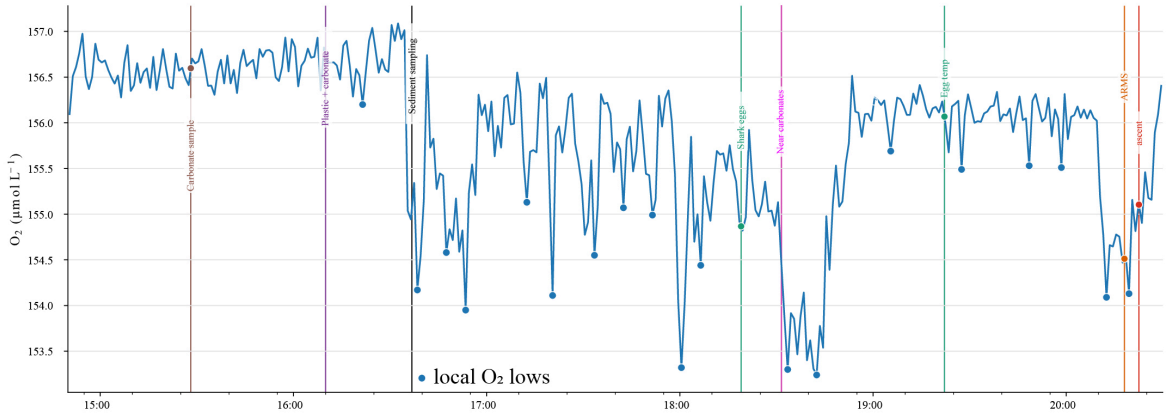
Supplementary Information for

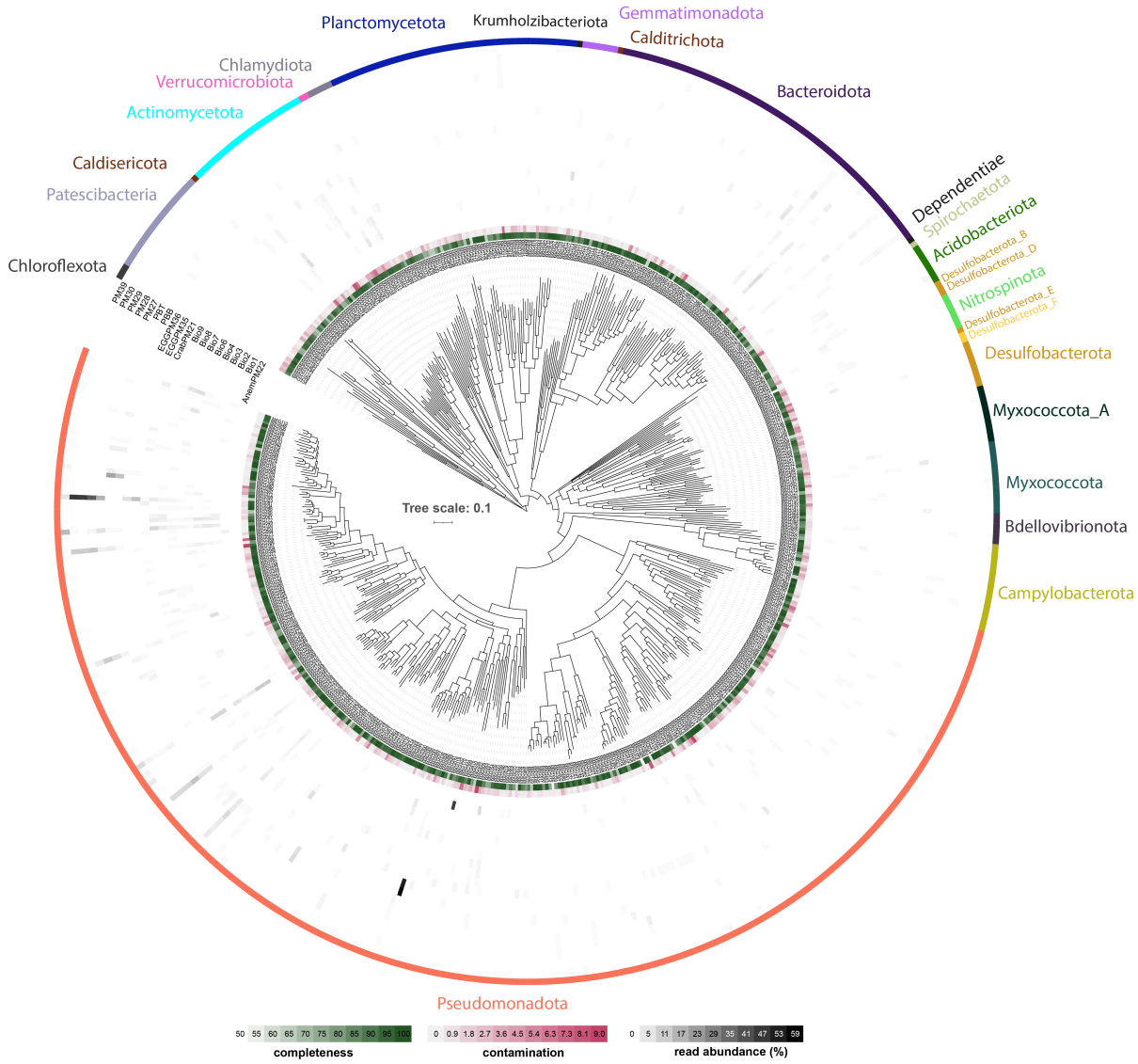
Hydrocarbon seep biofilms share a common functional organization across diverse substrates

Maxim Rubin-Blum, Tony Tendler, Stas Malavin, Lina Ratinskaia, Zoya Harbuzov, Yana Yudkovski, Lyubov Shmakova, Sheli Shlomo Itzhari, Matan Oren, Yizhaq Makovsky, Tamar Guy-Haim

