss201.

171. Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. *11*, 119. https://doi.org/10.1186/1471-2105-11-119.

172. Paysan-Lafosse, T., Andreeva, A., Blum, M., Chuguransky, S.R., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Llinares-López, F., Meng-Papaxanthos, L., et al. (2025). The Pfam protein families database: embracing AI/ML. Nucleic Acids Res. *53*, D523–D534. https://doi.org/10.1093/nar/gkae997.

173. Larralde, M., and Zeller, G. (2023). PyHMMER: a Python library binding to HMMER for efficient sequence analysis. Bioinformatics *39*, btad214. https://doi.org/10.1093/bioinformatics/btad214.

174. Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.-C., and Yates, J.R. (2013). Protein analysis by shotgun/bottom-up proteomics. Chem. Rev. *113*, 2343–2394. https://doi.org/10.1021/cr3003533.

175. Kong, Y., Zhang, R., Blain, S., and Obernosterer, I. (2024). Seasonal dynamics in microbial trace metals transporters during phytoplankton blooms in the Southern Ocean. Environ. Microbiol. *26*, e16695. https://doi.org/10.1111/1462-2920.16695.

176. Deutsch, E.W., Bandeira, N., Perez-Riverol, Y., Sharma, V., Carver, J.J., Mendoza, L., Kundu, D.J., Wang, S., Bandla, C., Kamatchinathan, S., et al. (2023). The ProteomeXchange consortium at 10 years: 2023 update. Nucleic Acids Res. *51*, D1539–D1548. https://doi.org/10.1093/nar/gkac1040.

177. Suzek, B.E., Wang, Y., Huang, H., McGarvey, P.B., and Wu, C.H.; The UniProt Consortium (2015). UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics *31*, 926–932. https://doi.org/10.1093/bioinformatics/btu739.

178. Roux, S., Adriaenssens, E.M., Dutilh, B.E., Koonin, E.V., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Lavigne, R., Brister, J.R., Varsani, A., et al. (2019). Minimum Information about an Uncultivated Virus Genome (MIUViG). Nat. Biotechnol. *37*, 29–37. https://doi.org/10.1038/nbt.4306.

179. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233. https://doi.org/10.1038/s41598-019-41695-z.

180. Luo, X.-Q., Wang, P., Li, J.-L., Ahmad, M., Duan, L., Yin, L.-Z., Deng, Q.-Q., Fang, B.-Z., Li, S.-H., and Li, W.-J. (2022). Viral community-wide auxiliary metabolic genes differ by lifestyles, habitats, and hosts. Microbiome *10*, 190. https://doi.org/10.1186/s40168-022-01384-y.

181. Cook, R., Hooton, S., Trivedi, U., King, L., Dodd, C.E.R., Hobman, J.L., Stekel, D.J., Jones, M.A., and Millard, A.D. (2021). Hybrid assembly of an agricultural slurry virome reveals a diverse and stable community with the potential to alter the metabolism and virulence of veterinary pathogens. Microbiome *9*, 65. https://doi.org/10.1186/s40168-021-01010-3.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Filtered hadal seawater samples	Collected from the Mariana Trench, the Yap Trench and the Philippine Basin	MEER
Chemicals, peptides, and recombinant proteins		
1 M Tris-HCl, pH 8.0	Thermo Fisher Scientific	cat#15568025
10% SDS solution	Invitrogen	cat#15553027
Chloroform-isoamyl alcohol, 24:1	Sigma-Aldrich	cat#25666
DNA Clean Beads	Vazyme	cat#N411-02
Ethanol	Sangon Biotech	cat#A500737-0500
Methanol	Macklin	cat#M813904-4L
Nuclease-free water	Invitrogen	cat#AM9937
Protease &Phosohatase Inhibitor single-Use Cocktail	Thermo Scientific	cat#1861280
RNase A	TransGen	cat#GE101-01
RNase-free Glycogen	ZOMANBIO	cat#ZP423-2
Tris-saturated phenol, pH 7.7-8.1	Macklin	cat#T917981-200ml
Trypsin/Lys-C	Promega	cat#V5073
Ammonium acetate	Macklin	cat#A801001-500g
Dithiothreitol (DTT)	Aladdin	cat#D104859-25g
EDTA-Na2	Solarbio	cat#E8030
Hexadecyl trimethyl ammonium bromide (CTAB)	Macklin	cat#H811117-100g
Polyethylene glycol 6000(PEG6000)	Macklin	cat#P815609-500g
PVP	Macklin	cat#P766521-100g
Sucrose	Macklin	cat#S6212-500g
Sodium chloride (NaCl)	Macklin	cat#S805275-500g
Sodium phosphate dibasic (Na_2HPO_4)	Macklin	cat#S818100-500g
Sodium phosphate monobasic (NaH_2PO_4)	Macklin	cat#S817780-500g
Urea	Macklin	cat#U6209-500g
Iodoacetamide	Sigma-Aldrich	cat#I1149-5G
Ammonium bicarbonate	Macklin	cat#A800859-500g
Critical commercial assays		
MGIEasy Fast FS DNA Library Prep Set	MGI-Tech, China	cat#940-000030-00
DNBSEQ-T10 x 4RS (FCL PE150)	MGI-Tech, China	cat#940-000100-00
Deposited data		
Raw data	This paper	NODE: OEX00025177
Raw data	This paper	iProX: PXD045060
Source code for metagenome-guided metaproteomic analysis	This paper	Zenodo: 10.5281/zenodo.20785637
Mariana Trench Environment and Ecology Research (MEER) project dataset	Xiao et al.[1]	https://www.biosino.org/mash/meer/home
Ocean Microbiomics Database	Paoli et al.[50]	https://microbiomics.io/ocean/
GTDB database release 220	Parks et al.[108]	https://gtdb.ecogenomic.org/
Pfam database	Mistry et al.[109]	http://pfam.xfam.org/clan/browse
KEGG database	Kanehisa et al.[110]	https://www.kegg.jp

