bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Temporal emergence of functional dark matter in microbial responses to PFAS revealed by materials-based cultivation

Marta Velaz Martín[1], Kersten S. Rabe[1], Laura Meisch[1], John Vollmers[2], Anne-Kristin Kaster[2], Christof M. Niemeyer[1]*

[1] Karlsruhe Institute of Technology (KIT), Institute for Biological Interfaces 1 (IBG-1), Biomolecular Micro- and Nanostructures, Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Germany.

[2] Karlsruhe Institute of Technology (KIT), Institute for Biological Interfaces 5 (IBG-5), Microbial Genetics & Biotechnology, Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Germany.

Abstract

Microbial responses to xenobiotic compounds are difficult to resolve due to environmental complexity and limited functional annotation. Here, we establish a materials-based cultivation framework using macroporous elastomeric silicone foams (MESIF) to capture microbial adaptation across environmental contexts and timescales. Using glyphosate as a model compound and per- and polyfluoroalkyl substances (PFAS) as a poorly understood class, we show that responses differ depending on the availability of established metabolic pathways. Glyphosate exposure induced rapid, pathway-specific functional enrichment with minimal taxonomic change. In contrast, PFAS exposure did not yield consistent taxonomic or annotation-based signals but instead produced responses that emerged over time and were primarily detectable at protein and genome-resolved levels. LC-MS analyses revealed transformation dynamics, including formation of shorter-chain products. These responses were not explained by known degrader taxa but involved uncharacterized proteins and microbial populations, highlighting the extent of functional “dark matter” in microbial responses to persistent contaminants.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

1. Introduction

Microorganisms play a central role in maintaining environmental balance[1-3]. However, increasing human activity has led to the widespread release of persistent contaminants into natural ecosystems, posing significant challenges for environmental and human health[4]. These xenobiotics disperse into aquatic systems, contaminating freshwater and drinking water sources and ultimately reaching wastewater treatment plants (WWTPs)[5].

Among these, glyphosate is one of the most widely used herbicides worldwide[6]. Beyond its effects on plants, glyphosate has been shown to impact microbial communities, including the gut microbiome[7], as well as higher organisms[8, 9], with documented bioaccumulation and entry into the human body[10-13]. In contrast, per- and polyfluoroalkyl substances (PFAS) represent a more recalcitrant class of contaminants. These synthetic compounds are characterized by highly stable carbon-fluorine bonds[14], resulting in extreme environmental persistence and resistance to degradation[15, 16]. PFAS exhibit strong bioaccumulation potential and have been associated with adverse effects in higher organisms, including endocrine disruption in humans[17, 18].

While microbial processes are known to mediate the transformation of many environmental xenobiotics, mechanistic understanding varies widely across compound classes. Glyphosate degradation proceeds via well-characterized pathways, including cleavage of the C-P bond by the C-P lyase system and oxidation by glyphosate oxidoreductase (gox), yielding aminomethylphosphonic acid (AMPA)[19-21]. These pathways enable microorganisms to utilize glyphosate as a source of phosphorus and nitrogen. In contrast, the physicochemical properties of PFAS pose fundamental challenges to microbial metabolism, and their biological transformation remains poorly understood, limiting current strategies for their remediation[22-24].

A major bottleneck in identifying microorganisms involved in xenobiotic transformation lies in their cultivation. Many environmentally relevant microbes remain uncultured due to unknown physiological requirements, slow growth, or dependence on community interactions[25, 26]. Moreover, environmental systems such as WWTPs are inherently dynamic, with continuous mixing and periodic turnover of influent, which can obscure the detection of selective microbial responses over time. To address these limitations, cultivation-based enrichment platforms have emerged as powerful tools to capture microbial diversity directly from environmental samples[27-29].

One such approach involves macroporous elastomeric silicone foams (MESIF) chips, which enable in situ cultivation under environmentally relevant conditions[30]. Previous work has shown that these platforms enrich distinct microbial communities compared to the native environment and respond to selective pressures such as glyphosate exposure[30, 31]. However, whether such systems can be systematically applied to resolve microbial dynamics over time, particularly under conditions of contrasting environmental stability, and to link specific organisms and functions to xenobiotic exposure remains unclear.

Here, we establish a materials-based cultivation framework to selectively capture and characterize microbial responses to xenobiotic exposure. We use glyphosate as a model compound with well-defined degradation pathways and PFAS as a poorly understood

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

compound class. MESIF chips were fabricated from a macroporous polydimethylsiloxane (PDMS) matrix generated by porogen leaching (Fig. 1a) and assembled into cultivation devices with an integrated reservoir enabling controlled and sustained exposure to xenobiotics (Fig. 1b,c). We implemented two complementary cultivation strategies: short-term in situ incubation in a WWTP to capture rapid responses under dynamic environmental conditions (Fig. 1d), and long-term laboratory mesocosm cultivation to resolve slower, emergent dynamics under more controlled conditions (Fig. 1e). All samples were processed using a unified workflow combining shotgun metagenomic sequencing with community-, protein-, and genome-resolved analyses (Fig. 1f). This integrated approach enables the identification of both rapid and temporally emerging microbial responses to xenobiotic exposure and reveals candidate organisms and functional traits, including previously uncharacterized protein space, associated with their transformation.

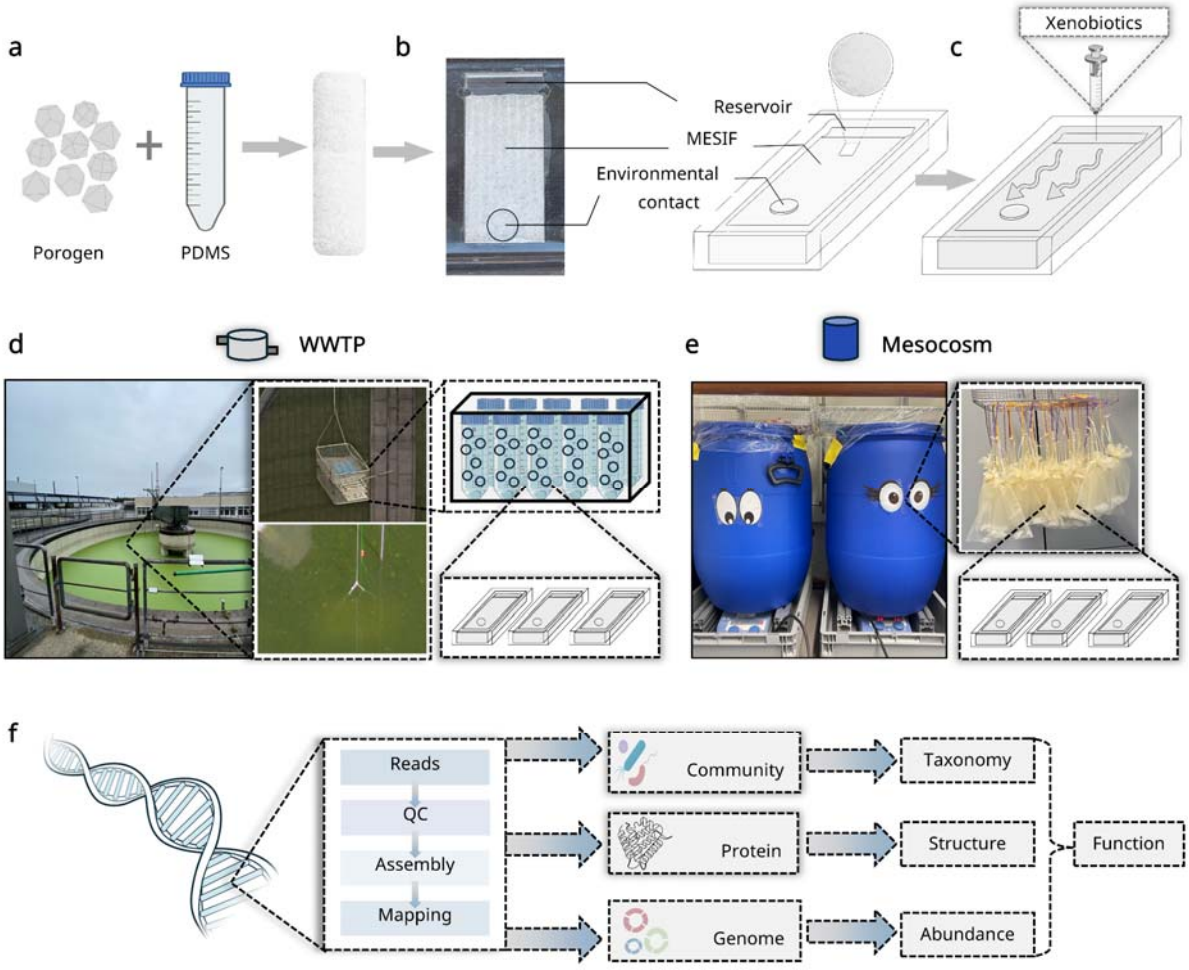


Fig. 1 | MESIF-based cultivation enables controlled and time-resolved enrichment of microbial communities under xenobiotic exposure. (a) MESIFs were fabricated by embedding porogens (salt crystals) into polydimethylsiloxane (PDMS), followed by leaching to generate a macroporous matrix. (b) MESIF chips consist of the porous matrix, a housing, and a perforated lid, allowing environmental exchange while maintaining a defined internal reservoir. (c) The reservoir enables controlled supplementation with basal media and xenobiotics. (d) For short-term in situ cultivation, MESIF chips were deployed in a wastewater treatment plant (WWTP) for 21 days. Chips were placed in perforated Falcon tubes, mounted in cages, and submerged in the treatment tank. (e) For long-term cultivation, MESIF chips were incubated in laboratory mesocosms for up to 100 days, placed in mesh bags and suspended in wastewater-filled tanks, using separate tanks for PFAS (left) and glyphosate (right) exposure (f) Following cultivation, samples were processed by DNA extraction

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

and shotgun metagenomic sequencing, and analyzed at the community, protein, and genome-resolved levels to assess microbial taxonomy, abundance, structural features, and functional potential.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

2. Results

To evaluate MESIF-based cultivation for resolving xenobiotic-associated microbial responses, we designed a multi-factorial framework spanning compound type, medium composition and cultivation timescale. MESIF chips were supplemented with xenobiotics either through the integrated reservoir or by adsorption to the matrix, depending on compound properties, and incubated in defined minimal medium (M9) or wastewater-derived medium (N). Cultivation was performed as short-term in situ incubation in a WWTP and as long-term laboratory mesocosm incubation to capture rapid and temporally evolving responses. We first used glyphosate as a benchmark compound with well-characterized degradation pathways, before extending the framework to PFAS, for which microbial transformation remains poorly understood.

2.1. Platform validation with glyphosate reveals functional adaptation

To evaluate selective enrichment of xenobiotic-responsive microbiomes, we used glyphosate as a model compound. MESIF chips were supplemented with glyphosate or left untreated and incubated in defined minimal medium (M9) or wastewater-derived medium (N). Given the high solubility of glyphosate, compound retention was assessed using the fluorescent tracer sodium fluorescein. Only minimal diffusion into the surrounding medium (~0.5%) was observed, indicating sustained retention within the matrix (Supplementary Fig. S1).

To establish a baseline, we compared microbial communities recovered from MESIF chips without xenobiotic supplementation (M9 and N) to those in the surrounding habitat. MESIF cultivation resulted in pronounced taxonomic and functional shifts, including increased recovery of low-abundance taxa and enhanced genome reconstruction (Extended Data Fig. 1). Taxonomic composition at the phylum level was assessed based on SSU and LSU rRNA genes (Extended Data Fig. 2). Community profiles differed across media, systems and time. In the WWTP, MESIF communities were dominated by *Pseudomonadota* and *Campylobacterota*, whereas habitat communities remained more evenly distributed. In mesocosms, MESIF communities showed stronger temporal turnover, with early dominance of *Pseudomonadota* followed by decline at later time points, particularly in N-derived samples (Extended Data Fig. 2b,c). Together, these results indicate that MESIF enrichment is context dependent and shaped by medium, incubation setting and time[30, 31].

Having established baseline differences between MESIF-derived communities and the surrounding habitat, subsequent analyses focused on xenobiotic-supplemented MESIF chips (M9-G and N-G) relative to their respective controls (M9 and N) within each system (WWTP and mesocosm), thereby minimizing cultivation-derived biases. At the taxonomic level, glyphosate supplementation resulted in only minor changes in overall community composition, with limited shifts at both phylum and genus levels, indicating that taxonomic restructuring is not the primary driver of glyphosate-associated responses (Extended Data Fig. 3, Supplementary Fig. S2).

We next assessed functional responses to glyphosate supplementation. Genes associated with glyphosate degradation, including C-P lyase systems, AMPA-related pathways, phosphonate transporters and sarcosine metabolism, were consistently enriched (Fig. 2a,b), indicating

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

selection on metabolic potential rather than taxonomic identity. Enrichment was more pronounced in M9, whereas N showed a weaker signal, consistent with differences in nutrient availability.

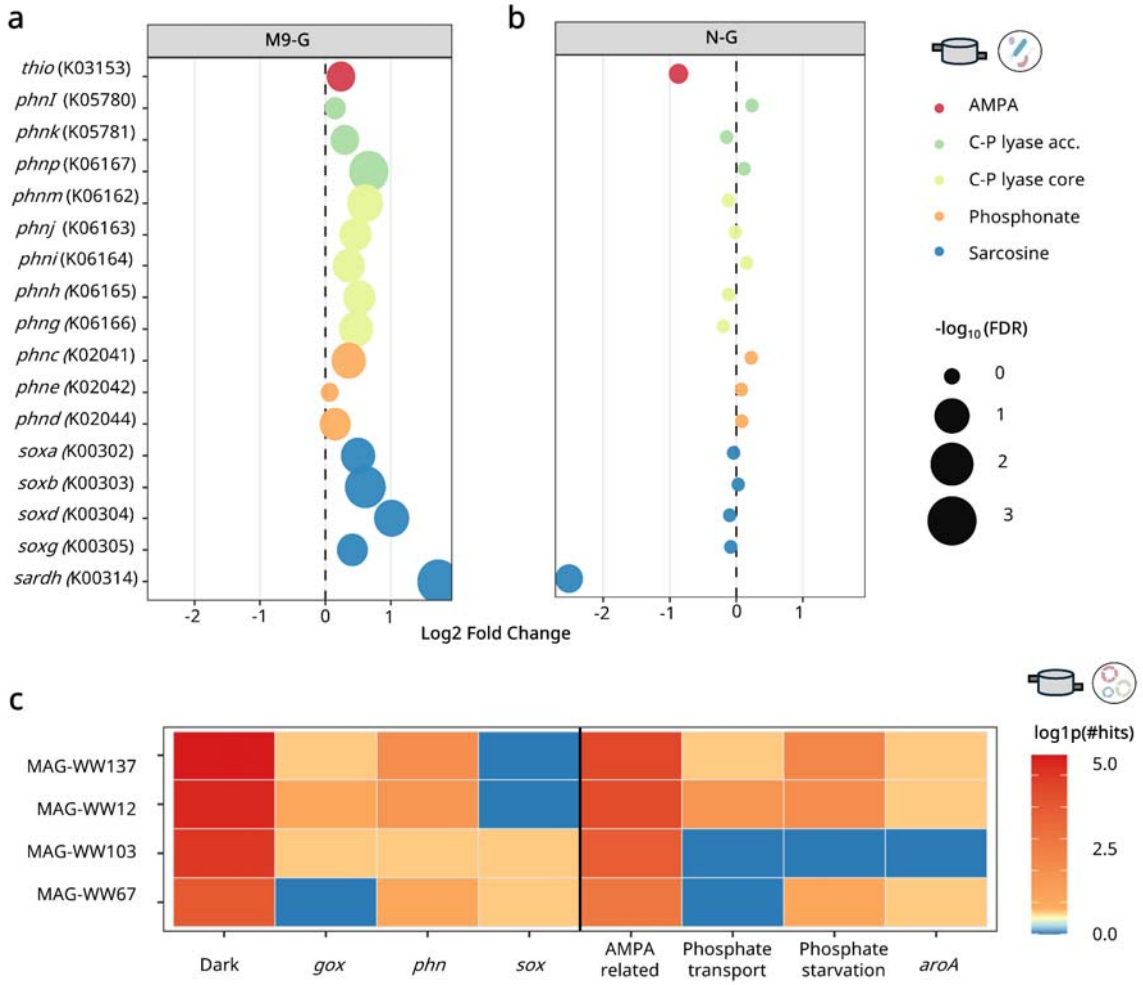


Fig. 2 | Glyphosate supplementation drives pathway-specific functional enrichment during short-term WWTP cultivation. (a,b) Functional response of microbial communities to glyphosate supplementation in (a) M9 and (b) native wastewater-derived medium (N). Log_2 fold changes are shown relative to the corresponding control without glyphosate supplementation. Each point represents a gene associated with key glyphosate-related pathways, including AMPA metabolism (red), the C-P lyase system (core and accessory components; green and yellow), phosphonate transport (orange), and sarcosine metabolism (blue). Point size reflects statistical significance ($-log_{10}$ FDR). (c) Functional profiling of metagenome-assembled genomes (MAGs) uniquely detected in glyphosate-supplemented samples. The heatmap shows the abundance of pathway-related gene hits per MAG, with color intensity indicating the number of detected hits.