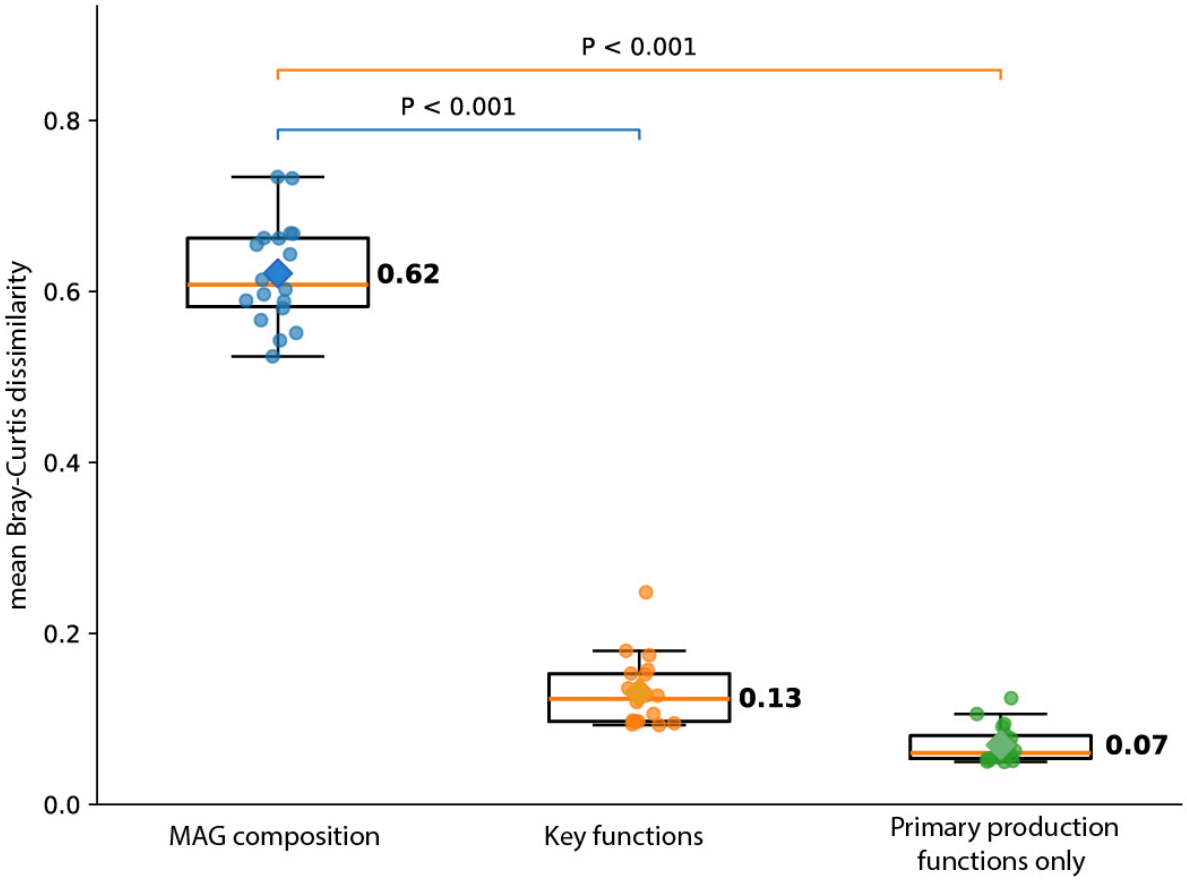
Supplementary figures:



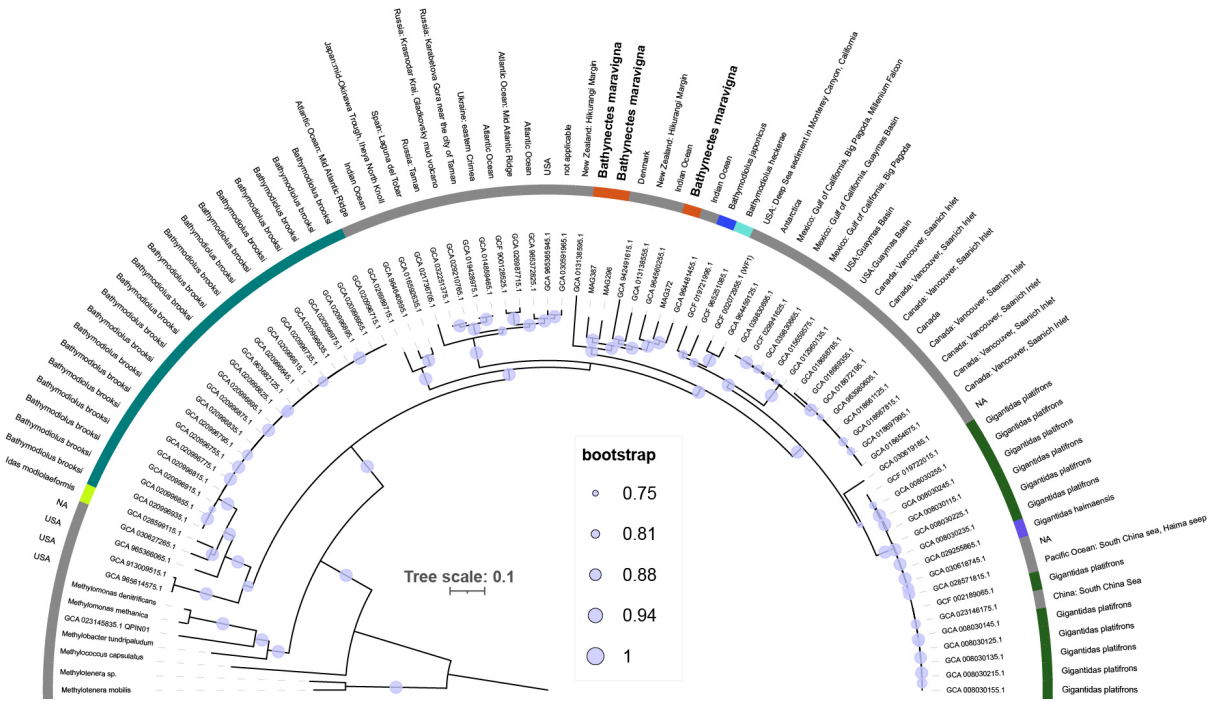
Supplementary Fig. S1 | Alignment of dissolved oxygen measurements with ROV operations during the ROV dive on 14 January 2025, in which we assessed all the substrates discussed in this paper. Continuous dissolved oxygen (O_2) concentrations recorded by the SBE 16plus CTD mounted on the ROV manipulator arm are shown together with major sampling and operational events identified from synchronized ROV video and dive logs. Vertical lines indicate the timing of carbonate, plastic, sediment, shark egg, and ARMS sampling, temperature measurements on shark eggs, and vehicle ascent. Filled circles denote local O_2 minima identified along the time series. The temporal correspondence between O_2 anomalies and sampling activities provided a basis for interpreting fine-scale oxygen variability across different seafloor substrates. It informed the selection of representative sampling locations discussed in the main text.