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Kaiju database (kaiju_db_nr_euk_2022-03-10)	Menzel et al.[111]	https://bioinformatics-centre.github.io/kaiju/
COG database	Galperin et al.[112]	https://www.ncbi.nlm.nih.gov/research/COG
VOG database	Trgovec-Greif et al.[113]	https://vogdb.org
PHROG database	Terzian et al.[114]	https://phrogs.lmge.uca.fr
Public upper ocean metaproteomic dataset	Giljan et al.[26]	PRIDE: PXD016702
Public upper ocean metaproteomic dataset	Brum et al.[28]	PRIDE: PXD000938
Public upper ocean metaproteomic dataset	Hawley et al.[29]	PRIDE: PXD004433
Public upper ocean metaproteomic dataset	McCain et al.[35]	PRIDE: PXD022995
Public upper ocean metaproteomic dataset	McIlvin et al.[36]	PRIDE: PXD027351
Heavy metal concentrations in the upper ocean water (0-1,000 m)	GEOTRACES database	Table S7
Heavy metal concentrations in the upper ocean water (0-1,000 m)	MBARI database	Table S7
Software and algorithms		
fastp (v0.23.2)	Chen et al.[115]	https://github.com/opengene/fastp
MEGAHIT (v1.2.9)	Li et al.[116]	https://github.com/voutcn/megahit
bbmap.sh (v39.01)	Bushnell[117]	https://sourceforge.net/projects/bbmap/
MetaBAT2 (v2.15)	Kang et al.[118]	https://bitbucket.org/berkeleylab/metabat
MaxBin2 (v2.2.7)	Wu et al.[119]	https://sourceforge.net/projects/maxbin2/
CONCOCT (v1.1.0)	Alneberg et al.[120]	https://github.com/binpro/concoct
VAMB (v3.0.2)	Nissen et al.[121]	https://github.com/RasmussenLab/vamb
MetaDecoder (v1.0.14)	Liu et al.[122]	https://github.com/liu-congcong/MetaDecoder
UniteM (v1.2.4)	Parks[123]	https://github.com/dparks1134/UniteM
GUNC (v1.0.5)	Orakov et al.[124]	https://github.com/grp-bork/gunc
dRep (v3.4.0)	Olm et al.[125]	https://github.com/MrOlm/drep
GTDB-Tk (v2.4.0)	Chaumeil et al.[126]	https://github.com/Ecogenomics/GTDBTk
GenomeSPOT (v1.0.1)	Barnum et al.[127]	https://github.com/cultivarium/genomespot
gRodon (v2.5.1)	Weissman et al.[128]	https://github.com/jlw-ecoevo/gRodon2
genome package	This paper	https://github.com/Hocnonsense/genome
pyrodigal-gv (v0.3.1)	Larralde[129]	https://github.com/althonos/pyrodigal-gv
MANTIS (v1.5.5)	Queirós et al.[130]	https://github.com/PedroMTQ/mantis
kaiju (v1.9.2)	Menzel et al.[111]	https://github.com/bioinformatics-centre/kaiju
pyrodigal (v3.3.0)	Larralde[131]	https://github.com/althonos/pyrodigal
pyhmmsearch (v2025.1.23)	Espinoza[132]	https://github.com/jolespin/pyhmmsearch
MetaCerberus (v1.4.0)	Figueroa et al.[133]	https://github.com/raw-lab/metacerberus
Phylophlan (v3.0.60)	Asnicar er al.[134]	https://github.com/biobakery/phylophlan
IQ-TREE (v2.3.0)	Minh et al.[135]	https://github.com/iqtree/iqtree2
FragPipe (v18.0)	Yu et al.[136]	https://github.com/Nesvilab/FragPipe
MSFragger (v3.5)	Kong et al.[137]	https://github.com/Nesvilab/MSFragger
Philosopher (v4.4.0)	da Veiga Leprevost et al.[138]	https://github.com/Nesvilab/philosopher
MSBooster (v1.1.4)	Yang et al.[139]	https://github.com/Nesvilab/MSBooster
Percolator (v3.05)	The et al.[140]	https://github.com/percolator/percolator
ProteinProphet (v4.4.0)	Nesvizhskii et al.[141]	https://proteinprophet.sourceforge.net
IonQuant (v1.8.0)	Yu et al.[142]	https://github.com/Nesvilab/IonQuant
featureCounts (v2.0.6)	Liao et al.[143]	https://subread.sourceforge.net/featureCounts.html
bwa (v0.7.17-r1188)	Li et al.[144]	https://github.com/lh3/bwa
CoverM (v0.6.1)	Aroney et al.[145]	https://github.com/wwood/CoverM
MMseqs (v15.6f452)	Steinegger et al.[146]	https://github.com/soedinglab/MMseqs2
adephylo (v1.1.17)	Jombart et al.[147]	https://github.com/adeverse/adephylo
phytools (v2.5.2)	Revell[148]	https://github.com/liamrevell/phytools

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
geNomad (v1.8.0)	Camargo et al.[149]	https://github.com/apcamargo/genomad
CheckV (v1.0.1)	Nayfach et al.[150]	https://bitbucket.org/berkeleylab/checkv/src/master/
BLAST+ (v2.16.0)	Camacho et al.[151]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
igraph (v0.11.8)	Csardi et al.[152]	https://igraph.org
Custom viral clustering pipeline	Camargo[153]	https://github.com/apcamargo/bioinformatics-snakemake-pipelines/tree/main/contig-ani-leiden-clustering-pipeline
hmmsearch (v3.4)	Eddy[154]	http://hmmer.org
VIBRANT (v1.2.1)	Kieft et al.[155]	https://github.com/AnantharamanLab/VIBRANT
PhaTYP	Shang et al.[156]	https://github.com/KennthShang/PhaBOX
PhaBOX2 (v2.1.9)	Shang et al.[157]	https://github.com/KennthShang/PhaBOX
ProkBERT PhaStyle (v0.1)	Ligeti et al.[158]	https://github.com/nbrg-ppcu/PhaStyle
iPHoP (v1.3.3)	Roux et al.[159]	https://bitbucket.org/srouxjgi/iphop/src/main/
tRNAscan-SE (v2.0.12)	Chan et al.[160]	https://github.com/UCSC-LoweLab/tRNAscan-SE
CHERRY	Shang et al.[161]	https://github.com/KennthShang/PhaBOX
PHAROKKA (v1.9.0)	Bouras et al.[162]	https://github.com/gbouras13/pharokka
CheckAMG	Kosmopoulos et al.[163]	https://github.com/AnantharamanLab/CheckAMG
R (v4.4)	R Core Team[164]	https://www.r-project.org
stats (v 4.4.3)	R Core Team[164]	https://www.r-project.org
clusterProfiler (v 4.14.6)	Yu et al.[165]	https://github.com/YuLab-SMU/clusterProfiler
ggplot2 (v 4.0.1)	Wickham[166]	https://ggplot2.tidyverse.org
ggpubr (v0.6.2)	Kassambara[167]	https://rpkgs.datanovia.com/ggpubr/
vegan (v2.7.2)	Oksanen et al.[168]	https://vegandevs.github.io/vegan/
ineq (v0.2.13)	Zeileis et al.[169]	https://CRAN.R-project.org/package=ineq