PFAM-based analysis further revealed enrichment of mobile genetic element-associated domains under glyphosate supplementation (Extended Data Fig. 4a,b), suggesting elevated stress responses and potential horizontal gene transfer, consistent with the dissemination of glyphosate-degrading genes across microbial communities as previously reported[32].

Next, we performed genome-resolved analyses to identify metagenome-assembled genomes (MAGs) associated with glyphosate supplementation. In total, 174 MAGs were reconstructed

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

across all WWTP samples, including both control and glyphosate-supplemented conditions (Supplementary Table S1). Among these, four were exclusively detected under glyphosate supplementation (Extended Data Fig. 4c) and encoded key phosphonate degradation pathways, including C-P lyase systems and associated transporters, indicating their potential involvement in glyphosate transformation (Fig. 2c).

A similar pattern was observed in the mesocosm, where 388 MAGs were reconstructed (Supplementary Table S2), including 13 uniquely associated with glyphosate supplementation (Supplementary Fig. S3). In contrast to the WWTP, this increase in genome-resolved diversity was not accompanied by consistent functional enrichment (Supplementary Fig. S4). Instead, community variation was primarily driven by temporal dynamics, with prolonged incubation attenuating initial selection effects and reducing differences between supplemented and control communities (Supplementary Figs. S5-S6). These results suggest that responses to compounds with well-defined degradation pathways can be rapidly established but become less distinct over time.

To validate the enrichment approach, we assessed the presence of known glyphosate degraders by mapping reference genomes to metagenomic datasets (Supplementary Table S3). In the WWTP, only a weak signal from a single degrader was detected, whereas in the mesocosm signals were stronger but limited to two of six members (Supplementary Fig. S7a,b). Detection in control samples indicates that these organisms are part of the background community rather than exclusively associated with glyphosate supplementation. A diffusion assay using fluorescein revealed minimal release into the surrounding medium, indicating sustained retention within the MESIF matrix (Supplementary Fig. S1). This suggests localized exposure conditions that may promote distributed metabolic activity across multiple taxa rather than dominance of specialist degraders. Consistent with this, functional enrichment of glyphosate-associated pathways was observed despite limited detection of canonical degraders, highlighting contributions from low-abundance or uncharacterized taxa.

Together, these results indicate that functional responses to glyphosate supplementation are primarily driven by well-characterized metabolic pathways and can emerge without pronounced shifts in community composition, reflecting a distributed metabolic potential across diverse and often low-abundance taxa. In contrast to glyphosate, where degradation mechanisms are well defined and rapidly activated[10, 20, 32-37], PFAS remain poorly understood with respect to their biological transformation[16, 38]. We therefore used PFAS as a second use case to assess whether microbial responses extend beyond known pathway-driven mechanisms and involve previously uncharacterized functional space.

2.2. Protein clustering uncovers PFAS-associated dark functional space

To extend the MESIF platform to xenobiotics lacking well-defined degradation pathways, we investigated PFAS using perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS). Due to their limited solubility, both compounds were introduced via physisorption onto the MESIF matrix. LC-MS measurements showed minimal release, indicating stable retention during cultivation (Supplementary Fig. S8,S9). MESIF chips were cultivated in WWTP and

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

mesocosm systems, followed by taxonomic and functional profiling across control (M9, N) and PFAS-supplemented conditions.

In contrast to glyphosate, PFAS exposure did not result in consistent taxonomic enrichment across conditions, with only minor genus-level shifts observed (Extended Data Figs. 5-6; Supplementary Figs. S10-S11). Community variation was primarily driven by medium rather than PFAS exposure. A limited number of genera showed reproducible increases, including *Pseudomonas* in the WWTP and *Aeromonas* in the mesocosm. Given the association of *Pseudomonas* with xenobiotic transformation[16, 39, 40], this may indicate PFAS-associated responses, although patterns remained taxonomically diffuse. Notably, taxonomic structure increased during prolonged mesocosm incubation but remained insufficient to explain PFAS-associated functional responses.

Functional changes in annotated PFAS-associated categories were weak and inconsistent (Extended Data Fig. 7a,b), indicating that PFAS-associated functional potential is not captured by annotation-based approaches and may instead reside within the “functional dark matter” of the community, i.e. proteins lacking functional annotation[41, 42]. To address this, we applied protein-level clustering across all conditions, including both annotated and unannotated sequences. Clusters present across all conditions were excluded to reduce background redundancy and enrich for condition-associated signals. Uncharacterized proteins from these clusters were selected for downstream analyses.

Protein-level clustering revealed treatment-associated groups, including clusters shared between PFOA- and PFOS-exposed samples in the WWTP (Fig. 3a,b). However, many of these clusters were also detected in control samples, indicating partial enrichment rather than strict PFAS specificity (Fig. 3c,d). Notably, this pattern shifted in the mesocosm, where the proportion of treatment-specific proteins increased compared to the WWTP incubation (Fig. 3e,f; Extended Data Fig. 7c,d), suggesting that prolonged incubation enhances the resolution of PFAS-associated protein modules and enables the emergence of condition-associated functional patterns. Temporal profiling of these proteins further showed that they were predominantly detected in PFAS-exposed samples and became more consistently resolved over time (Supplementary Figs. S12-S13), supporting their classification as condition-associated candidates.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

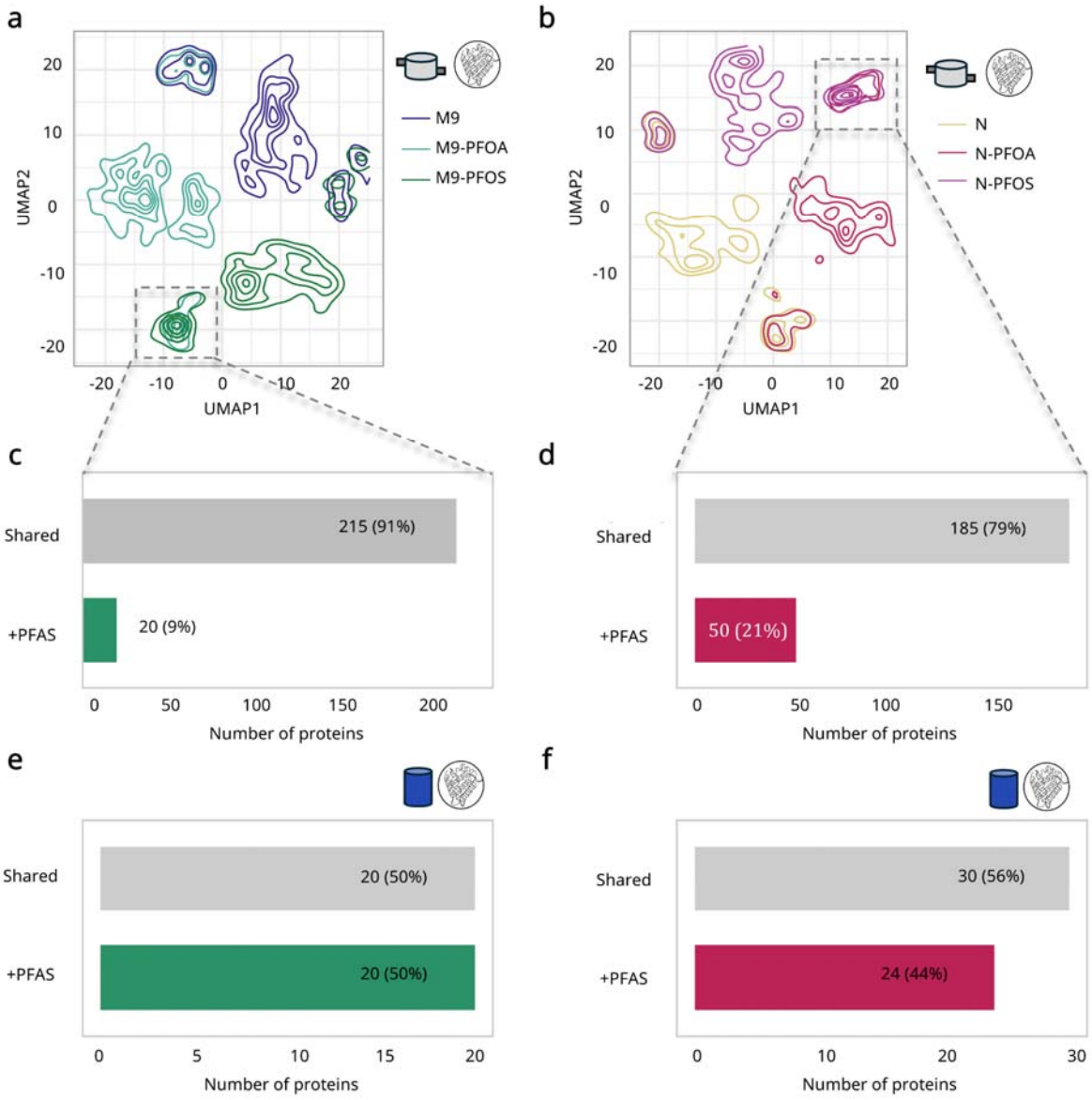


Fig. 3 | Protein-level clustering reveals treatment-associated modules and increased specificity during mesocosm cultivation. (a,b) UMAP projection of proteins from short-term WWTP cultivation after filtering to remove broadly shared proteins. Only proteins present in at most n-1 conditions were retained, enriching for condition-dependent signals. Density contours indicate the distribution across treatments (control, PFOA, PFOS) for (a) M9 and (b) native (N) communities, highlighting both partially shared and treatment-associated regions of functional space. (c,d) Quantification of PFAS-associated proteins in WWTP samples. Bars indicate the number and proportion of proteins detected in both PFOA- and PFOS-treated samples (+PFAS) compared to those also present in controls (Shared) for (c) M9 and (d) N. (e,f) Equivalent analysis for mesocosm samples, showing an increased proportion of treatment-associated proteins compared to WWTP for (e) M9 and (f) N.

To further characterize PFAS-associated uncharacterized proteins, we predicted structures using AlphaFold2 combined with Foldseek-based annotation (Supplementary Tables S4-S7). Across all treatments, at least 50% of candidates remained structurally uncharacterized, supporting the presence of functional dark matter (Extended Data Fig. 7e). Structurally dark proteins, defined here as sequences without detectable Foldseek hits, were preferentially

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

recovered in native (N) medium-derived samples but not in M9 (Supplementary Table S8), indicating that medium composition influences the recovery of uncharacterized proteins. Together, these results suggest that PFAS-associated functional potential may reside within previously uncharacterized protein modules. However, it remains unclear whether these signals reflect biodegradation or tolerance-based responses, and protein-level clustering alone does not resolve which community members encode these functions or are specifically enriched under PFAS exposure.

To address this, we performed genome-resolved analyses of metagenome-assembled genomes (MAGs) uniquely detected in PFAS-treated samples (PFOA, PFOS, or both), assessing their contribution to the identified clusters and their potential involvement in PFAS-associated processes.

2.3. Genome-resolved analysis identifies PFAS-associated populations and functions

Building on the identification of PFAS-associated protein clusters, we performed genome-resolved analyses to identify metagenome-assembled genomes (MAGs) preferentially detected in PFAS-treated samples. In total, 14 such MAGs were identified in the WWTP (Fig. 4a) and 34 in the mesocosm (Supplementary Fig. S14). We functionally annotated these genomes to assess their potential contribution to PFAS-associated processes[22].

PFAS-associated MAGs were more frequently detected at later timepoints in both WWTP and mesocosm cultivation (Fig. 4a; Supplementary Fig. S14), with similar temporal patterns observed across independent systems, suggesting that condition-associated populations become more apparent as incubation progresses, consistent with the temporal emergence of PFAS-associated functional signals. Functional annotation revealed genes linked to processes potentially relevant in PFAS-impacted systems, including dehalogenases, oxidoreductases, transport systems, as well as a substantial fraction of uncharacterized proteins (Fig. 4b; Supplementary Fig. S15).