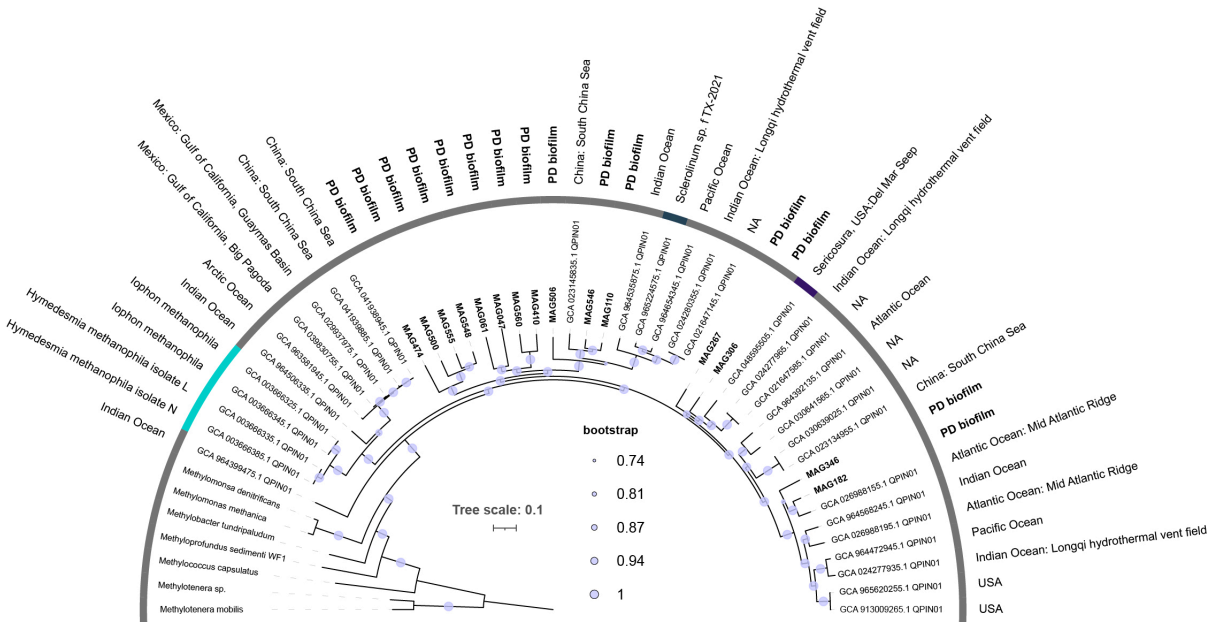
Supplementary Fig. 2 | Supplementary Fig. 2 | Phylogenomic tree of metagenome-assembled genomes recovered from deep-sea biofilms. Phylogenomic tree of bacterial metagenome-assembled genomes (MAGs) reconstructed in this study, based on their placement in the GTDB-Tk bacterial reference tree using the bac120 set of bacterial marker genes. Branch lengths represent substitutions per site (scale bar, 0.1 substitutions per site). The outer colored ring indicates phylum-level taxonomic affiliation. Concentric annotation tracks show estimated genome completeness and contamination determined using CheckM2, relative abundance across individual samples (normalized genome coverage, %), and the habitat distribution of each MAG. Relative abundance is shown for carbonate biofilms (Bio1–Bio9), shark egg capsules (EGGPM35–36), plastic biofilms (PM27–30 and PM39), ARMS biofilms (PBB and PBT), crab carapace biofilms (CrabPM21), and host-associated anemone samples (AnemPM22).



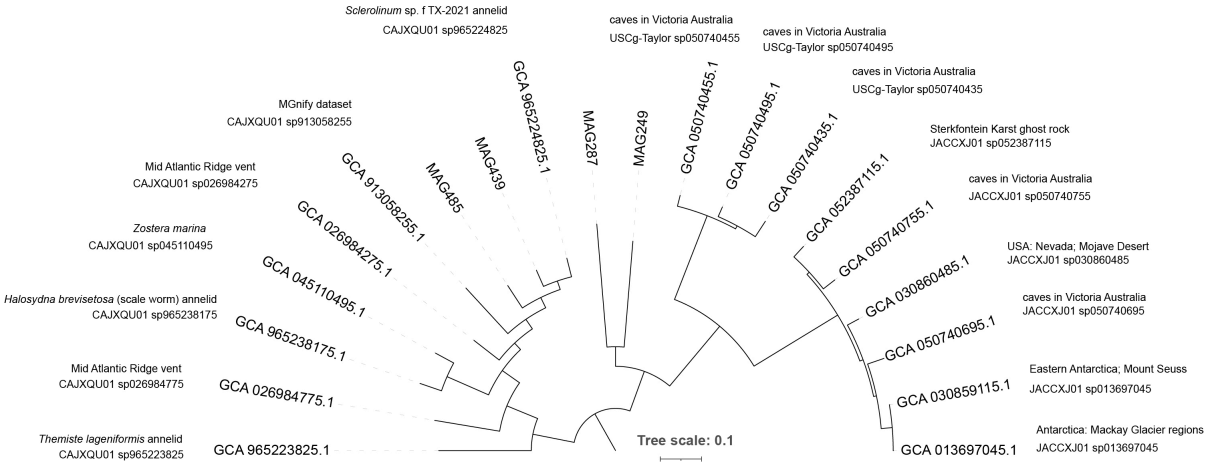
Supplementary Fig. 3 | Taxonomic turnover exceeds functional turnover across seep biofilms. Boxplots show the Distribution of mean Bray–Curtis dissimilarities for each biofilm calculated from MAG relative abundances, abundance-weighted ecological guilds, or abundance-weighted primary-production traits. For each sample, we calculated its mean dissimilarity to all other biofilm samples. Ecological guilds comprised methanotrophs, sulfur-oxidizing autotrophs, other autotrophs, and other community members, whereas primary-production profiles comprised methanotrophy, sulfur-oxidizing autotrophy, and autotrophic carbon fixation, as defined by the curated genome-resolved ecological annotations in **Supplementary Table S4**. Points represent individual biofilm samples; boxes show the interquartile range with the median; whiskers extend to 1.5 × the interquartile range; and diamonds indicate means, with values shown above each Distribution. We calculated Bray–Curtis dissimilarities after Hellinger transformation of relative abundance profiles. *P* values indicate one-sided paired Wilcoxon signed-rank tests comparing taxonomic and functional dissimilarities. **Detailed beta-diversity statistics, tests of taxonomic–functional concordance, and functional redundancy and core-guild metrics are provided in Supplementary Table S5.**