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

No experimental models nor human participants were included in this study. Samples used in this study were directly collected from natural oceanic environments using the HOV *Fendouzhe* and equipment onboard. No culturing of microbe strains, cell lines, nor primary cell cultures has been conducted.

METHOD DETAILS

Sample collection and processing

The ISMIFF was operated according to the manufacturer's instructions. In general, all tubes were thoroughly washed and prefilled with Milli-Q water before the PVDF membrane (142 mm, 0.22 μm, MilliporeSigma Company) was placed into the filtration cassette. The ISMIFF was mounted onto the rear of the human-occupied vehicle (HOV), which was approximately 1 m above the sea bed. Filtration programs were set to initiate one hour after the submersible was supposed to reach the bottom and terminate one hour before the end of bottom operation or if the volume of filtered seawater reached 200 L before the set time. Once the submersible was back on the deck, the membrane was removed from the filtration cassette with a sterile tweezer, transferred into a sterile sample bag and conserved at -80°C. In total, 12 *in situ* filtered samples were collected by the ISMIFF from October to November 2021, including seven samples from the Mariana Trench, one from the Yap Trench and four from the West Philippine Basin at the water depths ranging from 6-11 km b.s.l. Among the samples from the Mariana Trench, two were collected along the bottom axis of the trench (FDZ061 and FDZ077), two were collected on the northern slope (FDZ071 and FDZ076) and the remaining three were collected on the southern slope (FDZ062, FDZ072, and FDZ075). The sampling sites and submersible trajectories during ISMIFF operation are shown in Table S1.

Heavy metal concentration measurements

All seawater samples were diluted 15-fold with 2% nitric acid prior to analysis. Concentrations of arsenic (As), selenium (Se), copper (Cu), manganese (Mn), and zinc (Zn) were determined using inductively coupled plasma mass spectrometry (ICP-MS; NexION2000G, PerkinElmer, CA, USA), applying the standard addition method with rhodium (Rh) as an internal standard for matrix correction. Total mercury (Hg) concentrations were measured using an atomic fluorescence spectrometry speciation analyzer (AFS; SA-10, Beijing Ji-Tian Instrument Co., Ltd., Beijing, China). Samples were introduced into the reaction system using 2% (v/v) HNO_3 as the carrier

solution. Mercury was reduced to elemental vapor (Hg^0) using a freshly prepared reducing solution (10 g/L KBH_4 + 5 g/L KOH) prior to detection. The generated Hg^0 vapor was transported by argon gas (99.999%) into the fluorescence cell, where mercury atoms were excited by a hollow cathode lamp (λ = 254 nm). The emitted fluorescence was detected using a photomultiplier, and signal intensity was quantified against external calibration standards. Upper ocean reference data were obtained from publicly available databases covering the 0–1,000 m water column. Specifically, Cu, Mn, and Zn concentrations were retrieved from the GEOTRACES database (https://www.geotraces.org/), while As, Hg, and Se data were obtained from the MBARI database (https://www.mbari.org/).

Coextraction of DNA and protein

The filter membrane with retained microorganisms was put into a ceramic mortar, liquid nitrogen was added, and the membrane was ground to powder. 1/2 of the filter membrane powder, 1 g quartz sand (~1 mm diameter), 1 mL sodium dodecyl sulfate (SDS) buffer (1% SDS, 240 mM sodium phosphate buffer (pH 8.0), 20 mM ethylenediaminetetraacetic acid (EDTA), 40mM dithiothreitol (DTT), 60% sucrose), 3 mL phenol (Tris saturated phenol, pH 7.7-8.1, containing 0.1% 8-hydroxyquinoline), 0.06g dithiothreitol (DTT), and 30 μL protease inhibitor (Halt™ Protease and Phosphatase Inhibitor Single-Use Cocktail, Thermo Scientific) were added into 15 mL sterile polypropylene centrifuge tubes. Then, the samples were homogenized (4 m/s, 20 s) three times (3 min intervals) with a homogenizer (FastPrep-24, MP Biomedical). After grinding, the tubes were centrifuged (10,000 ×g, 5 min, 4 °C).

The upper organic phase was transferred to a 15 mL centrifuge tube, mixed by inversion with five times the volume of methanol-ammonium acetate (0.1 M ammonium acetate dissolved in methanol), and centrifuged (10,000 ×g, 10 min) after 16 hours of incubation at -20 °C. After suspension in 1 mL anhydrous ethanol, the precipitate was transferred to a 1.5 mL centrifuge tube and centrifuged (14,000 ×g, 5 min). Then, the precipitate was rinsed with anhydrous ethanol again. The protein precipitate was air-dried and finally dissolved in 50μL of 8 M urea (8 M urea dissolved in 0.1 M Tris-HCl, pH 8.0). Bradford reagent was used to determine the protein concentration. The extracted protein solution was used immediately for digestion and purification or stored at -70 °C for subsequent experiments.

After the phenol organic phase was removed, 1 mL of hexadecyl trimethyl ammonium bromide (CTAB) buffer (2% CTAB, 1.4 M NaCl, 240 mM sodium phosphate buffer (pH 8.0), 20 mM EDTA, 2% polyvinylpyrrolidone (average mol wt 40000)), were added to the residual system followed by brief vortexing and incubation at 55°C for 15~20 min with occasional gentle inversion.

The centrifuge tubes were kept at room temperature for approximately 3 min, followed by the addition of an equal volume of chloroform–isoamyl alcohol (24:1). After thorough vortexing, samples were centrifuged at 10,000 × g for 10 min at 4 °C.