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

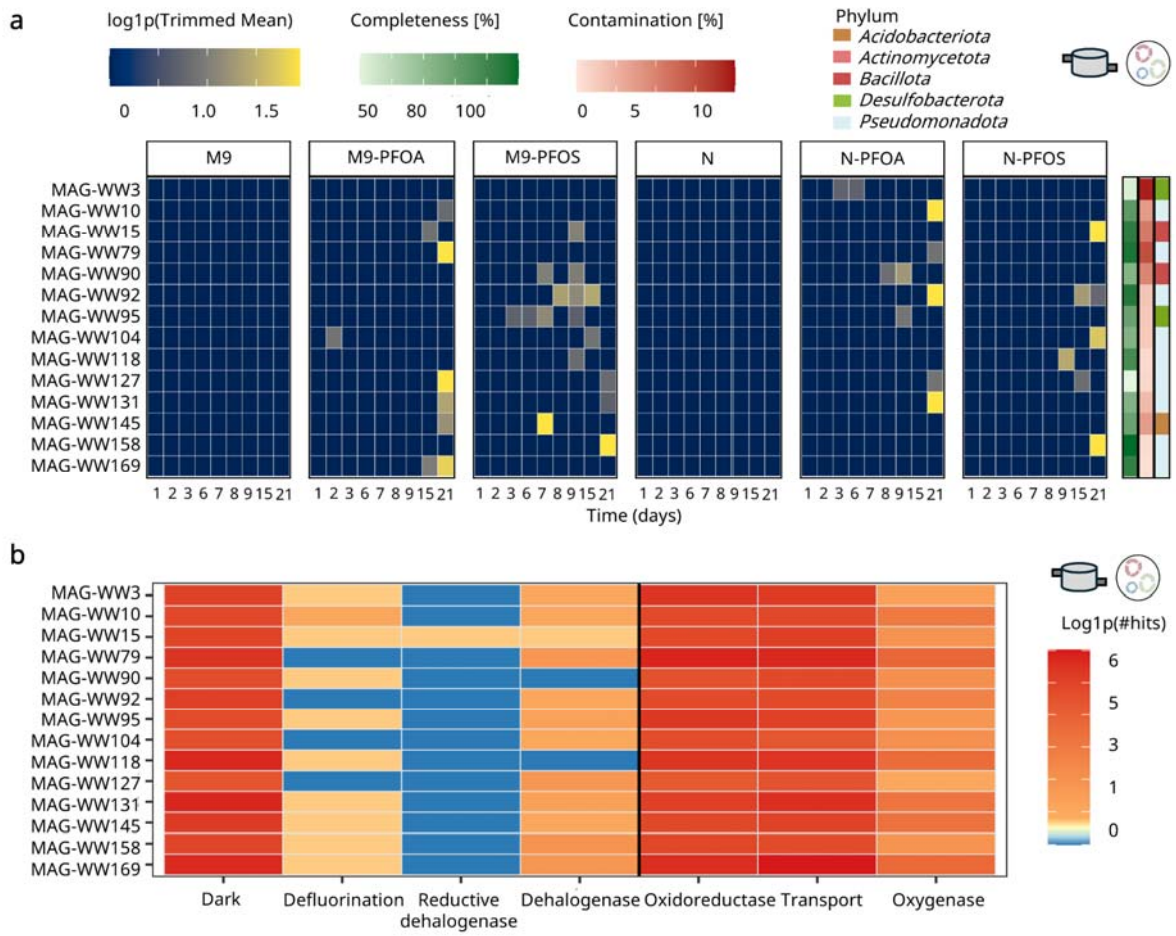


Fig. 4 | Genome-resolved analysis identifies PFAS-associated MAGs and their functional potential. (a) Temporal detection of metagenome-assembled genomes (MAGs) preferentially detected in PFAS-treated samples during short-term WWTP cultivation. Rows represent individual MAGs and columns correspond to sampling timepoints, with color intensity indicating relative abundance. Bars on the right indicate genome completeness and contamination, and phylum-level classification is shown alongside. (b) Functional annotation of PFAS-associated MAGs based on gene hits related to processes potentially relevant in PFAS-impacted systems, including dehalogenation, redox activity, and stress response. The heatmap shows the number of detected genes per functional category.

Notably, one MAG reconstructed from the WWTP encoded a reductive dehalogenase-like protein (Fig. 4b; MAG-WW15). Reductive dehalogenases are key enzymes mediating organohalide respiration[43], catalyzing the reductive cleavage of carbon-halogen bonds under anaerobic conditions. Although PFAS are highly recalcitrant due to the strength of the C-F bond[22], dehalogenase-mediated mechanisms have been proposed as potential biological routes for their transformation[43-47], suggesting a possible role of this MAG in PFAS-associated processes.

Inspection of the genomic neighborhood revealed a conserved locus organization, in which the rdh-like gene is flanked by multiple uncharacterized (dark) proteins, a membrane-associated MBOAT protein, and a downstream chemotaxis-related sensor (Extended Data Fig. 8a). Given recent evidence that PFAS transformation intermediates can be incorporated into bacterial membrane lipids[48], the presence of a membrane-associated enzyme may indicate potential

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

involvement in membrane-related adaptation or interaction processes. Finally, phylogenetic placement of this MAG showed that it clusters within the genus *Fusibacter* (Extended Data Fig. 8b), suggesting that it represents a previously uncharacterized lineage within this group.

Mapping of reference genomes from previously reported PFAS-associated degraders revealed detectable signals in PFAS-exposed samples, with generally higher abundance in the mesocosm compared to WWTP conditions (Supplementary Fig. S16). However, these signals were inconsistent across timepoints and treatments, indicating that known degraders do not dominate PFAS-associated community responses in this system. This observation is consistent with the emergence of distinct PFAS-associated MAGs and protein clusters identified above. This further supports the notion that PFAS-associated functional responses are not primarily driven by previously characterized degrader taxa.

Integration of protein- and genome-level data showed partial overlap between MAG-derived and reference proteins in WWTP samples, whereas mesocosm samples exhibited more distinct clustering, consistent with increased resolution of condition-associated protein modules (Supplementary Fig. S17). Unassigned proteins remained broadly distributed, highlighting functional diversity beyond characterized protein families.

To further assess PFAS-associated processes, we monitored transformation products during mesocosm cultivation using LC-MS (Fig. 5a). PFOA concentrations decreased over time in MESIF samples and were accompanied by the formation of shorter-chain products (C4-C8), whereas PFOS remained largely stable with limited byproduct formation, indicating compound-specific transformation dynamics (Fig. 5b,c).

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

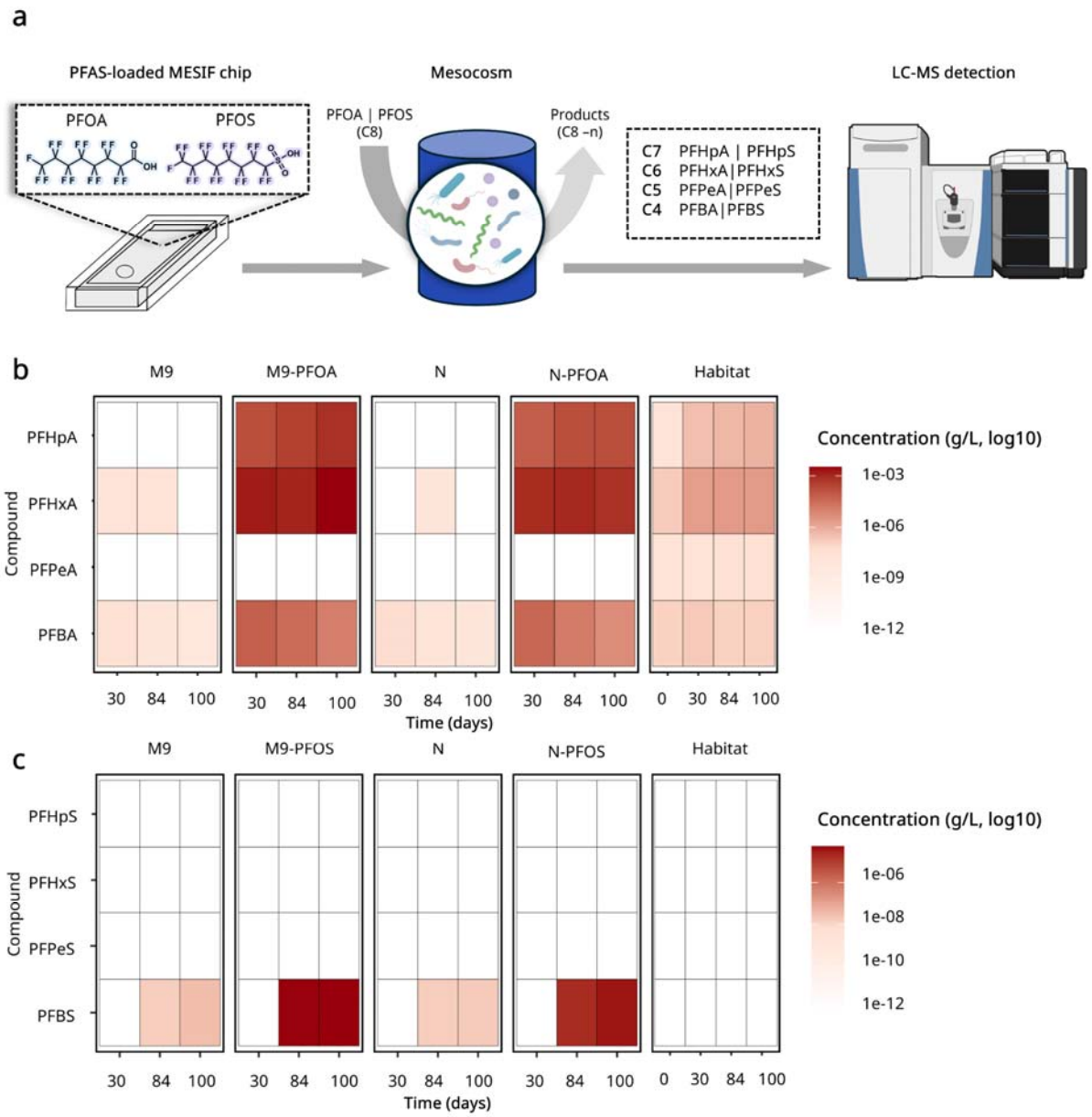


Fig. 5 | PFAS transformation dynamics during mesocosm cultivation in MESIF chips. (a) Schematic overview of the experimental setup. PFAS (PFOA, 18 mM; PFOS, 14.8 mM; initial loading concentration 7.4 g L^{-1}) were physisorbed onto MESIF chips and cultivated in mesocosms, followed by LC-MS monitoring of parent compounds and transformation products (C4-C8). (b,c) Heatmaps of (b) PFOA-derived and (c) PFOS-derived transformation products across time points and conditions. PFOA concentrations decreased over time and were accompanied by the formation of shorter-chain products, whereas PFOS remained largely stable with limited detectable byproducts.

2.4. Cross-system comparison reveals conserved PFAS-associated signatures

Finally, we investigated the presence and potential correlation of PFAS-associated signals between WWTP and mesocosm systems. To this end, we compared the most enriched and depleted genera across both systems relative to their respective controls to identify taxa consistently associated with PFAS exposure.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Overall, taxonomic responses differed substantially between experimental systems (Supplementary Fig. S18). Despite this variability, a subset of genera was reproducibly enriched across PFAS treatments and media, including *Pseudomonas*, a genus previously associated with xenobiotic transformation[49], as well as sulfate-reducing bacteria such as *Desulfovibrio* and *Desulfomicrobium* (Supplementary Fig. S19). These taxa are linked to anaerobic reductive metabolism, including dehalogenation-related processes relevant for the transformation of highly halogenated compounds[45]. Together, these observations indicate a limited but reproducible taxonomic response that becomes more pronounced over time. Consistent with this, the magnitude of enrichment was generally higher in mesocosm samples compared to WWTP-derived samples, suggesting that prolonged incubation enhances the resolution of PFAS-associated taxonomic patterns.

We next examined protein clusters shared across both systems. No consistent clustering patterns were observed, indicating the absence of a conserved global protein-level signature (Extended Data Fig. 9a,b). However, 64 proteins were consistently detected across both systems (Extended Data Fig. 9a,b; black triangles), including 11 sequence-dark proteins lacking functional annotation (Extended Data Fig. 9c; Supplementary Table S10). To further characterize these candidates, we used AlphaFold2-predicted structures in combination with Foldseek-based annotations. To assess their potential for PFAS interaction, all structures were analyzed using pocket prediction (fpocket)[50] followed by docking simulations (gnina)[51]. A control panel of known PFAS-binding and non-binding proteins clearly distinguished positive and negative controls (Supplementary Table S11), supporting the robustness of the approach. Binding affinities across candidates varied (Supplementary Table S12), with five proteins showing scores comparable to positive controls (Supplementary Table S13). Structural analysis of these candidates revealed diverse predicted binding pockets (Supplementary Fig. S20). While these results do not demonstrate catalytic activity, they suggest that these proteins have the capacity to interact with PFAS compounds.

Together, these findings indicate that PFAS-associated functional potential extends beyond annotated pathways and may reside in previously uncharacterized proteins, highlighting candidates for future experimental validation.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

3. Discussion

We established a materials-based cultivation framework to resolve microbial responses to xenobiotic exposure across environmental contexts and timescales. Using glyphosate as a model compound and PFAS as a poorly understood class, we show that MESIF-based cultivation captures distinct modes of microbial adaptation that depend on compound properties and temporal dynamics. The porous matrix enables surface-associated growth under sustained local exposure, facilitating recovery of slow-growing and environmentally adapted microorganisms. Across systems, responses were not consistently reflected at the taxonomic level but instead emerged at functional and genome-resolved levels.

For glyphosate, microbial responses were characterized by rapid and pathway-specific functional enrichment, consistent with the activation of well-established degradation mechanisms. Despite only minor shifts in community composition, genes associated with phosphonate metabolism, including C-P lyase systems and AMPA-related pathways, were consistently enriched, indicating that selection primarily acted on metabolic potential rather than taxonomic identity. These responses were most pronounced under nutrient-limited conditions, where glyphosate served as a key resource, highlighting the importance of environmental context in shaping functional adaptation. In contrast, prolonged incubation in the mesocosm led to a partial attenuation of these signals, suggesting that, for compounds with defined metabolic pathways, selective responses can be rapidly established but may become less distinct as communities equilibrate over time.

In contrast to glyphosate, PFAS exposure did not induce strong or consistent responses at the taxonomic or annotation-based level. Instead, PFAS-associated signals emerged over time and were primarily detectable through protein-level and genome-resolved analyses. Protein clustering revealed condition-associated modules that became more distinct during prolonged mesocosm cultivation, while genome-resolved analyses identified MAGs that increased in prevalence over time. Consistent with these observations, LC-MS analyses indicated compound-specific transformation dynamics, with changes in parent compounds and formation of shorter-chain products. Differences between PFAS types likely reflect variation in physicochemical properties and bioavailability, influencing microbial accessibility and transformation potential.

These findings point to a fundamental difference in how microbial communities respond to xenobiotics depending on the availability of established metabolic pathways. While glyphosate responses were driven by the activation of known and conserved biochemical routes, PFAS-associated responses appear to involve a broader and less well-defined functional repertoire. Notably, known PFAS-associated degrader taxa were detected only at low abundance and did not dominate community responses, indicating that the observed signals are not primarily explained by previously characterized organisms. Instead, both protein-level and genome-resolved analyses consistently pointed to the involvement of previously uncharacterized proteins and microbial populations.

A key insight from this study is the importance of temporal dynamics in resolving microbial responses to recalcitrant compounds. While short-term in situ cultivation in the WWTP captured rapid responses under highly dynamic conditions, longer-term mesocosm cultivation enabled

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

the emergence of more structured and reproducible patterns. This is consistent with the inherent differences between these systems, where continuous mixing and turnover in WWTP environments may obscure selective processes, whereas more stable conditions in mesocosms allow gradual enrichment of responsive populations. Importantly, even under prolonged incubation, taxonomic signals remained only partially informative, reinforcing the notion that functional and genome-resolved analyses are required to capture the full extent of PFAS-associated microbial responses.