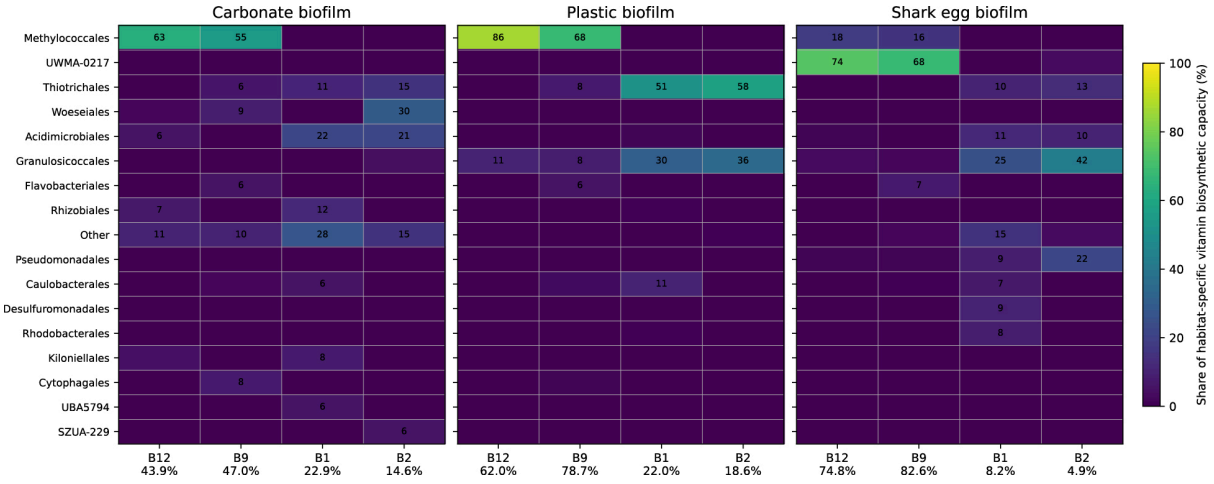
Supplementary Fig. 4 | Phylogenomic placement of biofilm *Methyloprofundus*. Maximum-likelihood phylogenomic tree of *Methyloprofundus* genomes recovered from public databases together with MAGs reconstructed in this study (bold). The tree was inferred from concatenated bacterial single-copy marker proteins. Branch lengths represent substitutions per site, and node support is shown as ultrafast bootstrap values, with circle size proportional to bootstrap support. The outer annotation indicates the host organism or environmental source of each genome. We included reference methylotrophs *Methylococcus capsulatus*, *Methylobacter tundripaludum*, *Methylomonas methanica*, *Methylomonas denitrificans* and *Methylotenera* spp. to root the tree.



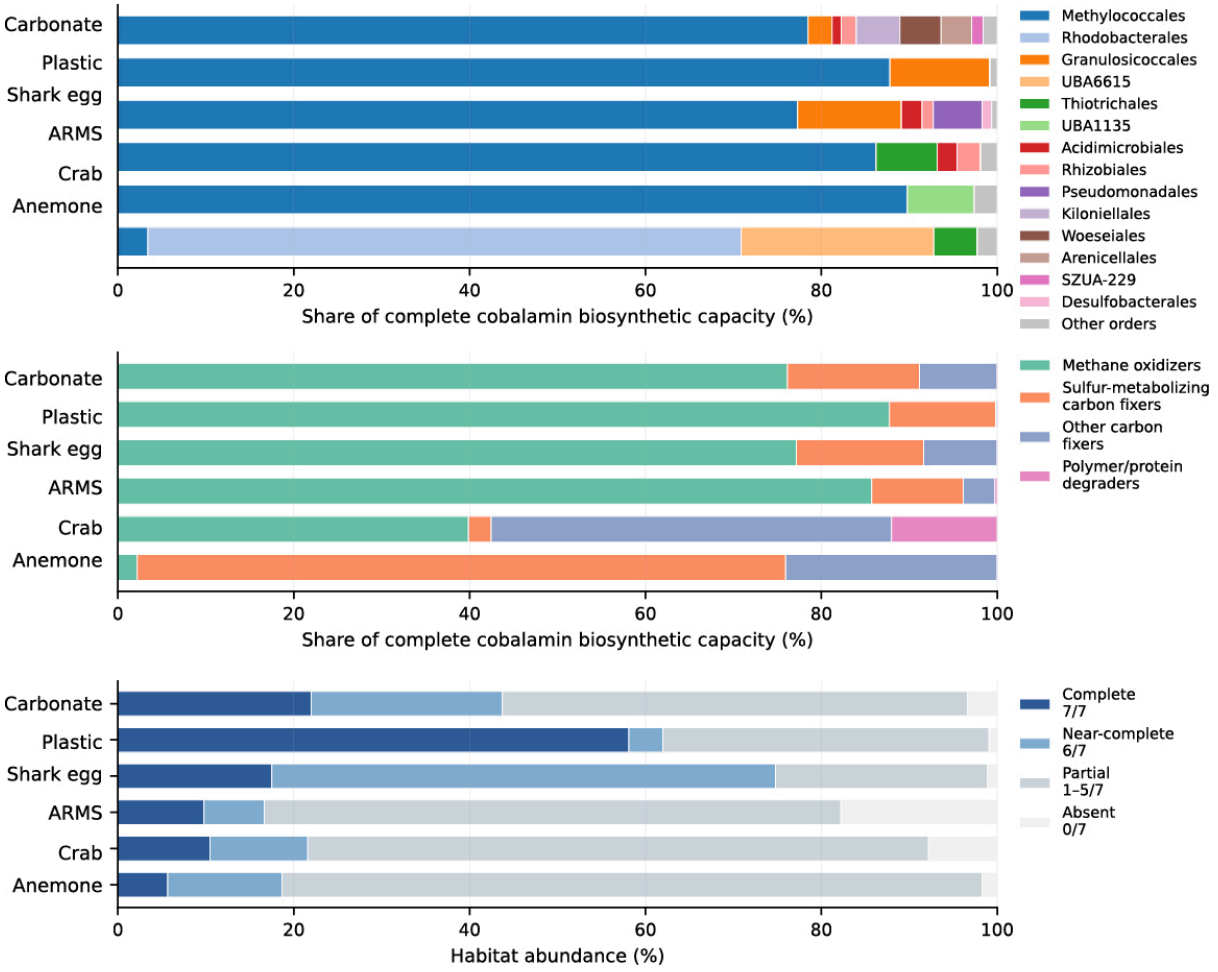
Supplementary Fig. 5 | Phylogenomic placement of biofilm QPIN01 (Marine Methylotrophic Group 2). Maximum-likelihood phylogenomic tree of genomes affiliated with the QPIN01 lineage (Marine Methylotrophic Group 2), including MAGs reconstructed in this study (bold) and publicly available genomes. The tree was inferred from concatenated bacterial single-copy marker proteins. Branch lengths represent substitutions per site, and node support is shown as ultrafast bootstrap values, with circle size proportional to bootstrap support. The outer annotation indicates the host organism or environmental source of each genome. Reference methylotrophs *Methylococcus capsulatus*, *Methylobacter tundripaludum*, *Methylomonas methanica*, *Methylomonas denitrificans* and *Methylotenera* spp. were included to root the tree.



Supplementary Fig. 6 | Genome phylogeny of the candidate methylotrophic order CAJXQU01. Maximum-likelihood phylogeny inferred from concatenated conserved bacterial marker genes. The tree includes all publicly available GTDB genomes assigned to *Ca. Methylocavales* together with four MAGs recovered from Eastern Mediterranean methane seep biofilms (MAG249, MAG287, MAG439, and MAG485) - representatives of the sister candidate order *Ca. Methyloligotrophales* (formerly JACCXJ01) and the atmospheric methanotroph lineage USCγ-Taylor were included as outgroups. Environmental origins of reference genomes are indicated alongside the tree, including cave, hydrothermal vent, marine invertebrate-associated, terrestrial, and environmental metagenomic sources.