The upper aqueous phase was then divided into 2 mL centrifuge tubes (no more than 900 μL of sample per tube), mixed with an equal volume of the sample volume of polyethylene glycol (PEG) solution (30% PEG 6000 dissolved in 1.6 M NaCl) and 1μL of glycogen (RNase free), incubated for 2 hours at 4 °C, and centrifuged (14,000 ×g, 15 min). The supernatant was carefully aspirated to avoid precipitate loss, and the precipitate was rinsed with 1 mL of cold 75% ethanol, followed by centrifugation (14,000 ×g, 3 min). Then, the supernatant was carefully aspirated and air dried. The nucleic acid precipitate was dissolved in TE buffer (Zymo Research). The nucleic acid solution was mixed with 1μL of RNase A (TransGen) and incubated at 37 °C for 30 min, and the DNA was purified using the DNA clean beads(vazyme) according to the manufacturer's instructions. The purified DNA was stored at -70 °C until library preparation and sequencing.

Enzymatic cleavage of proteins and purification of peptides

According to the concentrations determined by the Bradford method, 1-10 μg of protein in 8M urea into a 1.5mL centrifuge tube and the final volume was adjusted to 100 μL with 8 M urea. The protein solution was mixed with 1μL of 1M DTT (1 M DTT dissolved in 8 M urea) and incubated at 37 °C for 60 min. Then, the sample was mixed with 4μL of 1M IAA (1M IAA dissolved in 25 mM ammonium bicarbonate) and incubated for 30 min at room temperature while being protected from light.

Alkylated proteins were precipitated the a by adding five volumes of acetone/acetic acid (100:0.1, v/v), mixed thoroughly and incubated overnight at 4 °C. The samples were then centrifuged at 13,000 rpm for 60 min at 4 °C. Carefully remove the supernatant, the resulting pellet was washed with 1 mL of 90% acetone, followed by centrifugation at 13,000 rpm for 30 min at 4 °C. After removal of the supernatant, the pellet was air-dried at room temperature for approximately 30 min. The dried protein pellet was resuspended in 100 μL of trypsin/Lys-C dissolved in 25mM ammonium bicarbonate, with an enzyme-to-protein ratio of 1:10–1:20 (w/w) and incubate at 37°C for 16 hours.

Following digestion, the samples were centrifuged at 13,000 rpm for 20 min and the supernatant was carefully collected. The pH of the peptide solution was adjusted to approximately 2.0 by the addition of 100% formic acid. Peptides were desalted using a micro desalting column (Empore Products) and eluted with 15 μL of 0.1% formic acid. After vortexing for 5–10 seconds and centrifugation at 12,000 rpm for 1 min, 13 μL of the eluate was transferred to a sample vial for subsequent mass spectrometry analysis.

Mass spectrometry data acquisition

Based on the concentrations determined by Nanodrop, 8 μL of peptide solution was used per injection in the nanoscale liquid chromatographic tandem mass spectrometry (nLC–MS/MS) system comprised of a nanoElute® liquid chromatography instrument (Bruker) and timsTOF Pro mass spectrometry instrument (Bruker). Peptides were separated on a reversed-phase C18 analytical column (25 cm × 75 μm ID, APT) with an overall gradient from 2% to 95% acetonitrile at a flow rate of 300 nL/min over 90 min at 60 °C. Data-dependent acquisition (DDA) was performed in PASEF mode (oTOF control v6.0.0.12). The ion mobility window was set to 1/k0 start = 0.6 Vs/cm^2 to 1/k0 end = 1.6 Vs/cm^2, with a ramp time of 100 ms with a locked duty cycle and a mass range of 100-1,700 m/z.

MS/MS spectra were acquired with an accumulation time of 2 ms and a cycle time of 100 ms. The target intensity was set at 6,000, and the threshold intensity was set at 200.

Sequencing data processing, metagenomics assembly, and binning

Approximately 10 ng of metagenomic DNA was fragmented to ~500 bp for preparation of the DNA libraries. Then, the library DNA samples were constructed by the FAST PCR-free Kit (MGI, BGI, China) and sequenced on the DNBSEQ-T series machines to obtain 150 bp of read-length raw reads. Two libraries were constructed for each sample (Table S1).

Raw reads from each library were filtered using fastp (v0.23.2) with the parameter settings (params) "–length_required 100 -w 16 –cut_front –cut_right –cut_window_size 4 –cut_mean_quality 20" [115]. For assembly, overlapping paired reads were merged with fastp with params "-m –merged_out" and those described above and assembled using MEGAHIT (v1.2.9) with params "–presets meta-sensitive -r"[116]. Raw reads were filtered again without merging to generate clean reads for recall and mapped to contigs with lengths ≥1,000 bp to calculate contig coverage using bbmap.sh (v39.01) with params "nodisk k=13 minid=0.95 keepnames=t minaveragequality=5 trimreaddescriptions=t pairlen=350 rescuedist=650"[117]. Contigs of at least 2,500 bp were used in binning. Fourteen parallel-binned results for each sample were generated (9 with MetaBAT2 (v2.15),[118] 2 with MaxBin2 (v2.2.7),[119] and 1 each with CONCOCT (v1.1.0),[120] VAMB (v3.0.2),[121] and MetaDecoder (v1.0.14)[122]), ensembled using UniteM (v1.2.4)[123] with params "greedy" and filtered with completeness – 5 × contamination ≥ 50 and the GUNC (v1.0.5).[124] Filtered bins were collected and clustered using dRep (v3.4.0)[125] at 95% ANI, and the representative genome of each cluster was collected as the MAG set used in this study. Taxonomic classification of MAGs was performed using the GTDB-Tk pipeline (v2.4.0)[126] based on the Genome Taxonomy Database (GTDB, release 220).[108] The growth conditions of all representative MAGs were predicted using GenomeSPOT (v1.0.1).[127] Maximal growth rates were estimated with gRodon (v2.5.1),[128] using ribosomal proteins listed in KEGG BRITE ko03010 as the set of highly expressed genes. Coding density, codon usage bias, the average number of nitrogen atoms in amino acid side chains, the average number of sulfur atoms in amino acid side chains, and average protein weight were calculated using in-house scripts (https://github.com/Hocnonsense/genome), with codon usage bias estimated following previously described method.[170]