The identification of uncharacterized protein clusters and MAG-associated functions highlights the extent of functional “dark matter” involved in PFAS-associated processes. Structural analyses further indicated that a subset of these proteins may possess the capacity to interact with PFAS compounds, although their precise biochemical roles remain to be determined. Together, these findings suggest that microbial responses to PFAS are likely mediated by distributed and previously unrecognized functional traits rather than by a small number of specialized degraders.

Several limitations should be considered. While MESIF-based cultivation enables sustained local exposure to xenobiotics and facilitates the recovery of responsive communities, it represents a semi-controlled system that may not fully capture all environmental processes. In addition, although preliminary analyses suggest temporal changes in PFAS concentrations and potential transformation products, further targeted chemical analyses are required to directly link observed microbial responses to specific transformation pathways. Finally, functional predictions based on sequence clustering and structural modeling provide valuable hypotheses but require experimental validation to confirm biochemical activity.

In summary, MESIF-based cultivation provides a framework to resolve microbial responses to xenobiotic exposure across environmental and temporal gradients. By integrating community, protein-level and genome-resolved analyses, this approach reveals functional adaptation processes that remain inaccessible to annotation-based methods. These findings advance our understanding of microbial interactions with persistent contaminants and highlight the importance of integrating cultivation-based enrichment with multi-level molecular analyses to uncover previously hidden functional potential.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

4. Methods

4.1. MESIF cultivation platform

4.1.1. MESIF chip preparation

Macroporous elastomeric silicone foams (MESIF) were fabricated as previously described.[30] In brief, polydimethylsiloxane (PDMS; SYLGARD® 184) was mixed with curing agent at a 10:1 (w/w) ratio, degassed under vacuum, and cast into polymethylmethacrylate (PMMA) molds containing sieved salt crystals (500-707 µm) that served as porogens. After centrifugation to facilitate porogen embedding, the PDMS matrix was cured at 70 °C. The solidified PDMS-salt composites were removed from the molds and salt was leached by repeated washing in warm water to generate a macroporous structure (Fig. 1a). The resulting foams were dried and cut to the defined dimensions (18 x 10 x 3 mm). PDMS housings (cages and lids) were fabricated separately using PMMA molds and cured at 70 °C. Lids were perforated to allow direct medium contact with the porous matrix. MESIF foams and PDMS housings were subsequently treated with oxygen plasma to activate bonding surfaces and assembled into complete MESIF chips, followed by thermal curing under mechanical pressure to ensure stable sealing (Fig. 1b).

4.1.2. Cultivation media

In this study, two cultivation media were used: a defined mineral medium (M9) and a complex environmental medium, referred to hereafter as native medium (N). M9 consisted of 1× M9 salts supplemented with 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 1 mg l^{-1} biotin, 1 mg l^{-1} thiamine and 1× trace element solution. The M9 salts were prepared from a 10× stock solution (75.2 g l^{-1} $Na_2HPO_4 \cdot 2H_2O$, 30 g l^{-1} KH_2PO_4, 5 g l^{-1} NaCl and 5 g l^{-1} NH_4Cl), and the trace element solution was prepared from a 100× stock solution (5 g l^{-1} EDTA, 0.83 g l^{-1} $FeCl_3 \cdot 6H_2O$, 84 mg l^{-1} $ZnCl_2$, 13 mg l^{-1} $CuCl_2 \cdot 2H_2O$, 10 mg l^{-1} $CoCl_2 \cdot 2H_2O$, 10 mg l^{-1} H_3BO_3 and 1.6 mg l^{-1} $MnCl_2 \cdot 4H_2O$). The native medium consisted of wastewater that was autoclaved at 120 °C for 20 min and subsequently sterile filtered through a 0.2 µm membrane before use. Given its origin, the N medium was expected to contain a higher and more chemically diverse pool of organic carbon than M9, in which carbon availability was limited to the supplied xenobiotic compounds.

4.1.3. Supplementation of xenobiotics and medium

Xenobiotic supplementation was performed using two different strategies depending on the physicochemical properties of each compound, with all compounds delivered via the reservoir (Fig. 1c). The water-soluble herbicide glyphosate (Roundup™, Bayer, German) was directly added to the cultivation medium at a final concentration of 40 g L^{-1} (0.24 M) at the time of inoculation (Supplementary Fig. S22a). This concentration was selected to ensure strong selective pressure and therefore exceeds environmentally relevant levels.[12]

In contrast, per- and polyfluoroalkyl substances (PFAS) were obtained from Merck and introduced via physisorption onto the MESIF matrix due to their limited aqueous solubility. Stock solutions of perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS) (7.4 g L^{-1} in LC/MS-grade methanol) corresponding to 18 mM PFOA and 14.8 mM PFOS were injected into the macroporous MESIF structure (650 µL per chip). Chips were incubated for 72 h at room

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

temperature to allow methanol evaporation and PFAS adsorption onto the PDMS matrix (Supplementary Fig. S22b). To verify PFAS immobilization, a fluidic setup was used consisting of a PDMS housing with inlet and outlet ports and a central cavity containing PFAS-loaded MESIF (Supplementary Fig. S9a). The system was flushed with 1 mL of water at a flow rate of 10 µL min^{-1} to assess the stability of PFAS physisorption under flow conditions (Supplementary Fig. S9b). The effluent was collected in PFAS-free HPLC vials (Agilent, USA), and PFAS concentrations were subsequently quantified by HPLC-MS/MS using an Agilent 1290 Infinity II HPLC system coupled to an Agilent 6470 Triple Quadrupole LC/MS detector. Chromatographic separation was performed on an Agilent EclipsePlus C18 RRHD column (1.8 µm, 2.1 × 50 mm) with a Phenomenex C18 guard column, using HPLC-grade water containing 5 mM ammonium acetate and methanol as mobile phases at a flow rate of 0.5 ml min^{-1}.

For environmental cultivation, MESIF chips were filled with 1 mL of the corresponding cultivation medium and used for enrichment experiments In parallel, water samples from the corresponding system (wastewater treatment plant or mesocosm) were consistently collected as controls (hereafter referred to as “Habitat”) In total, nine different treatment conditions were established (Table 1).

Table 1: Experimental conditions used in this study.

Treatment	Medium	Stressor
M9	M9	-
M9-G		Glyphosate
M9-PFOA		PFOA
M9-PFOS		PFOS
Native	Native	-
N-G		Glyphosate
N-PFOA		PFOA
N-PFOS		PFOS
Habitat	-	-

4.2. Environmental sampling

Two experimental settings were employed: (I) a short-term *in situ* incubation in a wastewater treatment plant (hereafter referred to as “WWTP”), and (II) a long-term laboratory mesocosm incubation (hereafter referred to as “mesocosm”).

4.2.1. Wastewater treatment plant for in situ short-term incubation

MESIF chips were prepared as described above[30]. Chips were loaded with either M9 or N medium, resulting in nine treatment conditions (Table 2). For incubation, chips were placed in perforated Falcon tubes (nine openings in the body and five in the cap) to allow water exchange. The tubes were mounted in a cage and suspended within the chemical wastewater tank WP06 B01.1 at KIT Campus North (49.096683, 8.432755). Cultivation experiments were conducted for 21 days, and samples were collected at 9 time points (days 1, 2, 3, 6, 7, 8, 9, 15, and 21) (Fig. 1d).

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

4.2.2. *Mesocosm system for long-term incubation*

For long-term mesocosm incubation, a laboratory-scale mesocosm system was built allowing continuous cultivation of MESIF chips for up to 100 days. MESIF chips were prepared and supplemented as described above.[30] Mesocosms consisted of 60 L containers filled with 40 L of wastewater collected from the same wastewater treatment tank used for the *in situ* incubation experiments. Containers were maintained at room temperature under continuous stirring (180 rpm) and covered with plastic foil to minimize evaporation.

Two mesocosms were prepared and operated in parallel, one containing glyphosate-treated MESIF chips and the other PFAS-treated chips; both also included untreated control chips. MESIF chips were placed in permeable mesh bags allowing free water exchange with the surrounding wastewater. Each bag contained nine MESIF chips and three glass beads serving as weights to maintain submersion within the tanks. Samples were collected after 1, 2, 4, 7, 14, 21, 30, 50, 84, and 100 days of cultivation (Fig. 1e).

To assess whether water-soluble glyphosate could rapidly diffuse out of the MESIF into the system, we established small-scale mesocosms adapted to a 10:1 scale (6 L instead of 60 L). Accordingly, we placed only 6 MESIF chips instead of 60 in order to track the release of sodium fluorescein as a fluorescent dye tracer (Supplementary Fig. S1). For this purpose, a 1 mM fluorescein solution was prepared and MESIF chips were loaded with 1 mL per chip. Three chips were supplemented with fluorescein, while three served as controls, enabling monitoring of diffusion dynamics. Samples were collected at the same time points as in the mesocosm experiment over the initial 30 days. Fluorescence in liquid samples was measured using a Synergy plate reader (BioTek, Germany) with excitation at 485 nm and emission at 520 nm. In addition, MESIF matrices were imaged using a gel imaging system (Fusion FX, Vilber, France) to visualize fluorescein distribution within the material. Released fractions in the tank were calculated relative to the initially applied amount.

Finally, to assess physicochemical transformation of PFAS, chips from days 50, 84, and 100 were analyzed. For this purpose, MESIF matrices were extracted from their housing and centrifuged twice at 10,000 × g for 10 min to recover the liquid retained within the material. The collected liquid was subsequently filtered through 0.2 µm filters. The resulting samples were quantified by LC-MS regarding the occurrence shorter-chain PFAS as products of degradation. Due to its lower aqueous solubility, PFOS may exhibit stronger retention within the MESIF matrix, potentially affecting recovery.

4.3. Sample processing

4.3.1. *DNA Extraction*

Genomic DNA (gDNA) was extracted using the DNeasy PowerSoil Pro Kit (QIAGEN, Germany) following the manufacturer's instructions with minor modifications. For MESIF samples, chips were retrieved from their housings and cut in half. One half was immediately frozen at −20 °C for storage, while the other half was mechanically squeezed in the CD1 lysis buffer provided with the extraction kit to release matrix-associated biomass. Wastewater-derived samples (Habitat) were processed from 2 mL of liquid sample. The samples were centrifuged for 10 min at 10,000 × g, after which the resulting pellet was resuspended in the CD1 buffer and processed according to the extraction instruction.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Mechanical lysis was performed by bead-beating (two cycles of 1 min at 5.0 m s^{-1}; MP Biomedicals, USA). Genomic DNA was eluted in 50 µL nuclease-free water. DNA concentrations were determined spectrophotometrically (NanoDrop OneC, Thermo Fisher Scientific, Germany) and fluorometrically using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, Germany). All gDNA extracts were stored at −20 °C until further analysis.

4.3.2. Library Preparation and Sequencing

Metagenomic DNA libraries were prepared using the NEBNext® Ultra II FS DNA Library Prep Kit (New England Biolabs, Germany) according to the manufacturer's protocol for DNA inputs ≥100 ng. DNA fragmentation was performed for 10 min, followed by library amplification with 12 PCR cycles during the indexing step. Barcoding was carried out using NEBNext® Multiplex Oligos for Illumina (New England Biolabs, Germany). Library concentrations were quantified using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, Germany), and fragment size distributions were evaluated using the Agilent High Sensitivity DNA Kit on an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). After normalization to 2 nM, libraries were pooled and sequenced on an Illumina NextSeq 1000 platform (Illumina, USA) using a P2 kit (300 cycles).

4.4. Data analysis

4.4.1. Read preprocessing, assembly and coverage estimation

Raw reads were quality-filtered and adapter-trimmed using fastp.[52] Cleaned reads were merged with FLASH[53] and assembled using MEGAHIT[54], generating both individual assemblies and treatment-level co-assemblies. To estimate contig abundance across samples, quality-filtered reads were mapped back to the corresponding assemblies using CoverM[55], and alignment files were processed with samtools.[56] Resulting coverage profiles were used for downstream taxonomic, functional, and genome-resolved analyses.

4.4.2. Community taxonomic and functional profiling

Taxonomic profiling was performed by extracting phylogenetic marker genes from assembled contigs using MDMcleaner[57]. Recovered SSU and LSU rRNA genes were taxonomically classified with SINA against the SILVA database.[58, 59] Abundance-weighted taxonomic profiles were obtained by mapping reads back to the recovered marker genes using CoverM.

For functional profiling, open reading frames were predicted from assembled contigs using Prodigal[60], and functional annotation was performed with eggNOG-mapper[61] in metagenome mode. Functional annotations were combined with contig coverage values to generate abundance-weighted feature tables, including COG categories[62], KEGG orthologs[63-65], KEGG modules, KEGG pathways, and PFAM domains[66], which were used for downstream analyses of community metabolic potential.

4.4.3. Gene-family clustering, module detection and dark protein analysis

To identify shared and treatment-associated gene families, predicted proteins from co-assemblies were clustered into non-redundant families using MMseqs2[67]. Here, and throughout the study, "proteins" refer to predicted open reading frames derived from metagenomic sequence data rather than experimentally validated gene products. Cluster representatives were functionally annotated with eggNOG-mapper, and genes lacking functional annotation were defined as dark proteins. To identify co-occurring gene families, a binary treatment-by-cluster matrix was generated and pairwise Hamming distances were calculated[68]. These

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

relationships were visualized using UMAP, and co-occurrence modules, defined as groups of gene families with similar presence/absence patterns across treatments, were inferred using HDBSCAN.[69, 70] Modules composed exclusively of clusters lacking eggNOG-based functional annotation were classified as sequence-dark modules**.** Treatment-associated and sequence-dark modules were further examined, and representative uncharacterized proteins were structurally modeled with AlphaFold2[71] and compared against reference structure databases using Foldseek[72]. Finally, proteins that were identified as shared between the short-term WWTP incubation and the long-term mesocosm were further analyzed to predict binding pockets using fpocket[50] and subsequently subjected to docking simulations using gnina.[51] To validate and contextualize the docking results, a control framework was established comprising positive controls (*Homo sapiens* serum albumin (HSA), fatty acid-binding protein 1 (FABP1), and a predicted PFOS-binding reductive dehalogenase[44] and negative controls (triosephosphate isomerase, GAPDH, and Protein A).

4.4.4. Genome-resolved analysis and MAG characterization

Genome binning was performed using a multi-binner strategy combining MaxBin2[73], MetaBAT2[74], and CONCOCT[75], followed by refinement with DAS Tool[76]. Binning was applied to both individual assemblies and co-assemblies using coverage-informed approaches. Resulting bins were dereplicated with dRep[77], curated using MDMcleaner[57], and quality-assessed with CheckM2[78]. Taxonomic classification of MAGs was performed with GTDB-Tk[79], and phylogenetic placement was inferred using the GTDB-Tk de novo tree workflow. To assess genome-resolved functional potential, genes were predicted with Bakta[80] and annotated using eggNOG-mapper. Genome abundances were estimated by read mapping with CoverM. In addition, genomes of known xenobiotic degraders with available reference sequences were retrieved from NCBI[81] and mapped against the metagenomic datasets to assess their occurrence across treatments.

4.4.5. Data visualization and statistical analysis

All statistical analyses and data visualization were performed using R (RStudio version 2024.12.1; Posit Software, PBC).