Supplementary Fig. 7 | Genomic capacity of habitat-specific vitamin provider contributions. The heatmap is based on a METABOLIC-derived vitamin biosynthesis matrix and the habitat-specific MAG abundance table; for each habitat and vitamin, MAGs encoding a complete biosynthetic pathway were grouped by taxonomic order. Cell values show the percentage contribution of each order to the total habitat-specific biosynthetic capacity for that vitamin. Values below the vitamin labels indicate the percentage of total habitat abundance represented by MAGs that encode the complete pathway. Only orders contributing at least 5% to at least one habitat-vitamin combination are shown; remaining low-abundance providers are represented by 'other' where applicable.



Supplementary Fig. 8 | Community allocation of cobalamin biosynthetic capacity across seep biofilm habitats. a, Taxonomic contribution to complete cobalamin biosynthetic capacity. Bars show the percentage contribution of each taxonomic order to the summed abundance of metagenome-assembled genomes encoding all seven METABOLIC/KEGG M00122 steps for cobalamin biosynthesis from cobinamide to cobalamin. Orders contributing less than 1% each were combined into 'other' orders. b, Functional contribution to the same complete cobalamin biosynthetic capacity. Complete cobalamin-producing genomes were assigned once to hierarchical functional groups to avoid double counting: methane oxidizers, sulfur-metabolizing carbon fixers, other carbon fixers, polymer/protein degraders, and other cobalamin providers. c, Habitat-level Distribution of cobalamin pathway completeness, calculated as the percentage of total habitat abundance assigned to genomes with complete (7/7), near-complete (6/7), partial (1–5/7), or absent (0/7) M00122 step profiles. All bars are abundance-weighted using habitat-specific mean relative abundance.

Supplementary Excel Table legends:

Supplementary Table S1 | Operational definitions of metabolic and interaction-associated traits used for genome-resolved ecological classification. The table lists the criteria used to assign metabolic pathways, ecological functions, and interaction-associated traits to MAGs. For each trait, the table provides the annotation source, operational definition, and level of evidence required for assignment. Complete pathways required the specified set of diagnostic genes or pathway components, whereas diagnostic-gene assignments indicate functions supported by individual marker genes or gene combinations. These criteria provide the basis for the ecological classifications reported in Supplementary Table S4 and for abundance-weighted functional analyses.

Supplementary Table S2 | Software, databases, and parameters used for genome-resolved metagenomic and functional analyses. The table summarizes the computational workflow used for metagenome assembly, genome recovery and quality assessment, taxonomic classification, read mapping, functional annotation, and downstream characterization of metabolic and interaction-associated traits. For each analysis step, the table reports the software or database, version, and key parameters or settings used.

Supplementary Table S3 | ATLAS summary statistics, taxonomy and relative abundance of metagenome-assembled genomes recovered from seep biofilms.

Supplementary Table S4 | Genome-resolved ecological classification of seep biofilm microorganisms. The table summarizes taxonomy, ecological guild assignment, metabolic traits, respiratory capacities, nutrient-acquisition pathways, and interaction-associated traits for the 565 recovered MAGs. Binary columns indicate trait assignment (1, present; 0, not assigned) according to the operational criteria in Supplementary Table S1. Ecological_guild classifies MAGs as methanotrophs, sulfur-oxidizing autotrophs, other autotrophs, or other community members. Sulfur_oxidizing_autotroph identifies MAGs that combine diagnostic sulfur-oxidation capacity with autotrophic carbon fixation rather than MAGs carrying individual sulfur-metabolism genes alone. Autotrophic_carbon_fixation identifies MAGs with supported autotrophic carbon-fixation pathways, while Carbon_fixation_pathway, CBB_I, CBB_II, and rTCA specify the inferred pathway. Additional columns report methane oxidation, sulfur oxidation, terminal oxidases, nitrogen metabolism, urea utilization, pili, and secretion systems.

Supplementary Table S5 | Taxonomic and functional beta-diversity and functional redundancy across seep biofilms. We calculated taxonomic beta-diversity from MAG-level relative abundance profiles and functional beta-diversity from abundance-weighted ecological classifications (Supplementary Table S4). Strict ecological guild profiles comprised methanotrophs, sulfur-oxidizing autotrophs, other autotrophs, and other community members. Strict primary-production profiles included methanotrophy, sulfur-oxidizing autotrophy, and autotrophic carbon fixation. We defined sulfur-oxidizing autotrophs according to the curated “Sulfur oxidizing autotroph” classification rather than individual sulfur-metabolism genes alone, and autotrophic carbon fixers according to the curated “Autotrophic carbon fixation classification”. Carbon-fixation pathways distinguish CBB Form I, CBB Form II, and rTCA assignments.

We calculated Bray–Curtis dissimilarities after Hellinger-transforming sample profiles. Mean, median, and interquartile values summarize all pairwise dissimilarities. Within-substrate values include comparisons between samples belonging to the same substrate type, whereas between-substrate values include comparisons across substrate types. PERMANOVA tested differences among replicated

carbonate, plastic, shark egg capsule, and ARMS biofilms using 9,999 permutations; we excluded the single crustacean-carapace sample from this test. The paired Wilcoxon signed-rank test compares taxonomic and functional dissimilarity using each sample's mean distance to all other samples. Mantel tests compare the corresponding taxonomic and functional distance matrices, while Procrustes analyses compare their ordination configurations; both used 9,999 permutations.

For each ecological guild and carbon-fixation pathway, Carrier MAGs gives the number of MAGs assigned to that function and Carrier orders gives the number of taxonomic orders represented by those MAGs. Min, median, and max abundance report the summed relative abundance of the corresponding MAGs across samples. Samples ≥1% gives the number of the 18 biofilm samples in which the summed abundance reached at least 1%, and Core ≥1% all 18 indicates whether this threshold occurred in every sample. Median effective MAG number describes functional redundancy as the inverse Simpson effective number of MAGs contributing to that function within each sample, summarized as the median across samples.

Supplementary Table S6 | Genome-resolved metabolic profiles of heterotrophic microorganisms in seep biofilms. The table summarizes taxonomy, abundance, organic-matter degradation potential, nutrient and organic-substrate uptake, and fermentation-related functions for 379 MAGs classified as heterotrophic community members. Mean relative abundance and Max relative abundance report each MAG's mean and maximum relative abundance across biofilm samples, while Samples above 1% and Samples above 5% indicate the number of samples in which the MAG exceeded these abundance thresholds; Dominant sample identifies the sample in which each MAG reached its highest relative abundance. Transporter columns report gene-copy counts for TonB-dependent receptors and systems associated with oligopeptide, dipeptide, branched-chain amino acid, polyamine, organic-acid, phosphate, and taurine uptake. Fermentation-related columns summarize annotated pathways for acetate, acetaldehyde, and ethanol metabolism; lactate utilization; acetogenesis; and pyruvate conversion to acetyl-CoA and formate; the fermentation function count gives the number of these functions identified in each MAG. MEROPS hits and MEROPS families report the number and diversity of annotated peptidases, whereas CAZyme hits and CAZyme families report the corresponding abundance and diversity of carbohydrate-active enzymes. Transport KO total gives the total number of transporter-associated KEGG ortholog annotations included in the analysis.

