nature communications

Article

https://doi.org/10.1038/s41467-026-73084-2

Single nucleotide variants drive evolutionary phage-host arms race in anaerobic carbon dioxide-converting microbiome

Received: 18 June 2025

Accepted: 29 April 2026

Published online: 28 May 2026

Check for updates

G. Ghiotto[1], G. Zampieri[1], E. Orellana[1], A. Chatzis[2], P. G. Kougias[2], A. Camargo[3], S. Roux[3], S. Campanaro[1] ✉, N. C. Kyrpides[3,4] & L. Treu[1,4]

Microbial bioconversions are shaped by environmental perturbations and the adaptation of resident microbiomes. Prokaryotes coexist with bacteriophages, yet their coevolutionary trajectories remain underexplored. Here, we investigate the effects of a cultivation vessel leak on an anaerobic consortium performing carbon dioxide reduction. Using time-series shotgun metagenomic sequencing, we reconstruct microbial and viral genomes to track community shifts. We further apply single-nucleotide variant profiling and CRISPR array analysis to monitor viral microdiversity and host defense mechanisms. After bioaugmentation restores bioconversion efficiency, the consortium undergoes pronounced restructuring, with new dominant taxa emerging from the rare biosphere. We identify patterns consistent with phage predation selectively removing certain species, while others exhibit resilience to infection. This shift aligns with a widespread viral outbreak and a transient increased frequency of single nucleotide variants in bacterial CRISPR–Cas defense genes. Expansion of CRISPR spacers further supports that CRISPR-mediated processes influence microbial resilience. Concurrently, phages infecting resilient hosts exhibited adaptive evolution, marked by high genetic heterogeneity. Selective pressure varies across their genomes, targeting infectivity genes and protospacer-adjacent motifs. These findings highlight a dynamic evolutionary arms race driven by the selection of beneficial genetic variants, providing a mechanistic framework for multi-omics investigations, and informing biotechnological applications, including phage-based microbiome manipulation.

Microbial carbon capture and utilization systems represent a promising avenue for enhancing carbon recovery and mitigating environmental pollution, two key goals for sustainable development[1]. A pivotal branch of biotechnological strategy in this context is carbon dioxide (CO_2) conversion via biomethanation, wherein methanogenic archaea reduce CO_2 with hydrogen (H_2) to produce methane (CH_4) under anoxic conditions[2]. Recent advances in metagenomics and reduced sequencing costs have facilitated both small- and large-scale profiling of carbon-fixing microbiota across diverse environments[3–5]. The recovery of metagenome-assembled genomes (MAGs), coupled with transcriptomics and biochemical measurements, has further advanced community-scale modeling approaches such as flux balance analysis (FBA). These simulations facilitate the identification of core metabolic functions and potential auxotrophies shaped by cross-feeding

[1]Department of Biology, University of Padua, Padova, Italy. [2]Soil and Water Resources Institute, Hellenic Agricultural Organisation Dimitra, Thermi, Thessaloniki, Greece. [3]Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [4]These authors contributed equally: N. C. Kyrpides, L. Treu. ✉e-mail: stefano.campanaro@unipd.it

https://doi.org/10.1038/s41467-026-73084-2

interactions and substrate-driven competition[6,7]. Among these interactions, syntrophy between archaea and bacteria optimize community growth, shifting reaction equilibrium towards CH_4 production ($CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$).

Biomethanation performance and stability are tightly linked to the microbiome composition and the characteristics of the process setup[8]. Engineered bioreactors, serving as "ecosystems in a vessel", allow controlled perturbations to support process optimization. Modulating environmental variables such as temperature, pressure, and gas composition directly influence molecular processes, CH_4 production, and CO_2 fixation. For instance, increasing the flow rate of CO_2 and H_2, thereby reducing the gas retention time (GRT), can boost