Gene annotation and metaproteomic database construction

Coding sequences were predicted from contigs by pyrodigal-gv (v0.3.1),[129,131,149,171] and those encoding proteins with length < 33 amino acids were removed. Functional annotation including Pfam[109,172] and KEGG[110] was performed by MANTIS (v1.5.5),[130] and the taxonomy of the contig genes was annotated by kaiju (v1.9.2)[111] with a public database (kaiju_db_nr_euk_2022-03-10, downloaded from kaiju.binf.ku.dk). To build the contig protein dataset for metaproteomic searching, sequences with length < 50 amino acids were removed for better performance. Genes from MAGs were predicted by pyrodigal (v3.3.0),[131,171] filtered for at least 33 amino acids, and annotated with MANTIS. For COG annotation, pyhmmsearch (v2025.1.23)[132,154,173] was used to search against the COG HMM profiles provided by MetaCerberus (v1.4.0),[112,133] with an E-value cutoff of < 1e−5. Both MAG and contig genes were used for peptide searching, and 83.1% MAG-based peptides (383,134 of 453,341 peptides) were consistent with contig-based peptides (629,338 peptides), showing a high consistency.

Phylophlan (v3.0.60),[134] with the phylophlan database (downloaded automatically by the tool) with params "–diversity high –fast –min_num_markers 1", was then used to create a phylogeny alignment of the MAG set. The phylogenetic tree was reconstructed using IQ-TREE (v2.3.0) with params "-m MFP -B 1000 –bnni -redo"[135].

Metaproteomic searching, peptide validation, protein inference and quantification

Raw DDA files from three replicate injections of peptides acquired on a timsTOF Pro mass spectrometer (Bruker) were used. The FragPipe platform (v18.0)[136] with MSFragger (v3.5)[137] and Philosopher (v4.4.0)[138] were used to perform decoy generation, searching, peptide validation, and protein inference. Reversed protein sequences were appended to the original databases as decoys for MSFragger. The common repository of adventitious proteins (cRAP, https://www.thegpm.org/crap/) was used as the contaminant proteins. For the two-step database search, the initial search was performed with both precursor and fragment mass tolerances set to 40 ppm. Tryptic cleavage specificity was applied, and up to two missed trypsin cleavages were allowed. Isotope error was set to 0/1/2. Peptides with 750 amino acids and 500-5,000 Da were considered for identification. The search included variable methionine oxidation, variable protein N-terminal acetylation, and fixed carbamidomethyl cysteine modifications. The maximum number of variable modifications per peptide was set to 3. The peptide-spectrum matches (PSMs) were further rescored using deep learning prediction by the MSBooster (v1.1.4)[139] module, and discrimination between correct and decoy spectra was performed by Percolator (v3.05).[140] The validated peptides were assigned either as unique peptides to a particular protein or as razor peptides to a single protein that had the most peptide evidence by ProteinProphet (v4.4.0).[141] The results of the first step, including peptides and proteins, were filtered with a 25% false discovery rate (FDR). For the second step of the search, the inferred proteins generated from the first search were used as the database and then subjected to the same parameters, except they were filtered with a 1% FDR. The MaxLFQ intensity, including unique and razor peptides, was calculated using IonQuant (v1.8.0)[142] with a 1% match between run ion FDR and one MaxLFQ min ion. Proteins that had at least one unique peptide count were retained.

Metagenomic and metaproteomic relative abundance of contig genes/proteins

In this study, peptide abundance was defined as the sum of all unique peptide-spectrum matches (PSMs) corresponding to that peptide across all three injections, while protein abundance was defined as the total number of unique PSMs associated with all peptides belonging to a given protein. The metaproteomic relative abundance of contig proteins (subgroups) is presented in relative units across this dataset of unique PSM counts per sample,[174] which was calculated with the following equation:

$$[MP]_{protein} = \frac{[\text{unique PSM count}]_{protein}}{[\text{total unique PSM count}]} \cdot 100\%, \quad \text{(Equation 2)}$$

where $[MP]_{protein}$ refers to the metaproteomic relative abundance of a given protein (subgroup), and [total unique PSM count] refers to the sum of [unique PSM count] in one sample from all three injects.

The metaproteomic relative abundance of key functions was calculated with the following equation:

$$[MP]_{KO} = \frac{\sum [\text{unique PSM count}]_{protein}\sigma_{protein,KO}}{[\text{total unique PSM count}]} \cdot 100\%, \quad \text{(Equation 3)}$$

where $[MP]_{KO}$ refers to the metaproteomic relative abundance of a given KO, and $\sigma_{protein,KO} = 1$ if this protein was annotated as the key function else 0. The functional annotation of the protein subgroup was calculated as the lowest common taxonomy annotation of all related proteins.

The metagenomic relative abundance of individual genes on contigs was calculated as the GPM value by featureCounts (v2.0.6)[143] with params "-M—fraction" and formula

$$\text{GPM} = 10^6 \times \frac{\text{total reads mapped to gene/gene length in } bp}{\sum(\text{total reads mapped to gene/gene length in } bp)}, \quad \text{(Equation 4)}$$

and the metagenomic abundance of KO function $[MG]_{KO}$was calculated as the total GPM of all related genes.[175]