Supplementary Table S7 | Distribution of peptidase repertoires among major microbial orders in seep biofilms. The table summarizes MEROPS-annotated peptidases for microbial orders represented among the recovered MAGs. No. MAGs gives the number of genomes assigned to each order, while Mean MEROPS genes/MAG and Mean MEROPS families/MAG describe the size and diversity of their peptidase repertoires. Functional columns group MEROPS families into extracellular protein degradation, cell-wall and biofilm remodeling, peptide processing and nutrient recovery, and other peptidases according to the definitions provided in the Category_definitions sheet. Values give the mean number of genes per MAG followed by the predominant MEROPS families contributing to each category. The functional categories provide conservative ecological interpretations of peptidase repertoires rather than direct measurements of substrate degradation.

Supplementary Table S8 | Distribution of carbohydrate-active enzyme repertoires among major microbial orders in seep biofilms. The table summarizes CAZyme repertoires across microbial orders represented among the recovered MAGs. No. MAGs gives the number of genomes assigned to each order, while Mean CAZyme genes/MAG and Mean CAZyme families/MAG describe the abundance and diversity of carbohydrate-active enzymes. Functional columns group CAZyme families by conservative substrate associations, including bacterial cell-wall glycans, storage α-glucans, marine

polysaccharides and β-glucans, and complex glycoconjugates and oligosaccharides. Values give the mean number of genes per MAG followed by the principal CAZyme families contributing to each category. The Family_key sheet lists the families included in each category and their ecological interpretation, while the Summary sheet reports dataset-wide CAZyme statistics.

Supplementary Methods

ROV CTD measurements and environmental reconstruction: We monitored environmental conditions during the ROV dive with a Sea-Bird SBE 16plus V2 CTD equipped with temperature, conductivity, pressure, and dissolved-oxygen sensors. A flexible hose extended the water intake toward the manipulator arm so that measurements reflected water surrounding sampling operations rather than the vehicle body. The instrument recorded measurements at 60-s intervals throughout the dive.

We converted raw hexadecimal (.hex) files to engineering units using the instrument-specific calibration coefficients in the accompanying XMLCON configuration file. We calculated temperature, conductivity, salinity (PSS-78), pressure and dissolved oxygen using the manufacturer's calibration equations and expressed dissolved oxygen as μmol L^{-1} after applying the instrument calibration and oxygen-solubility corrections.

We synchronized the CTD record with the ROV dive log using dive duration and logged operational events, including seabed touchdown, video transitions, sampling operations, takeoff and vehicle recovery. Independently documented events provided additional alignment constraints. We then associated CTD measurements with the corresponding dive phase, video segment, and sampling activity, and extracted environmental measurements surrounding carbonate, plastic, shark egg capsule, and ARMS collections. Because the CTD sampled once per minute, we focused on sustained deviations in water properties rather than short-lived excursions that may have occurred below the sampling interval. We used the reconstructed timeline to provide environmental context for biological sampling rather than to assign precise plume-encounter times.

Genome reconstruction, quality assessment, taxonomy and abundance: We processed metagenomic libraries with ATLAS v2.18.1[1]. ATLAS assembled reads with metaSPAdes v3.15.3 using k-mer sizes 21, 33, 66, 99 and 127 and binned metagenome-assembled genomes (MAGs) with VAMB v3.0.4[2]. We dereplicated MAGs with dRep[3] at 97.5% average nucleotide identity and retained the highest-quality representative from each cluster. CheckM2[4] provided completeness and contamination estimates, and GTDB-Tk v2.3 classified genomes against the Genome Taxonomy Database. **Supplementary Table S3** reports genome quality, taxonomy, and sample-level abundance information for the final MAG catalog.

We estimated MAG abundance by competitively recruiting metagenomic reads against the dereplicated catalog with Minimap2 and retained mappings at ≥99% nucleotide identity. We normalized recruited reads by genome length and sequencing depth and expressed values as relative abundance within each sample. Sample-level relative abundances supported community-level statistical analyses. For habitat-level summaries and abundance-weighted visualizations, we calculated the mean relative abundance across biological replicates belonging to the same habitat.

Functional annotation and ecological trait reconstruction: We predicted protein-coding genes with Prodigal[5] and annotated proteins with RAST[6], eggNOG-mapper[7] and METABOLIC[8]. METABOLIC supplied KEGG-module and HMM-based evidence for metabolic reconstruction. We annotated carbohydrate-active enzymes with dbCAN2 and peptidases with the MEROPS database. MacSyFinder v2.1.4 identified secretion systems, pili and related surface-associated machineries with the TXSScan and TFF-SF models, and QSAP identified quorum-sensing and signaling systems. Software versions and principal parameters are summarized in **Supplementary Table S2**.

We assigned methane oxidation, methanol oxidation, sulfur oxidation, autotrophic carbon fixation, nitrogen metabolism, urea utilization, vitamin biosynthesis, polymer degradation, and interaction-associated traits using conservative operational definitions that required the diagnostic genes or pathway combinations listed in **Supplementary Table S1**. We then integrated these calls with GTDB taxonomy to assign each MAG to the ecological classes reported in **Supplementary Table S4**. We treated sulfur oxidation and sulfur-oxidizing autotrophy separately: the sulfur-oxidizing autotroph class required both diagnostic sulfur-oxidation capacity. It supported autotrophic carbon fixation rather than the presence of individual sulfur-metabolism genes alone. Likewise, we assigned autotrophic carbon fixation only when the diagnostic components of a supported fixation pathway were present; **Supplementary Table S4** distinguishes CBB Form I, CBB Form II, and rTCA assignments.

Methane oxidation and methane-assimilation pathway classification: We classified particulate methane oxidation only when a MAG encoded the complete *pmoABC* operon. We retained partial pmo or soluble methane monooxygenase marker hits for annotation audits but did not use isolated hits as evidence of complete methane oxidation. In the C1 interaction analysis, the methane-oxidizer group comprised Methylococcales MAGs that met the methane-oxidation criteria together with four non-Methylococcales MAGs (MAG066, MAG249, MAG287 and MAG295) encoding complete *pmoABC* operons.