4.5. Data availability

Metagenomic sequencing reads generated in this study have been submitted to RADAR (DOI 10.35097/fau7wstc5sfgb1xe).

5. Acknowledgements

This work was financially supported through the Helmholtz program “Materials Systems Engineering” under the topic “Adaptive and Bioinstructive Materials Systems” (43.33.11), KIT EXU project DigiteLiSE and BMBF project 161L0284A MicroMatrix. The authors thank Simone Weigel and Cornelia Ziegler for their experimental support in sample processing as well as Rafael Peschke and Michael Wagner from Engler-Bunte Institut (EBI), Wasserchemie und Wassertechnologie for conducting the LC-MS measurements.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Author Information

Corresponding Author

Christof M. Niemeyer - Institute for Biological Interfaces 1 (IBG-1), Biomolecular Micro- and Nanostructures, Karlsruhe Institute of Technology (KIT), D-76344 Eggenstein-Leopoldshafen, Germany; https://orcid.org/0000-0002-8837-081X ; E-mail: niemeyer@kit.edu

Authors

Marta Velaz Martín − Institute for Biological Interfaces 1 (IBG-1), Biomolecular Micro- and Nanostructures, Karlsruhe Institute of Technology (KIT), D-76344 Eggenstein-Leopoldshafen, Germany; https://orcid.org/0000-0002-6960-7287; E-mail: marta.martin@kit.edu

Kersten S. Rabe − Institute for Biological Interfaces 1 (IBG1), Biomolecular Micro- and Nanostructures, Karlsruhe Institute of Technology (KIT), D-76344 Eggenstein-Leopoldshafen, Germany; https://orcid.org/0000-0001-7909-8191 ; E-mail: kersten.rabe@kit.edu

Laura Meisch − Institute for Biological Interfaces 1 (IBG-1), Biomolecular Micro- and Nanostructures, Karlsruhe Institute of Technology (KIT), D-76344 Eggenstein-Leopoldshafen, Germany; E-mail: laura.meisch@kit.edu

John Vollmers - Institute for Biological Interfaces 5 (IBG-5), Microbial Genetics and Biotechnology, Karslruhe Institute of Technology (KIT), D-76344 Eggenstein-Leopoldshafen, Germany; https://orcid.org/0000-0001-5980-5178 ; E-mail: john.vollmers@kit.edu

Anne-Kristin Kaster - Institute for Biological Interfaces 5 (IBG-5), Microbial Genetics and Biotechnology, Karslruhe Institute of Technology (KIT), D-76344 Eggenstein-Leopoldshafen, Germany; https://orcid.org/0000-0002-6361-9551 ; E-mail: kaster@kit.edu

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

6. References

[1] M. S. Ayilara, O. O. Babalola, Bioremediation of environmental wastes: the role of microorganisms, *Front. Agron.* **2023**, *Volume 5 - 2023*.

[2] N. Sonune, in *Rhizobiont in Bioremediation of Hazardous Waste* (Eds.: V. Kumar, R. Prasad, M. Kumar), Springer Singapore, Singapore, **2021**, pp. 391-407.

[3] G. Maglione, P. Zinno, A. Tropea, C. U. Mussagy, L. Dufossé, D. Giuffrida, A. Mondello, Microbes' role in environmental pollution and remediation: a bioeconomy focus approach, *AIMS Microbiol* **2024**, *10*, 723-755.

[4] R. Fuller, P. J. Landrigan, K. Balakrishnan, G. Bathan, S. Bose-O'Reilly, M. Brauer, J. Caravanos, T. Chiles, A. Cohen, L. Corra, M. Cropper, G. Ferraro, J. Hanna, D. Hanrahan, H. Hu, D. Hunter, G. Janata, R. Kupka, B. Lanphear, M. Lichtveld, K. Martin, A. Mustapha, E. Sanchez-Triana, K. Sandilya, L. Schaefli, J. Shaw, J. Seddon, W. Suk, M. M. Téllez-Rojo, C. Yan, Pollution and health: a progress update, *The Lancet Planetary Health* **2022**, *6*, e535-e547.

[5] I. Bashir, F. A. Lone, R. A. Bhat, S. A. Mir, Z. A. Dar, S. A. Dar, Concerns and Threats of Contamination on Aquatic Ecosystems.

[6] S. O. Duke, S. B. Powles, Glyphosate: a once-in-a-century herbicide, *Pest Manage. Sci.* **2008**, *64*, 319-325.

[7] P. Puigbò, L. I. Leino, M. J. Rainio, K. Saikkonen, I. Saloniemi, M. Helander, Does Glyphosate Affect the Human Microbiota?, *Life (Basel, Switz.)* **2022**, *12*, 707.

[8] M. E. Richmond, Glyphosate: A review of its global use, environmental impact, and potential health effects on humans and other species, *J. Environ. Stud. Sci.* **2018**, *8*, 416-434.

[9] F. Maggi, D. la Cecilia, F. H. M. Tang, A. McBratney, The global environmental hazard of glyphosate use, *Science of The Total Environment* **2020**, *717*, 137167.

[10] J. Fan, G. Yang, H. Zhao, G. Shi, Y. Geng, T. Hou, K. Tao, Isolation, identification and characterization of a glyphosate-degrading bacterium, *Bacillus cereus* CB4, from soil, *J. Gen. Appl. Microbiol.* **2012**, *58*, 263-271.

[11] J. V. Tarazona, D. Court-Marques, M. Tiramani, H. Reich, R. Pfeil, F. Istace, F. Crivellente, Glyphosate toxicity and carcinogenicity: a review of the scientific basis of the European Union assessment and its differences with IARC, *Arch. Toxicol.* **2017**, *91*, 2723-2743.

[12] K. Gandhi, S. Khan, M. Patrikar, A. Markad, N. Kumar, A. Choudhari, P. Sagar, S. Indurkar, Exposure risk and environmental impacts of glyphosate: Highlights on the toxicity of herbicide co-formulants, *Environ. Chall.* **2021**, *4*, 100149.

[13] F. S. Galli, M. Mollari, V. Tassinari, C. Alimonti, A. Ubaldi, C. Cuva, D. Marcoccia, Overview of human health effects related to glyphosate exposure, *Front Toxicol* **2024**, *6*, 1474792.

[14] D. O'Hagan, Understanding organofluorine chemistry. An introduction to the C-F bond, *Chem. Soc. Rev.* **2008**, *37*, 308-319.

[15] R. B. Stockbridge, L. P. Wackett, The link between ancient microbial fluoride resistance mechanisms and bioengineering organofluorine degradation or synthesis, *Nat Commun* **2024**, *15*, 4593.

[16] L. P. Wackett, Nothing lasts forever: understanding microbial biodegradation of polyfluorinated compounds and perfluorinated alkyl substances, *Microb. Biotechnol.* **2022**, *15*, 773-792.

[17] H. Pezeshki, S. Rajabi, M. Hashemi, S. Moradalizadeh, H. Nasab, Per- and poly-fluoroalkyl substances as forever chemicals in drinking water: Unraveling the nexus with obesity and endocrine disruption - A mini review, *Heliyon* **2025**, *11*, e42782.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

[18] F. Pérez, M. Nadal, A. Navarro-Ortega, F. Fàbrega, J. L. Domingo, D. Barceló, M. Farré, Accumulation of perfluoroalkyl substances in human tissues*, Environ. Int.* **2013**, *59*, 354-362.

[19] L. P. Wackett, S. L. Shames, C. P. Venditti, C. T. Walsh, Bacterial carbon-phosphorus lyase: products, rates, and regulation of phosphonic and phosphinic acid metabolism*, J. Bacteriol.* **1987**, *169*, 710-717.

[20] S. Singh, V. Kumar, J. P. K. Gill, S. Datta, S. Singh, V. Dhaka, D. Kapoor, A. B. Wani, D. S. Dhanjal, M. Kumar, S. L. Harikumar, J. Singh, Herbicide Glyphosate: Toxicity and Microbial Degradation*, Int. J. Environ. Res. Public Health* **2020**, *17*, 1-18.

[21] X. Huang, J. He, X. Yan, Q. Hong, K. Chen, Q. He, L. Zhang, X. Liu, S. Chuang, S. Li, J. Jiang, Microbial catabolism of chemical herbicides: Microbial resources, metabolic pathways and catabolic genes*, Pestic Biochem Physiol* **2017**, *143*, 272-297.

[22] L. P. Wackett, Why Is the Biodegradation of Polyfluorinated Compounds So Rare?*, mSphere* **2021**, *6*, 10.1128/msphere.00721-00721.

[23] S. Nasrollahpour, S. Purewal, R. K. Das, S. K. Brar, Biodegradation of per- and polyfluoroalkyl substances: mechanisms, challenges, and emerging strategies for sustainable remediation*, Environ. Sci.: Water Res. Technol.* **2026**, *12*, 397-420.

[24] S. I. Probst, F. D. Felder, V. Poltorak, R. Mewalal, I. K. Blaby, S. L. Robinson, Enzymatic carbon-fluorine bond cleavage by human gut microbes*, Proc. Natl. Acad. Sci. U. S. A.* **2025**, *122*, e2504122122.

[25] S. R. Vartoukian, R. M. Palmer, W. G. Wade, Strategies for culture of ‘unculturable’ bacteria*, FEMS Microbiol. Lett.* **2010**, *309*, 1-7.

[26] W. H. Lewis, G. Tahon, P. Geesink, D. Z. Sousa, T. J. G. Ettema, Innovations to culturing the uncultured microbial majority*, Nat. Rev. Microbiol.* **2021**, *19*, 225-240.

[27] B. Berdy, A. L. Spoering, L. L. Ling, S. S. Epstein, In situ cultivation of previously uncultivable microorganisms using the ichip*, Nat. Protoc.* **2017**, *12*, 2232-2242.

[28] D. Nichols, N. Cahoon, E. M. Trakhtenberg, L. Pham, A. Mehta, A. Belanger, T. Kanigan, K. Lewis, S. S. Epstein, Use of ichip for high-throughput in situ cultivation of "uncultivable" microbial species*, Appl. Environ. Microbiol.* **2010**, *76*, 2445-2450.

[29] D. Jung, E. Y. Seo, J. S. Owen, Y. Aoi, S. Yong, E. V. Lavrentyeva, T. S. Ahn, Application of the filter plate microbial trap (FPMT), for cultivating thermophilic bacteria from thermal springs in Barguzin area, eastern Baikal, Russia*, Biosci. Biotechnol. Biochem.* **2018**, *82*, 1624-1632.

[30] A. E. Zoheir, L. Meisch, M. V. Martin, C. Bickmann, A. Kiselev, F. Lenk, A. K. Kaster, K. S. Rabe, C. M. Niemeyer, Macroporous Silicone Chips for Decoding Microbial Dark Matter in Environmental Microbiomes*, ACS Appl Mater Interfaces* **2022**, *14*, 49592-49603.

[31] A. Itani, M. Velaz Martín, L. Meisch, P. Lemke, T. Scharnweber, I. M. Khattab, K. S. Rabe, C. M. Niemeyer, Materials-Based spatiotemporal analysis of microbial responses to glyphosate in Winogradsky columns*, Methods* **2026**, *245*, 1-12.

[32] K. S. Souza, M. R. F. da Silva, M. A. Candido, H. T. S. Lins, G. de Lima Torres, K. C. da Silva Felix, K. C. C. Silva, R. M. N. Filho, R. Bhadouria, S. Tripathi, R. Singh, M. D. V. Santos, I. P. S. Silva, A. V. de Barros, L. C. A. de Araújo, F. Motteran, M. B. M. de Oliveira, Biodegradation Potential of Glyphosate by Bacteria: A Systematic Review on Metabolic Mechanisms and Application Strategies*, Agronomy* **2025**, *15*, 1247.

[33] V. Korkmaz, N. Yildirim, G. O. Erguven, B. Durmus, Y. Nuhoglu, The bioremediation of glyphosate in soil media by some newly isolated bacteria: The COD, TOC removal efficiency and mortality assessment for Daphnia magna*, Environ. Technol. Innov.* **2021**, *22*.

[34] T. Z. Esikova, T. O. Anokhina, N. E. Suzina, T. V. Shushkova, Y. Wu, I. P. Solyanikova, Characterization of a New Pseudomonas Putida Strain Ch2, a Degrader of Toxic

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Anthropogenic Compounds Epsilon-Caprolactam and Glyphosate, *Microorganisms* **2023**, *11*, 650.

[35] R. E. Dick, J. P. Quinn, Control of glyphosate uptake and metabolism in Pseudomonas sp. 4ASW, *FEMS Microbiol. Lett.* **1995**, *134*, 177-182.

[36] H. W. Talbot, L. M. Johnson, D. M. Munnecke, Glyphosate utilization byPseudomonas sp. andAlcaligenes sp. isolated from environmental sources, *Curr. Microbiol.* **1984**, *10*, 255-259.

[37] S. Singh, V. Kumar, J. Singh, Kinetic study of the biodegradation of glyphosate by indigenous soil bacterial isolates in presence of humic acid, Fe(III) and Cu(II) ions, *J. Environ. Chem. Eng.* **2019**, *7*, 103098.

[38] L. P. Wackett, S. L. Robinson, A prescription for engineering PFAS biodegradation, *Biochemical Journal* **2024**, *481*, 1757-1770.

[39] S. Chetverikov, G. Hkudaygulov, D. Sharipov, S. Starikov, D. Chetverikova, Biodegradation Potential of C7-C10 Perfluorocarboxylic Acids and Data from the Genome of a New Strain of Pseudomonas mosselii 5(3), *Toxics* **2023**, *11*, 1001.

[40] L. Yi, C. Tang, Q. Peng, Q. Peng, L. Chai, Draft Genome Sequence of Perfluorooctane Acid-Degrading Bacterium Pseudomonas parafulva YAB-1, *Genome Announc* **2015**, *3*.

[41] Z. Ardern, S. Chakraborty, F. Lenk, A.-K. Kaster, Elucidating the functional roles of prokaryotic proteins using big data and artificial intelligence, *FEMS Microbiol. Rev.* **2023**, *47*.

[42] P. Escudeiro, C. S. Henry, R. P. M. Dias, Functional characterization of prokaryotic dark matter: the road so far and what lies ahead, *Curr. Res. Microb. Sci.* **2022**, *3*, 100159.

[43] M. Bommer, C. Kunze, J. Fesseler, T. Schubert, G. Diekert, H. Dobbek, Structural basis for organohalide respiration, *Science* **2014**, *346*, 455-458.

[44] H.-B. Guo, V. A. Varaljay, G. Kedziora, K. Taylor, S. Farajollahi, N. Lombardo, E. Harper, C. Hung, M. Gross, A. Perminov, P. Dennis, N. Kelley-Loughnane, R. Berry, Accurate prediction by AlphaFold2 for ligand binding in a reductive dehalogenase and implications for PFAS (per- and polyfluoroalkyl substance) biodegradation, *Sci. Rep.* **2023**, *13*, 4082.

[45] M. Fincker, A. M. Spormann, Biochemistry of Catabolic Reductive Dehalogenation, *Annu. Rev. Biochem* **2017**, *86*, 357-386.