Metagenomic and metaproteomic relative abundance of MAGs

Clean reads from two libraries of each sample were concatenated and mapped to those representative MAGs with bwa (v0.7.17-r1188)[144] for CoverM (v0.6.1)[145] to calculate the metagenomic relative abundance ($[MG]_{MAG}$, or $[MG]$) of MAGs with params "-m relative_abundance". For metaproteomic-level abundance estimation, all MAG-derived proteins were compiled to construct a MAG protein database. A few protein clusters contained genes from two or more MAGs and were ignored in the analysis. Among the 552 representative MAGs, 527 were used for quantitative analysis. These MAGs were supported by at least three unique proteins in at least one sample and exhibited non-zero proteomic abundance in at least three samples. Notably, ~60% were detected in at least 11 of the 12 samples, indicating consistent population-level activity in the environment. The metaproteomic abundance of MAGs was presented considering both unique PSM counts and genomic content with the following equation, i.e., PPGM:

$$[MP]_{MAG} = \frac{\sum [\text{unique PSM count}]_{protein}\sigma_{protein,MAG}}{\sum \sigma_{protein,MAG}[\text{total MP}]} \cdot 100\%, \quad \text{(Equation 5)}$$

where $[MP]_{MAG}$ (or simpler, $[MP]$) refers to the metaproteomic abundance of MAGs, $\sigma_{protein,MAG} = 1$ if this protein was only annotated from the MAG else 0, $\sum \sigma_{protein,MAG}$ refers to the number of genes predicted from the MAG, and [total MP] was the sum of all $[MP]_{MAG}$ in a sample. The relative abundance of higher taxa was calculated as the sum of the abundance of related MAGs. We next defined the expression level a as the ratio of metaproteomic relative abundance and metagenomic relative abundance with the following equation:

$$a = \frac{[MP]}{[MG].} \quad \text{(Equation 6)}$$

For a higher taxon u consisting of n MAGs, the expression level a_u was defined with the following equation:

$$a_u = \sum_{i=1}^{n} a_i \cdot c_i, \quad \text{(Equation 7)}$$

where c_i was defined as the weighting factor of MAG i proportional to $[MG]_i$. Then we have

$$a_u = \sum_{i=1}^{n} a_i \cdot \left(\frac{[MG]_i}{\sum_{j=1}^{n} [MG]_j} \right),$$

$$= \sum_{i=1}^{n} \frac{a_i \cdot [MG]_i}{[MG]_u},$$

$$= \frac{\sum_{i=1}^{n} a_i \cdot [MG]_i}{[MG]_u},$$

$$= \frac{\sum_{i=1}^{n} \frac{[MP]_i}{[MG]_i} \cdot [MG]_i}{[MG]_u},$$

$$= \frac{\sum_{i=1}^{n} [MP]_i}{[MG]_u},$$

$$= \frac{[MP]_u}{[MG]_u}, \quad \text{(Equation 8)}$$

where $[MP]_u$ and $[MG]_u$ refer to the relative abundance of MAG set u, and for the entire community, we have

$$a_{total} = \frac{[MP]_{total}}{[MG]_{total}} = 1. \quad \text{(Equation 9)}$$

Correlations between metagenomic abundance and metaproteomic expression

The gene expression profile of a MAG measured by the metaproteomic method may be correlated to metaproteomic abundance $[MG]$, and we transformed the protein results inferred from the metaproteome to several quantitative indexes for each MAG. Any protein subgroup potentially referring to two or more MAGs was ignored as described above. Three important items from ''combined_protein.tsv'' provided by FragPipe were used, i.e., ''[experiment] spectral peptide count'', ''[experiment] unique spectral count'', and ''[experiment] MaxLFQ intensity'', which were obtained from the protein summary generated from FragPipe according to these formulas:

$$[SampleSpectralCount]_{protein} = \sum_{Injection} [[\text{experiment}]\ \text{spectral peptide count}], \quad \text{(Equation 10)}$$

$$[SampleUniqueCount]_{protein} = \sum_{Injection} [[\text{experiment}]\ \text{unique PSM count}], \quad \text{(Equation 11)}$$

$$[SampleIntensityFreq]_{protein} = \sum_{Injection} ([[\text{experiment}]\ \text{unique PSM count}] > 0), \quad \text{(Equation 12)}$$

$$[MaxLFQ]_{protein} = \begin{cases} 0 & if\ [SampleIntensityFreq] < 2, \\ \dfrac{\sum_{Injection} [[\text{experiment}]\ \text{MaxLFQ intensity}]}{[SampleIntensityFreq]} & otherwise. \end{cases} \quad \text{(Equation 13)}$$

We used these indexes to describe the protein expression of each MAG in one sample:

$$[SpectralCount]_{MAG} = \sum [SampleSpectralCount]_{protein} \sigma_{protein,MAG}, \quad \text{(Equation 14)}$$

$$[SpectralMean]_{MAG} = \frac{[SpectralCount]_{MAG}}{\sum ([SampleSpectralCount]_{protein} \sigma_{protein,MAG} > 0)}, \quad \text{(Equation 15)}$$

$$[UniqueCount]_{MAG} = \sum [SampleUniqueCount]_{protein} \sigma_{protein,MAG}, \quad \text{(Equation 16)}$$

$$[UniqueMean]_{MAG} = \frac{[UniqueCount]_{MAG}}{\sum ([SampleUniqueCount]_{protein} \sigma_{protein,MAG} > 0)}, \quad \text{(Equation 17)}$$

$$[ProteinCount]_{MAG} = \sum ([SampleUniqueCount]_{protein} \sigma_{protein,MAG} > 0), \quad \text{(Equation 18)}$$

$$[QuantProteinCount]_{MAG} = \sum ([MaxLFQ]_{protein} > 0) \sigma_{protein,MAG}, \quad \text{(Equation 19)}$$

$$[QuantSum]_{MAG} = \sum [MaxLFQ]_{protein}\sigma_{protein,MAG}, \quad \text{(Equation 20)}$$

$$[QuantMean]_{MAG} = \frac{[QuantSum]_{MAG}}{\sum([MaxLFQ]_{protein}\sigma_{protein,MAG} > 0)}, \quad \text{(Equation 21)}$$

$$[QuantMedian]_{MAG} = Q([MaxLFQ]_{protein}\sigma_{protein,MAG}, 50\%), \quad \text{(Equation 22)}$$

$$[QuantSumTop3]_{MAG} = \sum MAX\ 3([MaxLFQ]_{protein}\sigma_{protein,MAG}), \quad \text{(Equation 23)}$$

$$[QuantTop1\_4]_{MAG} = Q([MaxLFQ]_{protein}\sigma_{protein,MAG}, 25\%), \quad \text{(Equation 24)}$$

We also adjusted protein expression according to MAG gene content:

$$[PerSpectral]_{MAG} = \frac{[SpectralCount]_{MAG}}{\sum \sigma_{protein,MAG}}, \quad \text{(Equation 25)}$$

$$[PerUnique]_{MAG} = \frac{[UniqueCount]_{MAG}}{\sum \sigma_{protein,MAG}}, \quad \text{(Equation 26)}$$

$$[PerProtein]_{MAG} = \frac{[ProteinCount]_{MAG}}{\sum \sigma_{protein,MAG}}, \quad \text{(Equation 27)}$$

$$[PerQuantSum]_{MAG} = \frac{[QuantSum]_{MAG}}{\sum \sigma_{protein,MAG}}, \quad \text{(Equation 28)}$$

where $Q(x, 50\%)$ refers to the median of number set $x \in R^+$ and $MAX\ 3(x)$ refers to the highest three numbers of x (or all numbers if $\sum(x > 0) < 3$).