We classified methane-assimilation pathways from KEGG orthologs in the METABOLIC KEGGModuleStepHit output. We identified the ribulose monophosphate (RuMP) core from hexulose-phosphate synthase (HPS; K08093) and phosphohexuloisomerase (PHI; K08094). We assigned the Entner-Doudoroff (EDD) regeneration variant when both 6-phosphogluconate dehydratase (Edd; K01690) and 2-keto-3-deoxy-6-phosphogluconate aldolase (Eda; K01625) were present. We assigned the Embden-Meyerhof-Parnas (EMP) variant when a fructose-bisphosphate aldolase (K01622, K11645, or K16305), a fructose-1,6-bisphosphatase (K03841, K02446, or K04041), and phosphofructokinase (PfkA, K00850, or PfkB, K16370) were present. MAGs encoding both regeneration routes were classified as RuMP-EDD+EMP, whereas MAGs with the RuMP core but without complete evidence for either branch were classified as RuMP-unresolved. We preferred this gene-based classification to KEGG-module completion when branch-specific enzymes were present despite incomplete module calls.

Phylogenomic analyses: We used the bacterial bac120 phylogeny generated by GTDB-Tk to represent relationships among the recovered bacterial MAGs (**Supplementary Fig. S2**). We extracted the Methylococcales subtree from this phylogeny for the focused analysis shown in **Fig. 5**. For lineage-specific phylogenies of Methyloprofundus, QPIN01/MMG2 and CAJXQU01 (**Supplementary Figs. S4–S6**), we used GToTree v1.7.10[9] with the Gammaproteobacteria HMM set comprising 172 conserved single-copy markers. For lineage-specific analyses of Methylococcales and related methane-oxidizing lineages, including QPIN01, we selected reference genomes spanning the relevant GTDB diversity and environmental origins. GToTree excluded genomes containing fewer than 50% of the targeted markers and inferred phylogenies from concatenated amino acid alignments with FastTreeMP v2.1.11 under the JTT+CAT model, with SH-like support values based on 1,000 resamples. We visualized and annotated trees in iTOL v7[10].

Genome-resolved metatranscriptomic analyses: We mapped quality-filtered paired-end metatranscriptomic reads from two shark egg biofilms and the PM39 plastic biofilm against the complete MAG catalog with Bowtie2 in --very-sensitive mode[11]. We retained properly paired alignments with mapping quality ≥20 using SAMtools[12]. Prodigal-predicted coding sequences served as genomic features for featureCounts[13]. We standardized gene identifiers across protein, GFF, and annotation files before integrating read counts with functional annotations.

We normalized gene expression as transcripts per million (TPM) after correcting read counts for gene length. We summed gene-level TPM values within each MAG to estimate genome-level transcriptional activity. We integrated TPM values with eggNOG annotations, including KEGG orthologs, PFAM

domains, CAZyme annotations and functional descriptions, to generate genome-resolved expression matrices. Unless otherwise stated, comparative ecological trait matrices included MAGs contributing at least 1% of the total transcript abundance in a library. We applied the same conservative metabolic definitions used for the metagenomic analyses when assigning expressed pathways.

Genome-resolved characterization of heterotrophic functions: We characterized heterotrophic MAGs by integrating sample-level relative abundance with transporter repertoires, fermentation-related functions, MEROPS peptidases, and carbohydrate-active enzymes (**Supplementary Tables S6-S8**). For each MAG, we calculated mean and maximum relative abundance across biofilm samples, counted the number of samples in which the MAG exceeded 1% or 5% relative abundance, and recorded the sample in which it reached its maximum abundance.

We summarized transporter-associated KEGG orthologs for TonB-dependent receptors and systems involved in oligopeptide, dipeptide, branched-chain amino acid, polyamine, organic acid, phosphate, and taurine uptake. We summarized fermentation-related potential from curated pathway calls for acetate, acetaldehyde, and ethanol metabolism; lactate utilization; acetogenesis; and pyruvate conversion to acetyl-CoA and formate. For each MAG, we counted MEROPS-annotated peptidase hits and distinct MEROPS families and dbCAN-annotated CAZyme hits and families. For order-level summaries, we calculated the mean numbers of genes and families per MAG. We grouped MEROPS and CAZyme families into the conservative functional categories described in **Supplementary Tables S7 and S8**. These categories describe genomic potential and do not by themselves demonstrate degradation of a specific extracellular substrate.

Abundance-weighted analysis of vitamin biosynthetic capacity: We identified complete vitamin biosynthetic pathways according to the operational definitions in **Supplementary Table S1** and combined these assignments with habitat-specific MAG relative abundances. For each vitamin and habitat, we summed the abundance of MAGs encoding the complete pathway and partitioned this abundance by taxonomic order. We grouped low-contribution orders as described in the corresponding figure legends.

For cobalamin, we evaluated the seven METABOLIC/KEGG M00122 steps for conversion of cobinamide to cobalamin and classified MAGs as complete (7/7 steps), near-complete (6/7), partial (1-5/7) or absent (0/7). To summarize the ecological identity of complete cobalamin producers without double counting, we assigned each MAG once to a hierarchical functional category comprising methane oxidizers, sulfur-oxidizing autotrophs, other autotrophs, polymer/protein degraders, or other

cobalamin providers. All habitat-level vitamin summaries used abundance-weighted MAG contributions.

Construction of abundance-weighted alluvial plots: We generated the abundance-weighted alluvial plot in Fig. 3 with custom Python scripts using pandas, NumPy, and svgwrite. We ranked taxonomic orders by cumulative MAG relative abundance, displayed the 15 most abundant orders individually, and collapsed the remainder as “other taxa”. For each displayed taxon-feature pair, we summed the relative abundance of MAGs carrying the feature. Node heights and ribbon widths used the square root of cumulative MAG abundance, matching the figure legend, to preserve quantitative ranking while retaining lower-abundance links.

We calculated phi coefficients from genome-level binary taxon-feature membership. We tested taxon-feature associations with two-sided Fisher’s exact tests followed by Benjamini-Hochberg correction across the complete set of tests. We retained positive associations for visualization, and ribbon opacity in Fig. 3 scales with the positive phi coefficient.

Construction of methane-derived one-carbon interaction networks: We reconstructed potential methane-derived C1 interactions from genome-resolved metabolic traits and habitat occurrence. The final workflow used MAG relative-abundance profiles together with METABOLIC HMM annotations. We assigned MAGs to methane-donor, methanol-consumer, formate-consumer, or dual C1-consumer guilds. Methane donors comprised Methylococcales with methane-oxidation evidence and four additional non-Methylococcales MAGs (MAG066, MAG249, MAG287 and MAG295) carrying complete PmoABC operons; isolated MmoB hits did not qualify. Methanol-consumer assignments used Mxa-like methanol-dehydrogenase or methanol-oxidation HMM evidence, whereas formate consumers required FDH-O/fdo HMM evidence. MAGs carrying both capacities formed the dual-consumer class. Formaldehyde-processing capacity, identified from Fae, H_4MPT-C_1 or related formaldehyde-processing HMMs, remained a node attribute and did not define a separate exchange edge.