[46] P. R. Jaffé, S. Huang, J. Park, M. Ruiz-Urigüen, W. Shuai, M. Sima, in *Methods in Enzymology, Vol. 696* (Ed.: R. B. Stockbridge), Academic Press, **2024**, pp. 287-320.

[47] M. Hu, C. Scott, Toward the development of a molecular toolkit for the microbial remediation of per-and polyfluoroalkyl substances, *Appl. Environ. Microbiol.* **2024**, *90*, e00157-00124.

[48] Y. Xie, G. Chen, M. J. Keller, B. M. Wong, D. Ramirez, C. M. Swift, D. L. Carper, P. E. Abraham, A. J. Stai, R. L. Hettich, A. L. May, S. R. Campagna, J. A. LaFond, J. L. Guelfo, A. Thapalia, T. A. Key, F. E. Löffler, Bacteria covalently incorporate polyfluoroalkyl carboxylates into membrane lipids, *Nat. Microbiol.* **2026**, *11*, 929-939.

[49] L. P. Wackett, Pseudomonas: versatile biocatalysts for PFAS, *Environ. Microbiol.* **2022**, *24*, 2882-2889.

[50] V. Le Guilloux, P. Schmidtke, P. Tuffery, Fpocket: an open source platform for ligand pocket detection, *BMC Bioinformatics* **2009**, *10*, 168.

[51] A. T. McNutt, Y. Li, R. Meli, R. Aggarwal, D. R. Koes, GNINA 1.3: the next increment in molecular docking with deep learning, *J. Cheminform.* **2025**, *17*, 28.

[52] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, *Bioinformatics* **2018**, *34*, i884-i890.

[53] T. Magoč, S. L. Salzberg, FLASH: fast length adjustment of short reads to improve genome assemblies, *Bioinformatics* **2011**, *27*, 2957-2963.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

[54] D. Li, C.-M. Liu, R. Luo, K. Sadakane, T.-W. Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph, *Bioinformatics* **2015**, *31*, 1674-1676.

[55] R. J. P. N. Samuel T. N. Aroney, Jakob N. Nissen, Antonio Pedro Camargo, Gene W. Tyson, Ben J. Woodcroft, CoverM: Read alignment statistics for metagenomics, *arXiv* **2025**, *2501.11217*.

[56] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, S. Genome Project Data Processing, The Sequence Alignment/Map format and SAMtools, *Bioinformatics* **2009**, *25*, 2078-2079.

[57] J. Vollmers, S. Wiegand, F. Lenk, A.-K. Kaster, How clear is our current view on microbial dark matter? (Re-)assessing public MAG & SAG datasets with MDMcleaner, *Nucleic Acids Res.* **2022**, *50*, e76-e76.

[58] C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F. O. Glockner, The SILVA ribosomal RNA gene database project: improved data processin g and web-based tools, *Nucleic Acids Res.* **2013**, *41*, D590-596.

[59] E. Pruesse, C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, F. O. Glockner, SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB, *Nucleic Acids Res.* **2007**, *35*, 7188-7196.

[60] D. Hyatt, G. L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer, L. J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification, *BMC Bioinformatics* **2010**, *11*, 119.

[61] C. P. Cantalapiedra, A. Hernández-Plaza, I. Letunic, P. Bork, J. Huerta-Cepas, eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale, *Mol. Biol. Evol.* **2021**, *38*, 5825-5829.

[62] M. Y. Galperin, R. Vera Alvarez, S. Karamycheva, K. S. Makarova, Yuri I. Wolf, D. Landsman, E. V. Koonin, COG database update 2024, *Nucleic Acids Res.* **2024**, *53*, D356-D363.

[63] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation, *Nucleic Acids Res.* **2015**, *44*, D457-D462.

[64] M. Kanehisa, S. Goto, S. Kawashima, A. Nakaya, The KEGG databases at GenomeNet, *Nucleic Acids Res.* **2002**, *30*, 42-46.

[65] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res.* **2000**, *28*, 27-30.

[66] J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, Gustavo A. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn, A. Bateman, Pfam: The protein families database in 2021, *Nucleic Acids Res.* **2020**, *49*, D412-D419.

[67] M. Hauser, M. Steinegger, J. Söding, MMseqs software suite for fast and deep clustering and searching of large protein sequence sets, *Bioinformatics* **2016**, *32*, 1323-1330.

[68] S. Z. Li, A. Jain, *Hamming Distance*, Springer US, Boston, MA, **2009**.

[69] L. McInnes, J. Healy, N. Saul, L. Großberger, UMAP: Uniform Manifold Approximation and Projection, *J. Open Source Softw.* **2018**, *3*.

[70] L. McInnes, J. Healy, S. Astels, hdbscan: Hierarchical density based clustering, *J. Open Source Softw.* **2017**, *2*.

[71] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature* **2021**, *596*, 583-589.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

[72] M. van Kempen, S. S. Kim, C. Tumescheit, M. Mirdita, J. Lee, C. L. M. Gilchrist, J. Söding, M. Steinegger, Fast and accurate protein structure search with Foldseek*, Nat. Biotechnol.* **2024**, *42*, 243-246.

[73] Y.-W. Wu, B. A. Simmons, S. W. Singer, MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets*, Bioinformatics* **2016**, *32*, 605-607.

[74] D. D. Kang, F. Li, E. Kirton, A. Thomas, R. Egan, H. An, Z. Wang, MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies*, PeerJ* **2019**, *7*, e7359.

[75] J. Alneberg, B. S. Bjarnason, I. de Bruijn, M. Schirmer, J. Quick, U. Z. Ijaz, L. Lahti, N. J. Loman, A. F. Andersson, C. Quince, Binning metagenomic contigs by coverage and composition*, Nat. Methods* **2014**, *11*, 1144-1146.

[76] C. M. K. Sieber, A. J. Probst, A. Sharrar, B. C. Thomas, M. Hess, S. G. Tringe, J. F. Banfield, Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy*, Nat. Microbiol.* **2018**, *3*, 836-843.

[77] M. R. Olm, C. T. Brown, B. Brooks, J. F. Banfield, dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication*, Isme j* **2017**, *11*, 2864-2868.

[78] A. Chklovski, D. H. Parks, B. J. Woodcroft, G. W. Tyson, CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning*, Nat. Methods* **2023**, *20*, 1203-1212.

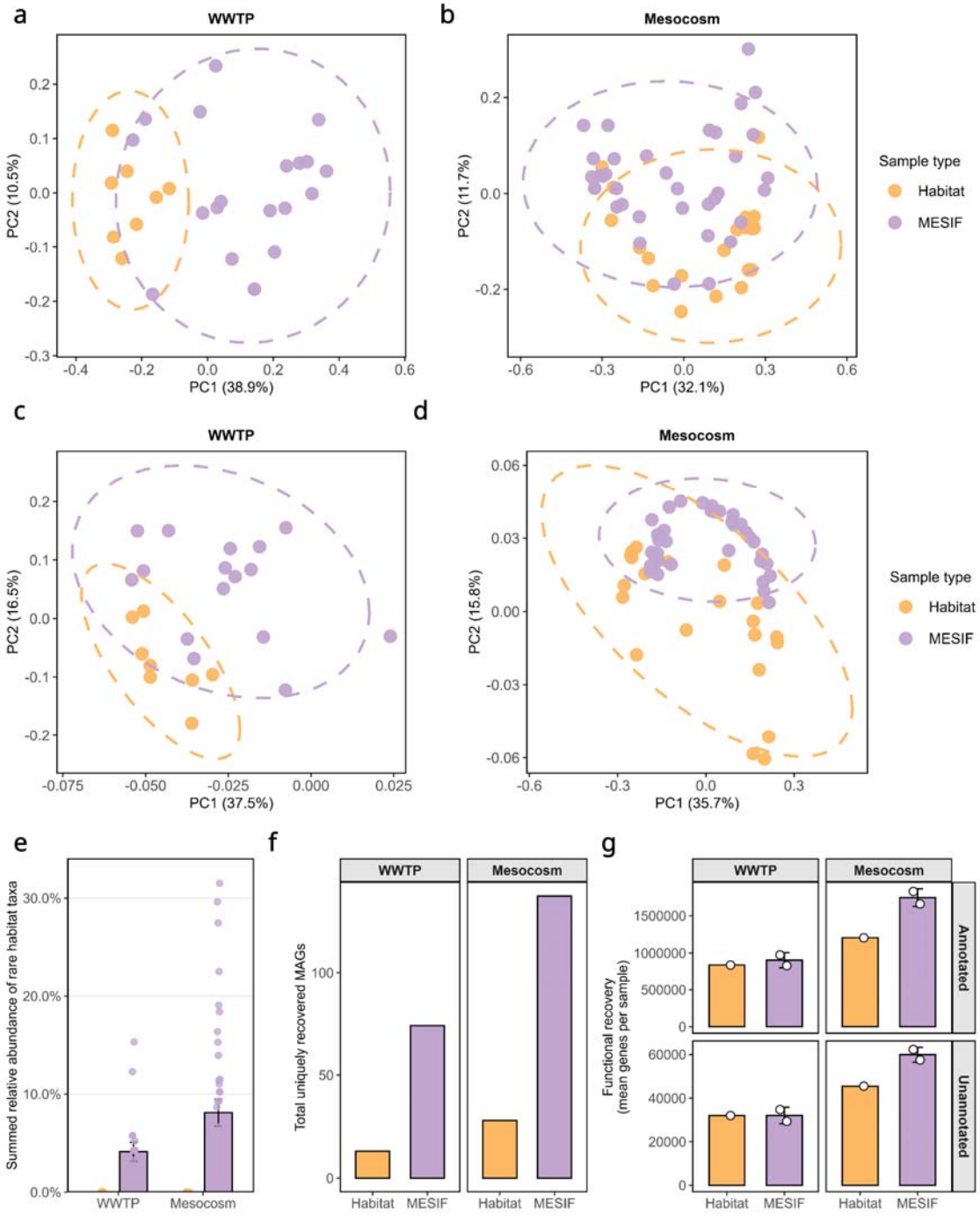
[79] P.-A. Chaumeil, A. J. Mussig, P. Hugenholtz, D. H. Parks, GTDB-Tk v2: memory friendly classification with the genome taxonomy database*, Bioinformatics* **2022**, *38*, 5315-5316.

[80] O. Schwengers, L. Jelonek, M. A. Dieckmann, S. Beyvers, J. Blom, A. Goesmann, Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification*, Microb Genom* **2021**, *7*.

[81] N. A. O'Leary, E. Cox, J. B. Holmes, W. R. Anderson, R. Falk, V. Hem, M. T. N. Tsuchiya, G. D. Schuler, X. Zhang, J. Torcivia, A. Ketter, L. Breen, J. Cothran, H. Bajwa, J. Tinne, P. A. Meric, W. Hlavina, V. A. Schneider, Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets*, Sci. Data* **2024**, *11*, 732.

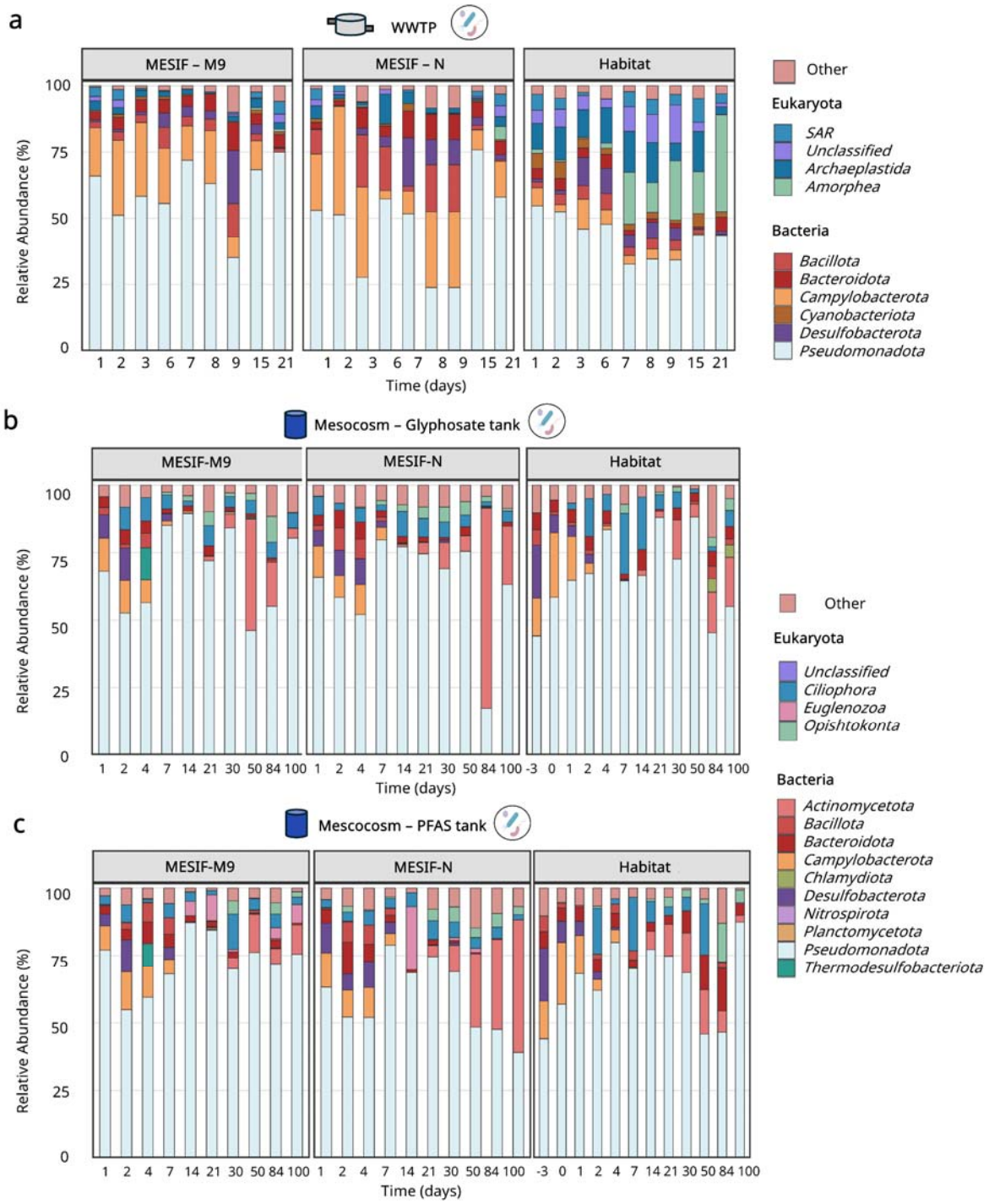
bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Extended Data Figures:



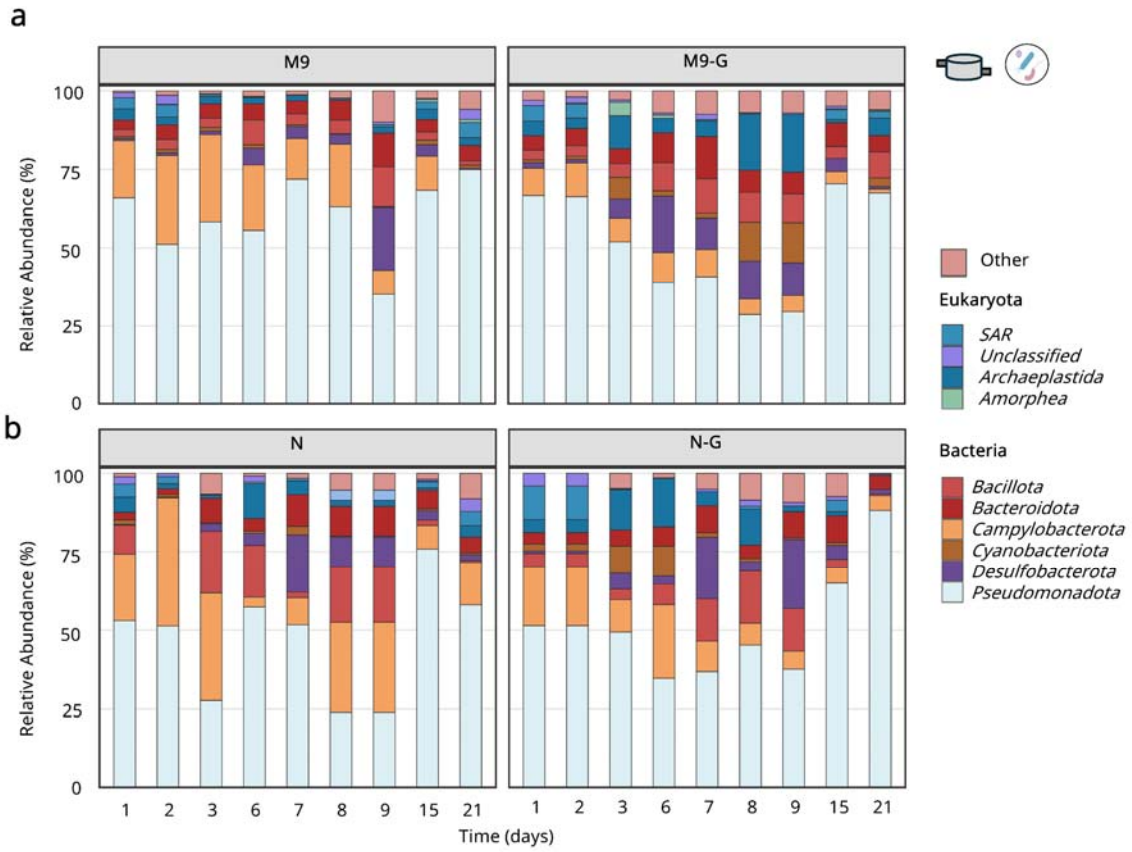
Extended Data Fig. 1 | MESIF cultivation drives taxonomic, functional and genome-level divergence from habitat communities. (a,b) Taxonomy-based community dissimilarity between MESIF-associated and habitat-derived microbiomes assessed using Bray-Curtis distances for (a) WWTP and (b) mesocosm samples. (c,d) Functional dissimilarity based on predicted gene content for (c) WWTP and (d) mesocosm samples. (e) Enrichment of low-abundance taxa in MESIF samples relative to habitat communities, indicating selective recruitment of rare populations. (f) Number of reconstructed metagenome-assembled genomes (MAGs) across sample types, reflecting enhanced genome recovery in MESIF compared to habitat conditions. (g) Number of annotated (top) and unannoted genes (bottom) across systems. Center values represent means and error bars indicate standard deviation.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



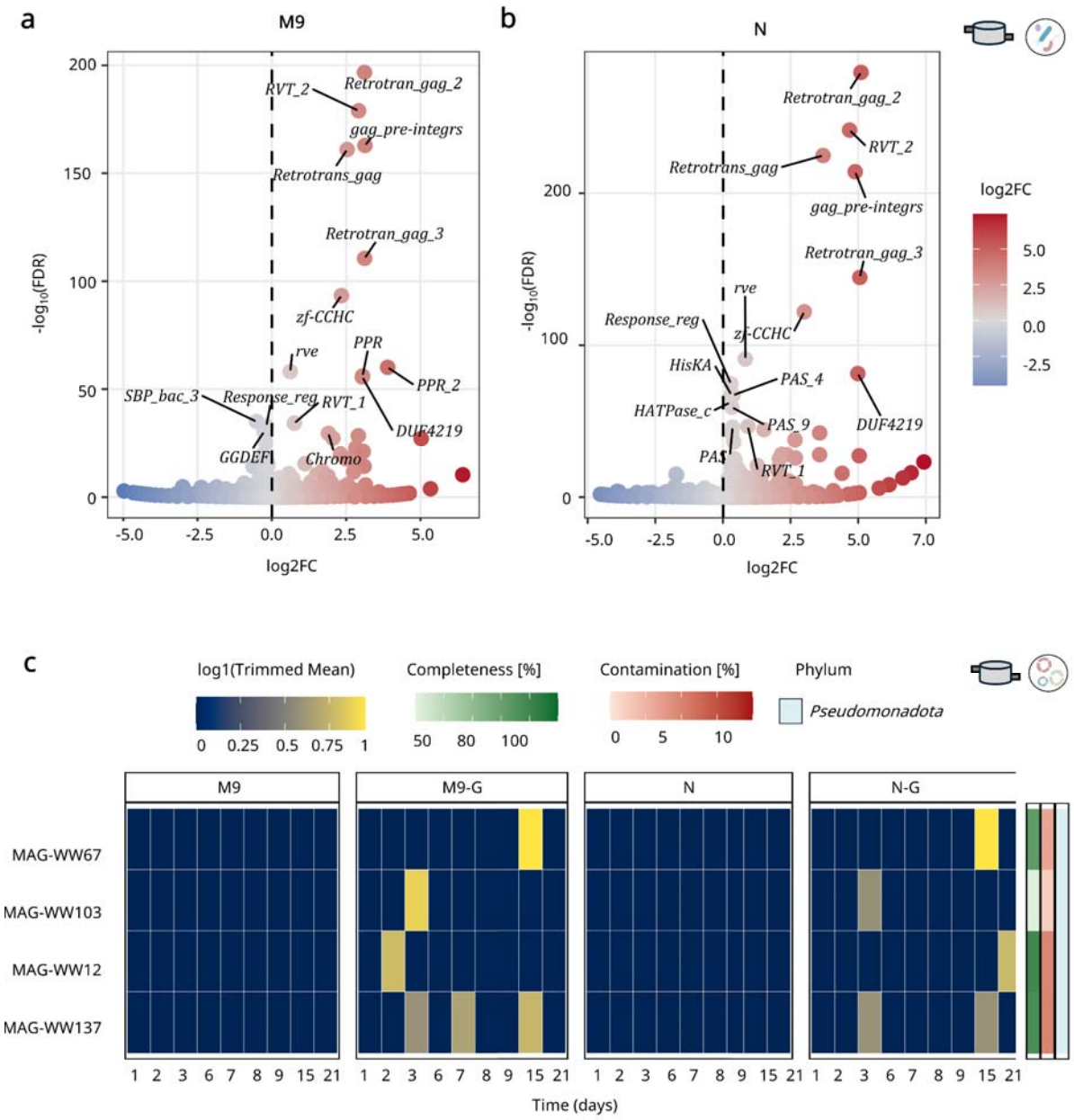
Extended Data Fig. 2 | MESIF cultivation restructures microbial community composition in a context-dependent manner across systems and time. (a) Short-term in situ cultivation in a wastewater treatment plant (WWTP). Long-term mesocosm cultivation in the (b) glyphosate-derived or (c) PFAS-derived system, shown here without xenobiotic supplementation to assess system-specific baseline community trajectories; panels (b) and (c) therefore represent independent mesocosm systems and highlight baseline dynamics irrespective of xenobiotic exposure. Stacked bar plots show the relative abundance of microbial phyla over time for control MESIF chips (without xenobiotic supplementation) containing M9 (left) or native wastewater-derived medium (N, middle), compared to the surrounding wastewater (Habitat, right). Each bar represents a sampling time point, and colors indicate the relative abundance of individual phyla. In mesocosm Habitat samples, time point −3 corresponds to the initial WWTP inoculum used to seed the mesocosms, and time point 0 marks the start of sampling following MESIF deployment.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



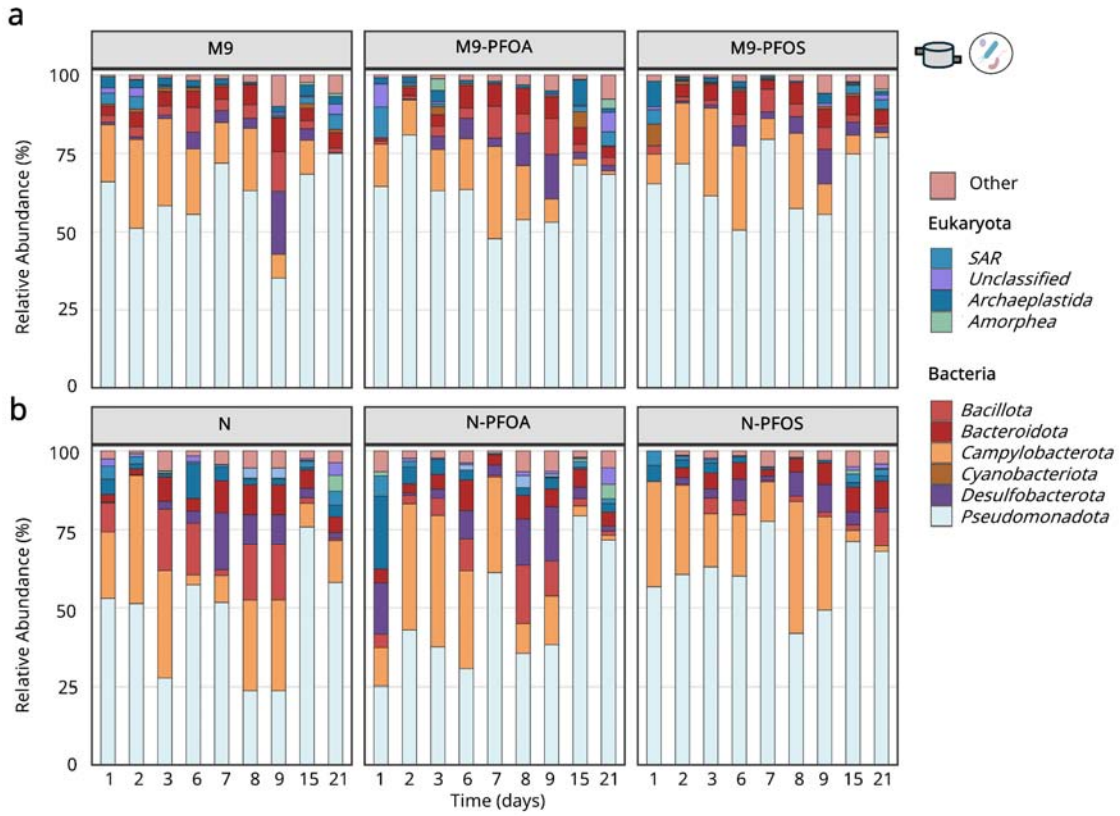
Extended Data Fig. 3 | Glyphosate exposure does not induce pronounced taxonomic shifts at the phylum level during short-term WWTP cultivation. Stacked bar plots show the relative abundance of microbial phyla over time for (a) M9 medium comparing control (M9) and glyphosate-supplemented conditions (M9-G), and (b) native wastewater-derived medium comparing control (N) and glyphosate-supplemented conditions (N-G). Each bar represents a sampling time point (days 1-21), and colors indicate the relative abundance of individual phyla. Across both media, community composition remained largely stable at the phylum level, with only minor variation between control and glyphosate-treated samples.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



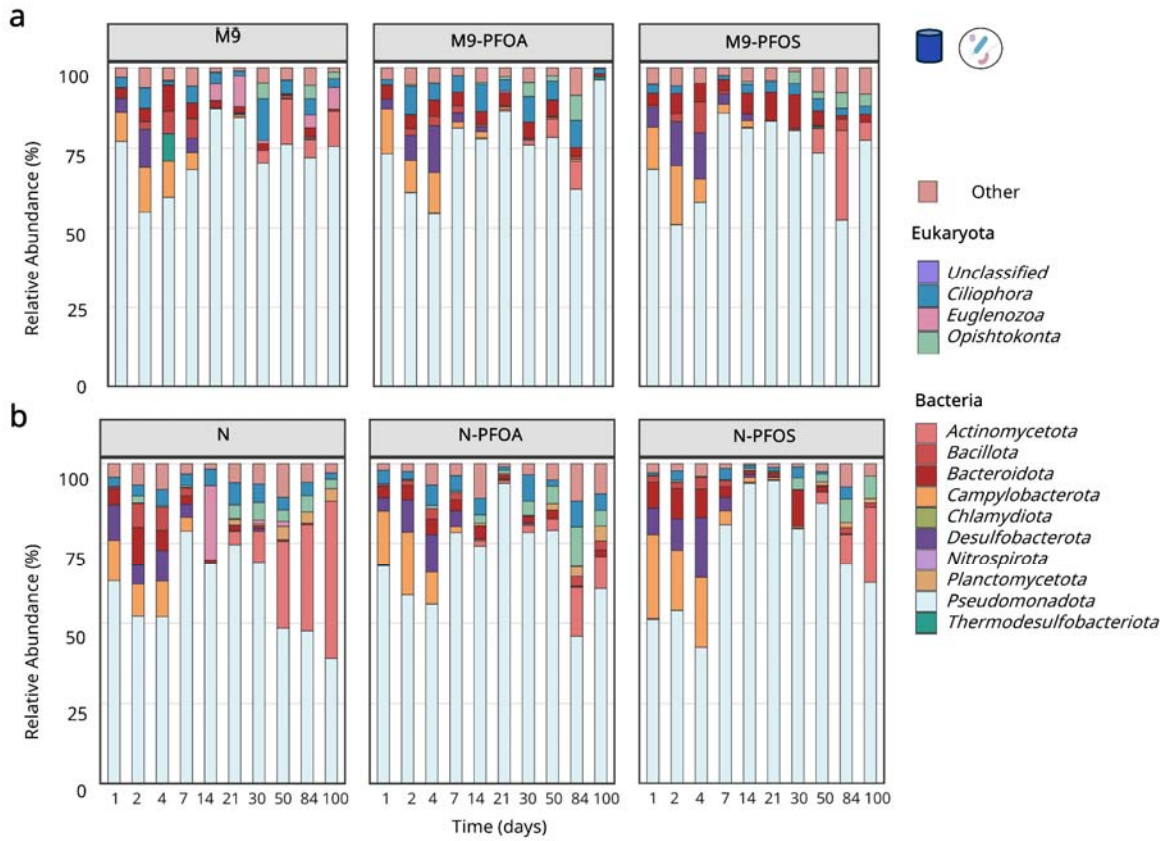
Extended Data Fig. 4 | PFAM domain enrichment and genome-resolved dynamics under glyphosate supplementation during short-term WWTP cultivation. (a,b) Volcano plots showing PFAM domain enrichment in glyphosate-supplemented samples relative to controls for (a) M9 and (b) native wastewater-derived medium (N). Each point represents a PFAM domain, with the x axis indicating $\log_2$ fold change (glyphosate-supplemented vs control) and the y axis representing statistical significance ($-\log_{10}$ FDR). Positive values indicate enrichment under glyphosate supplementation, whereas negative values indicate depletion. (c) Genome-resolved analysis of metagenome-assembled genomes (MAGs) uniquely detected in glyphosate-supplemented samples. Heatmaps show the temporal abundance of MAGs across sampling time points, with rows representing individual MAGs and columns representing sampling time points. Values indicate normalized abundance across control (M9, N) and glyphosate-supplemented conditions (M9-G, N-G). Side panels indicate genome completeness, contamination and phylum-level classification.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