All indexes described here were compared with $[MG]$, and we found that $[PerUnique]_{MAG}$ obtained the highest correlation (Spearman correlation = 0.65), followed by $[PerSpectral]_{MAG}$. We next scaled $[PerUnique]_{MAG}$ to 100% in each sample to represent the intensity of genomic activity, which is referred to as $[MP]$ in other parts of this article. For indexes using MaxLFQ intensity, we found that $[QuantSumTop3]_{MAG}$ obtained high correlation with both $[QuantSum]_{MAG}$ and $[QuantMean]_{MAG}$, indicating that the highest intensity of the protein predominated over all other protein expression information.

The correlation between metagenomic relative abundance and metaproteomic relative abundance was also determined at different taxa levels, and both $[PerSpectral]_{order}$ and $[PerUnique]_{order}$ performed better than all other methods. While $[PerSpectral]_{order}$ was superior to $[PerUnique]_{order}$ for most taxonomic levels, we still chose $[PerUnique]_{order}$ because the results were better defined. In summary, $[PerSpectral]$ and $[PerUnique]$ performed well in calculating (unique) PSMs per gene per MAG at different taxonomic levels and built a bridge between the metagenome and metaproteome. We also found that when the metagenomic abundance was below 0.03%, the corresponding order might only be detected with a few proteins. We believe this abundance data may have considerable errors, so we excluded these points during fitting.

Comparing the metaproteome from hadal seawater with other metaproteomic datasets

To compare MEER-APD with other environmental metaproteomic studies, public metaproteomic studies were retrieved from the ProteomeXchange (https://www.proteomexchange.org/)[176] database using the keyword filter (species=environmental). Data from five studies on environmental marine samples, which publicly released both their mass spectrometry data and the corresponding expressed protein amino acid sequences, were downloaded.[26,28,29,35,36] The functional annotation of these publicly available protein sequences was performed using MANTIS, as described above. For the dataset from Hawley et al.,[29] only protein sequences supported by unique peptides were retained to avoid redundancy in the metagenomic database. All retained protein sequences from these studies, together with those detected in the contig-based and MAG-based dataset in this study, were clustered using MMseqs (v15.6f452)[146] following UniRef standard.[177] Briefly, protein sequences were clustered successively with params "-c 1.0 –cov-mode 1 –min-seq-id 1.0 –exact-kmer-matching 1", "–cov-mode 0 -c 0.8 –min-seq-id 0.9", and "–cov-mode 0 -c 0.8 –min-seq-id 0.5".

Phylogenetic inference of evolutionary dynamics and biogeographic patterns

To investigate the evolutionary dynamics of microbial lineages between hadal and non-hadal environments, we integrated representative genomes from three datasets: genomes from the Ocean Microbiomics Database (OMD),[50] hadal sediment genomes from the MEER project,[1] and hadal seawater genomes generated in this study. This combined dataset was dereplicated into species-level clusters at a 95% ANI threshold using dRep. We first inferred the environmental provenance of the hadal seawater MAGs based on their co-clustering patterns. Specifically, a hadal seawater MAG was designated as co-occurring in the upper ocean if it clustered with representative genomes from the OMD, whereas it was defined as present in hadal sediments if it co-clustered with genomes from the

MEER dataset. Subsequently, each species cluster could be placed into a single genus, and phylogenetic trees of all non-redundant genomes were built at the genus level for higher resolution. For each species cluster containing genomes from both hadal (> 6,000 m) and non-hadal environments, we inferred evolutionary directionality using two complementary approaches: (1) phylogenetic proximity analysis, where pairwise distances were calculated using adephylo (v1.1.17)[147] and the relative positions of hadal versus non-hadal genomes to the cluster root were compared using Wilcoxon rank-sum tests (Benjamini-Hochberg adjusted); and (2) ancestral state reconstruction via stochastic character mapping (phytools (v2.5.2),[148] SYM model, 100 replicates) to estimate the posterior probability of a hadal root state and transition frequencies. Species clusters were assigned to evolutionary pattern groups based on the following hierarchical criteria: (1) "hadal to non-hadal": clusters where the root was inferred to be hadal (p_hadal_root > 0.9) or hadal genomes were significantly closer to the root position (Wilcoxon rank-sum tests, adjusted $p < 0.05$); (2) "widespread": clusters with frequent depth-layer transitions (>1 transition per branch on average); (3) "non-hadal to hadal": remaining clusters showing gradual colonization of hadal environments. Clusters with insufficient sampling (only one genome in either environment) or ambiguous rooting (midpoint rooting due to the absence of sister species within the genus) were assigned as "unknown" and excluded from downstream analyses. All other species clusters found exclusively in the hadal zone were assigned as "hadal-specific".

Viral sequence identification, clustering, taxonomic assignment, and lifestyle prediction