We considered three directed interaction classes: methane-to-methanol, methane-to-formate and methanol-to-formate. Candidate edges required compatible source and recipient metabolic capacities and the presence of both MAGs somewhere within at least one shared habitat. Habitat classes comprised carbonate, plastic, shark egg capsule, ARMS, crab carapace and anemone-associated communities. Same-sample co-occurrence was not required for edge inclusion but contributed to edge weighting. For methanol-to-formate interactions, methane donors qualified as sources only when they

additionally encoded Mxa-like methanol dehydrogenase, FDH-O/Fdo, or formaldehyde-processing capacity.

For every compatible source-target pair within each habitat, we calculated habitat-specific joint relative abundance as the geometric mean of the summed source and target relative abundances and joint raw abundance as the geometric mean of their summed raw abundance values. The interaction score was $\log\left(1+10^6\sqrt{R_s R_t}\right)\times\log\left(1+\sqrt{C_s C_t}\right)\times B$where R_s and R_t denote habitat-summed source and target relative abundances, C_s and C_t denote the corresponding raw abundance sums, and B represents sample overlap. When source and target co-occurred in at least one replicate within the habitat, B equaled 1 plus the fraction of habitat replicates in which both occurred; pairs occurring within the same habitat but never in the same replicate received B = 0.65. Self-edges were excluded. When a source-target-interaction combination occurred in multiple habitats, we retained the habitat with the highest interaction score, using shared-sample number as a secondary ranking criterion.

We evaluated co-abundance as supporting evidence rather than as an edge-inclusion criterion. Global centered-log-ratio (CLR) Spearman correlations supported an edge when rho was at least 0.60, and the FDR-adjusted q value was at most 0.05. For habitats containing at least four samples, we also calculated habitat-specific CLR Spearman correlations and considered rho $\geq$ 0.60 with P $\leq$ 0.05 as local support. We classified edges according to combined global/local CLR support, same-sample occurrence, or shared-habitat occurrence alone.

For visualization in Cytoscape, we reduced the complete interaction network by retaining representative high-scoring edges across interaction classes and individual source and target MAGs, thereby limiting visual complexity while maintaining the overall network structure. Node size represents mean MAG relative abundance, node shape represents C1 guild, and node color represents either dominant habitat or taxonomic order. Edge color indicates the predicted C1 interaction class, and edge width scales with the interaction score. The network represents potential C1 transfer routes supported by genome-encoded metabolic complementarity and ecological co-occurrence rather than direct measurements of metabolite transfer or carbon flux.

In the final network, node size represents mean MAG relative abundance, node shape represents C1 guild, and node color represents either dominant habitat or taxonomic order in the two corresponding panels. Edge color identifies the predicted C1 interaction class and edge width scales with the interaction score. The network represents potential C1 transfer routes supported by genome-encoded

metabolic complementarity and ecological co-occurrence rather than direct measurements of metabolite transfer or carbon flux.

Taxonomic and functional beta-diversity, functional redundancy and core guilds: To compare taxonomic and ecological turnover among biofilms, we constructed three sample-level abundance matrices from the 18 biofilm samples; we excluded the host-associated *Sagartiogeton* sp. anemone from this comparison. The taxonomic matrix contained MAG relative abundances. The ecological-guild matrix summed the relative abundances of MAGs classified in **Supplementary Table S4** as methanotrophs, sulfur-oxidizing autotrophs, other autotrophs, or other community members. The primary-production matrix summed the relative abundances of MAGs assigned to methanotrophy, sulfur-oxidizing autotrophy, or autotrophic carbon fixation. Thus, both functional matrices retained abundance information rather than reducing traits to sample-level presence/absence.

We Hellinger-transformed each abundance matrix by converting each sample to relative abundance and taking the square root of each proportion, and then calculated Bray-Curtis dissimilarities among samples. For **Supplementary Fig. 3**, we represented each biofilm by its mean Bray-Curtis dissimilarity to the other 17 biofilms. We compared the paired taxonomic and functional values with one-sided Wilcoxon signed-rank tests. This sample-level summary avoids treating the 153 pairwise distances as independent observations. For substrate-level PERMANOVA, we used only replicated substrate classes (carbonate, plastic, shark egg capsule and ARMS biofilms), excluded the single crab-carapace sample, and used 9,999 permutations. Mantel and Procrustes permutation tests (9,999 permutations) quantified concordance between taxonomic and functional distance structures. **Supplementary Table S5** reports the resulting summary statistics and tests.

For each curated ecological guild or carbon-fixation pathway, we quantified functional redundancy within each sample using the inverse Simpson effective number. We report the median effective number across samples, together with the number of contributing MAGs and taxonomic orders. We defined a function as core when the MAGs assigned to that function collectively accounted for at least 1% of community relative abundance in each of the 18 biofilm samples.

Statistical analyses and visualization: We performed community analyses on normalized MAG relative abundance profiles. We calculated MAG-level Shannon diversity and tested habitat differences with a Kruskal-Wallis test followed, where appropriate, by pairwise Wilcoxon rank-sum tests with Benjamini-Hochberg correction. We evaluated taxonomic community composition from Hellinger-transformed MAG relative abundance matrices using Bray-Curtis dissimilarity, principal coordinate analysis

(PCoA), PERMANOVA and PERMDISP. We implemented the community analyses in R using vegan, including diversity, vegdist, adonis2, betadisper/permutest, Mantel and Procrustes procedures as applicable; base-R functions provided rank-based tests, and p.adjust (method = “BH”) provided false-discovery-rate correction. We implemented the taxonomic-functional beta-diversity analysis in Python using pandas, NumPy and SciPy, with Bray-Curtis distances calculated using scipy.spatial.distance and paired Wilcoxon signed-rank tests using scipy.stats. Matplotlib generated **Supplementary Fig. 3**. Custom Python scripts generated the abundance-weighted alluvial plot and network input tables, and Cytoscape rendered the final C1 network. We provide the corresponding analysis and figure-generation scripts in the accompanying Markdown file.

References

1. Kieser, S., Brown, J., Zdobnov, E. M., Trajkovski, M. & McCue, L. A. ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. *BMC Bioinformatics* **21**, 257 (2020).

2. Nissen, J. N. *et al.* Improved metagenome binning and assembly using deep variational autoencoders. *Nat. Biotechnol.* **39**, 555–560 (2021).

3. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).

4. Chklovski, A. *et al.* CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat. Methods* **20**, 1203–1212 (2023).

5. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

6. Overbeek, R. *et al.* The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**, D206–D214 (2014).

7. Huerta-Cepas, J. *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).

8. Zhou, Z. *et al.* METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* **10**, 33 (2022).

9. Lee, M. D. & Ponty, Y. GToTree: A user-friendly workflow for phylogenomics.

Bioinformatics **35**, 4162–4164 (2019).

10. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

11. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

12. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

13. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).