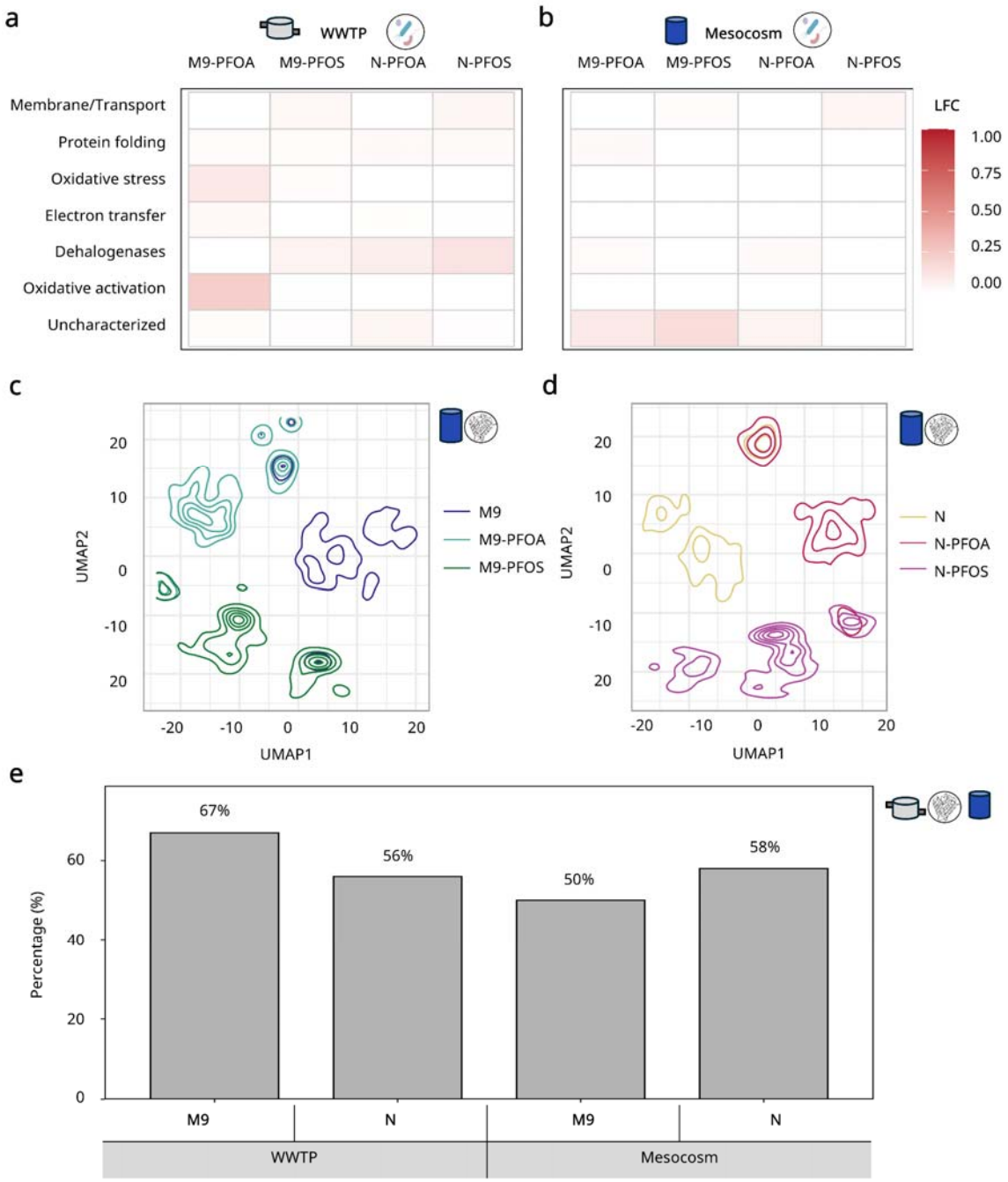
Extended Data Fig. 5 | Phylum-level community composition under PFAS exposure during short-term WWTP cultivation. Stacked bar plots showing the temporal distribution of microbial phyla across sampling timepoints for control (left), PFOA-treated (middle), and PFOS-treated conditions (right) in (a) M9 and (b) native (N) medium-supplemented MESIF chips. Each bar represents a sampling timepoint (days 1-21), and colors indicate the relative abundance of individual phyla. Across both media, phylum-level community composition remained largely similar between control and PFAS-exposed samples.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



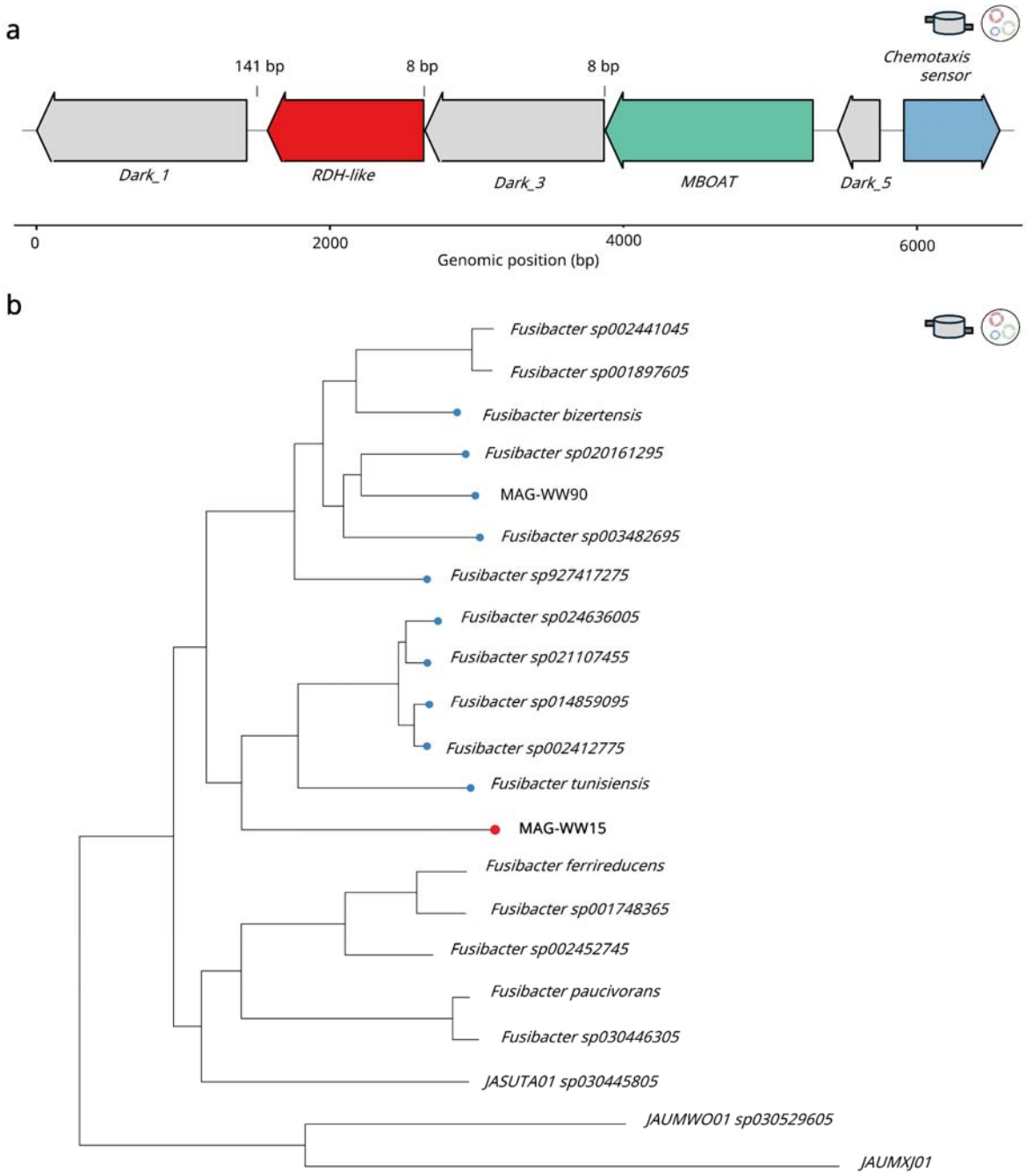
Extended Data Fig. 6 | Phylum-level community composition under PFAS exposure during long-term mesocosm cultivation. Stacked bar plots showing the temporal distribution of microbial phyla across sampling timepoints for control (left), PFOA-treated (middle), and PFOS-treated conditions (right) in (a) M9 and (b) native (N) medium-supplemented MESIF chips. Each bar represents a sampling timepoint (days 1-100), and colors indicate the relative abundance of individual phyla. Across both media, phylum-level community composition remained broadly comparable between control and PFAS-exposed samples over time. Although gradual shifts in relative abundance were observed, these were not consistently associated with PFAS treatment.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



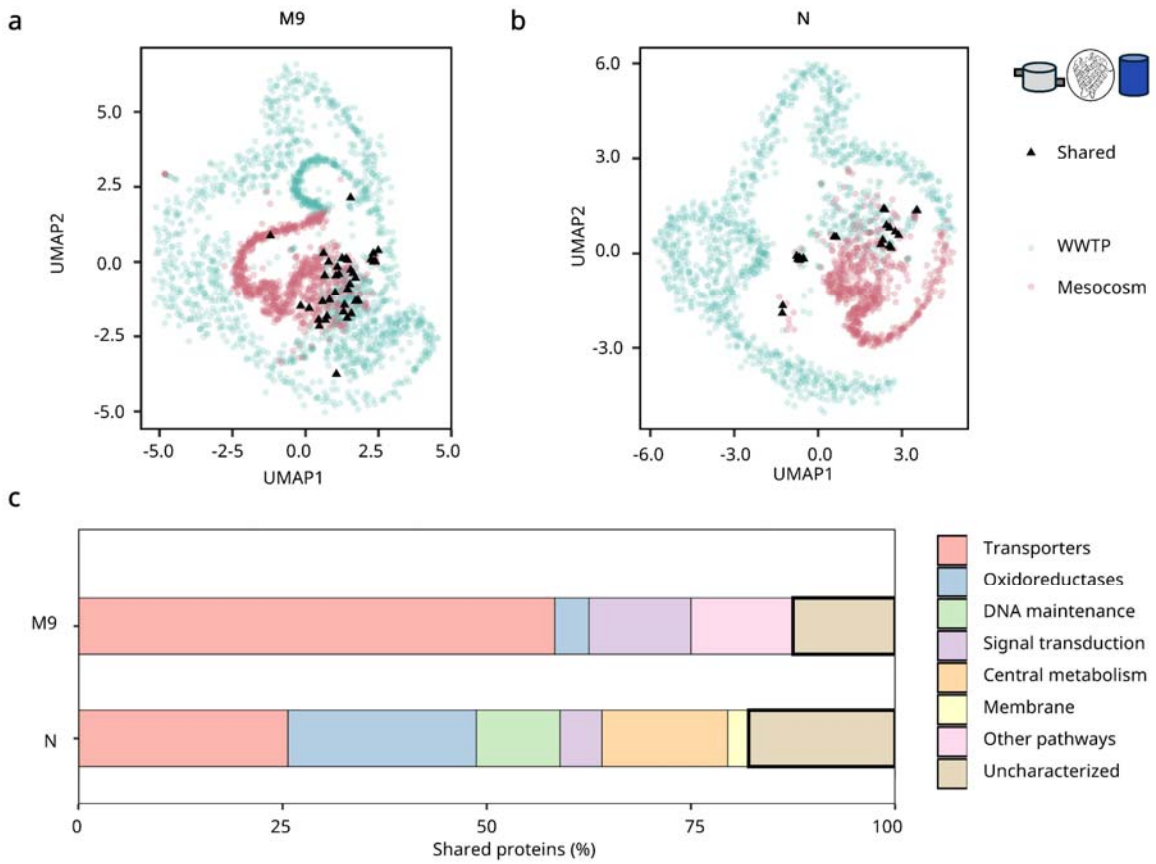
Extended Data Fig. 7 | Functional and protein-level responses to PFAS exposure across cultivation systems. (a,b) Functional enrichment of PFAS-associated categories across cultivation systems. Heatmaps show differential abundance (log_2 fold change relative to controls) of selected functional categories potentially linked to PFAS-related processes in (a) short-term WWTP cultivation and (b) long-term mesocosm cultivation. Columns represent treatment conditions and rows correspond to functional categories derived from eggNOG-based annotations. Color intensity reflects enrichment relative to the corresponding controls (M9 or N). Across both systems, most categories remained close to baseline levels, indicating limited functional differentiation at the level of annotated pathways. (c,d) Protein-level clustering of mesocosm samples. UMAP-based clustering of proteins after filtering to reduce redundancy, retaining only proteins present in at most n-1 conditions and excluding those shared across all treatments. Density contours illustrate the distribution of protein clusters across control, PFOA-, and PFOS-treated samples for (c) M9 and (d) N communities, highlighting both shared and condition-specific regions of functional space. (e) Structural annotation coverage of PFAS-associated proteins. Bar plot showing the proportion of PFAS-associated proteins that remained uncharacterized after structure-based annotation across experimental systems (WWTP and mesocosm) and media (M9 and N). Despite structural modeling, a substantial fraction of proteins (>50%) remained unannotated, indicating that many sequence-level dark proteins also lack structural annotation.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



Extended Data Fig. 8 | Genomic context and phylogenetic placement of a PFAS-associated MAG encoding a reductive dehalogenase-like protein. (a) Genomic organization of the locus containing a reductive dehalogenase-like (rdh-like) gene (red). The rdh-like gene is flanked by multiple uncharacterized (dark) proteins (grey), a membrane-associated MBOAT protein (green), and a downstream chemotaxis-related sensor (blue), suggesting a functionally linked genomic region potentially involved in environmental sensing or response. (b) Phylogenetic placement of the corresponding MAG (MAG-WW15) based on conserved marker genes. The genome clusters within the genus *Fusibacter*, indicating that it likely represents a previously uncharacterized lineage within this group.

bioRxiv preprint doi: https://doi.org/10.64898/2026.05.27.728134; this version posted May 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



Extended Data Fig. 9 | Cross-system comparison reveals limited overlap of PFAS-associated protein space and enrichment of uncharacterized functions. (a,b) UMAP projection of proteins detected across short-term (WWTP) and long-term (mesocosm) cultivation for (a) M9 and (b) native (N) communities. Points represent individual proteins, colored by origin, with proteins detected in both systems highlighted. Limited overlap indicates the absence of a conserved global protein-level signature across systems. (c) Functional classification of proteins detected in both systems, showing a substantial fraction of uncharacterized ("dark") proteins.