Viral sequences were identified from assembled contigs ($\geq$ 5,000 bp) across 12 samples using geNomad (v1.8.0).[149] Quality assessment and trimming of non-viral regions were performed with CheckV (v1.0.1),[150] and only sequences $\geq$ 5,000 bp after trimming were retained for downstream analysis. Viral sequences were clustered into species-level vOTUs following the MIUViG criteria (95% ANI and $\geq$85% alignment fraction).[178] Pairwise similarities were calculated using all-versus-all BLAST+ (v2.16.0, -task megablast -evalue 1e-5 -max_target_seqs 25000)[151] and clusters were defined using the Leiden algorithm[179] (igraph v0.11.8,[152] resolution = 1.0) with scripts adapted from a publicly available pipeline (https://github.com/apcamargo/bioinformatics-snakemake-pipelines/tree/main/contig-ani-leiden-clustering-pipeline).[153] Representative vOTUs were re-evaluated with CheckV, and taxonomic assignment was performed using the annotate function in geNomad. Viral lifestyle was predicted using multiple complementary approaches. Temperate viruses were identified based on: (i) prophage signals detected by geNomad or CheckV; (ii) presence of lysogeny-related genes (e.g., integrase, recombinase, transposase, repressor, ParA/ParB) identified using Pfam HMMs with hmmsearch (v3.4, –cut_ga)[180,181] (iii) prediction by VIBRANT (v1.2.1)[155]; (iv) concordant temperate prediction by PhaTYP (score > 0.75, included in the PhaBOX2 software, v2.1.9)[156,157] and ProkBERT PhaStyle (v0.1, PhaStyle-mini model)[158]; or (v) detection of lysogeny-associated VOGs in MetaCerberus annotations.[113] vOTUs lacking evidence for lysogeny were classified as putatively virulent. A full list of lysogeny-related marker genes is provided in the Zenodo repository described in the Data and code availability section.

Prediction of virus–host linkage

All reconstructed MAGs were used as the host reference database. Virus-host linkages were predicted based on four different in silico strategies: CRISPR spacer matching, nucleotide homology, k-mer composition, and tRNA matching. Host prediction was first conducted using iPHoP (v1.3.3)[159] with default parameters. Shared tRNAs between viruses and hosts were identified by annotating tRNAs with tRNAscan-SE (v2.0.12)[160] ("-B" for bacterial MAGs, "-A" for archaeal MAGs, and "-G" for vOTUs). Viral and host tRNAs were compared using BLASTn with stringent criteria ("-perc_identity 100 -qcov_hsp_perc 100"), requiring 100% identity over 100% coverage. Host prediction was further performed using CHERRY[161] implemented in PhaBOX2. Virus–host associations supported by at least one method were retained as putative linkages.

vOTU abundance calculation

Relative abundance was used to quantify vOTUs at both the metagenomic and metaproteomic levels. Metaproteomic abundance was calculated using the previously defined PPGM metric. For metagenomic data, reads were first mapped to vOTUs using bwa. Subsequently, vOTU relative abundance was estimated using the contig mode in CoverM, applying thresholds of 95% sequence identity and 90% aligned fraction. Contigs with less than 70% of their length covered by mapped reads in a given sample were assigned zero coverage. Average contig coverage was calculated using the trimmed_mean method in CoverM, which excludes the highest and lowest 10% of coverage values. Normalization was performed by dividing the average coverage of each contig by the number of trimmed reads in the corresponding sample and then scaling by the mean read count across all samples.

Viral gene annotation

Coding sequences were predicted from representative vOTUs using the annotate function in geNomad. To improve gene identification and functional annotation, multiple complementary databases and tools were applied. Annotations from geNomad were retained, while additional annotations were generated using PHAROKKA (v1.9.0, pharokka_proteins.py function with default settings)[162] against the PHROG database,[114] pyhmmsearch against the COG HMM profiles provided by MetaCerberus (E-value < 1e-05), and MANTIS for Pfam and KEGG assignments. Viral hallmark genes were identified if any of the following criteria were met: (i) geNomad_virus_hallmark = 1; (ii) PHROG annotations belonging to the categories 'head and packaging', 'connector', 'tail', 'integration and excision', or 'lysis'; (iii) COG annotations assigned to category X (Mobilome: prophages, transposons); or (iv) KEGG classification corresponding to Viral protein families. In addition, a supplementary list of potential hallmark genes based on annotations from the KEGG, Pfam, and PHROG databases was compiled following the CheckAMG framework (https://github.com/AnantharamanLab/CheckAMG),[163] with the list provided in the Zenodo repository described in the Data and code availability

section. Viral core genes, defined as genes closely associated with viral replication and structural functions, included hallmark genes, genes matching a curated list of lysogeny-related markers, and genes whose annotations contained specific keywords (also see Data and code availability section). Auxiliary metabolic genes (AMGs) were identified when located between two viral core genes on a viral contig or when at least one hallmark gene was present on the same contig. Candidate AMGs were further manually inspected, and functional assignments were prioritized in the order KEGG > COG > Pfam > PHROG.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed in R (v 4.4, https://www.r-project.org/).[164] KEGG functional enrichment analysis was performed using the clusterProfiler package.[165] Metaproteomic and metagenomic relative abundances were modeled using a linear regression forced through the origin (Equation 1) with stats::lm() (v 4.4.3),[164] and model significance was assessed via F-test using stats::pf(). Orders with non-zero proteomic abundance in at least three samples were included, with those below 0.03% metagenomic abundance excluded from the analysis. Taxa with adjusted $R^2 < 0.5$ or $p > 0.01$ were designated as "unfitted"; remaining taxa were classified as "highly active" ($a > 1$) or "lowly active" ($a < 1$). For virus–host regression analyses, linear regression and Pearson correlation coefficients were computed using ggplot2::geom_smooth(method = "lm") (v 4.0.1)[166] and stats::cor.test(), respectively. Where applicable, differences between two groups were assessed using two-tailed Wilcoxon rank-sum tests via ggpubr::stat_compare_means() (v 0.6.2),[167] with $p \leq 0.05$ considered statistically significant. Data distributions are displayed as violin plots with embedded boxplots, with the center line indicating the median and box bounds indicating the interquartile range. For COG functional category enrichment analysis, a 2×2 contingency table was constructed for each functional category in highly active and lowly active groups independently, comparing the number of COGs with detected proteins in the metaproteome versus total COGs annotated in the metagenome. Two-tailed Fisher's exact test was applied using stats::fisher.test(), and resulting p-values were corrected using the Benjamini–Hochberg (BH) method. Community dissimilarity between samples was assessed using non-metric multidimensional scaling (NMDS) based on Bray–Curtis dissimilarity at the order level using vegan::metaMDS() (v 2.7.2).[168] Permutational multivariate analysis of variance (PERMANOVA) was conducted using vegan::adonis2() with location as the grouping variable (Bray–Curtis dissimilarity, 999 permutations) to test for significant differences among groups. The number of observations (n) for each statistical analysis is reported in the corresponding figure legends. Gini index was calculated using the R package ineq (v0.2.13).[169]