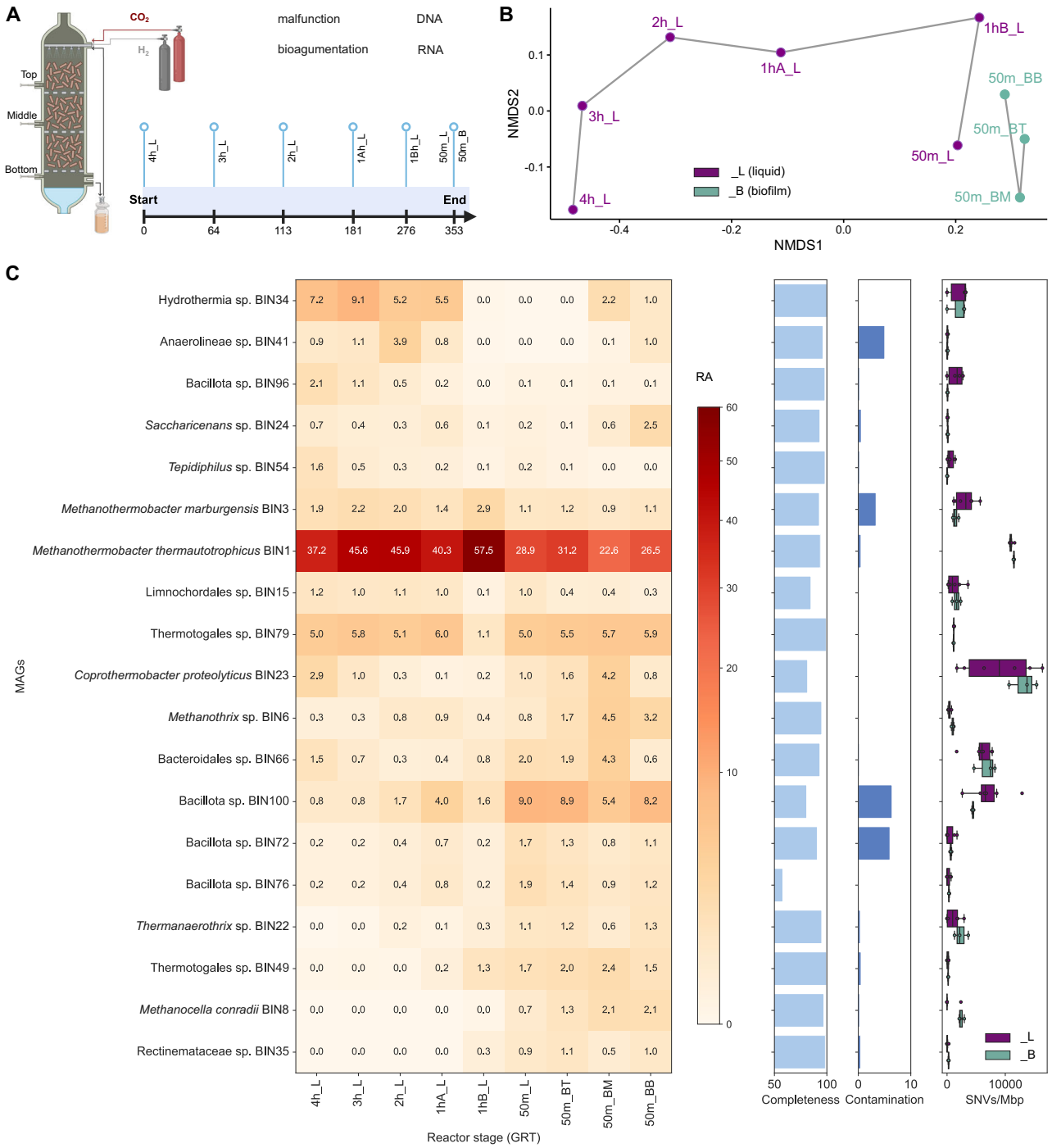

Fig. 1 | Microbial composition and activity at different reactor stages. **A** Experimental design and sample collection. Samples were grouped as liquid (_L) or biofilm (_B), according to their origin. Biofilm samples were differentiated spatially, since they were collected from the top (BT), middle (BM), and bottom (BB) section of the bioreactor column. The bioreactor malfunction (1hA_L) is indicated with a lightning bolt, while the bioaugmentation (1hB_L) with a syringe. Created in BioRender. Treu, L. (2026) https://BioRender.com/tbpsgo5. **B** Beta diversity of reconstructed MAGs was calculated using the robust Bray–Curtis metric. **C** Taxonomic assignment, relative abundance (RA), completeness, and contamination of the dominant microbiome fraction (only selected MAGs with RA > 1% in at least one sample are reported). The rightmost panel illustrates within-species nucleotide diversity (SNVs/Mbp) for each time point, quantified by the fraction of SNVs with intermediate allele frequencies ($0.05 < f < 0.95$). For box plots, the middle line denotes the median SNVs/Mbp, the box denotes the interquartile range, and the whiskers denote 1.5× the interquartile range.

the CH_4 yield[8]. However, excessive reductions may impair gas dissolution, ultimately limiting the bioconversion performance and imposing physiological stress on microbial populations. Recent studies leveraging long-term tracking of single nucleotide variants (SNVs) have shown the role of genetic heterogeneity in microbial adaptation to fluctuating environmental conditions[9,10]. Following perturbations, selection acts as a key driver shaping microbial communities through de novo mutations and generating new haplotypes within populations[11]. The frequencies of SNVs fluctuate over time due to genetic drift or selective pressures, influencing evolutionary trajectories. Newly emerging strains may acquire enhanced metabolic capabilities through beneficial mutations, driving haplotype diversification and strain replacement[12]. In parallel, purifying selection acts to preserve functional genomic integrity by eliminating deleterious mutations, ensuring the maintenance of essential cellular function[13].

Moreover, when considering population dynamics upon ecosystem perturbations, it is essential to account for the role of mobile genetic elements[14,15]. Such entities, including plasmids, phages, and integrated prophages, depend on host machinery for replication and contribute to genetic innovation and ecological adaptation, profoundly influencing the interplay within microbiota. The virosphere significantly contributes to altering population structure[16], for instance as prophages transition to the lytic cycle in response to stressor (e.g., shifts in temperature or pH)[17] and hijack microbial hosts, triggering a cascade of events having evolutionary consequences. This pressure fosters an ongoing arms race, where microbial defense systems counteract viral infections[18], while viruses evolve mechanisms to evade host detection[19]. Among these defense strategies, the Cas machinery plays a central role by capturing and processing fragments of foreign nucleic acids encountered, ultimately integrating them into a clustered regularly interspaced short palindromic repeats (CRISPR) array within the host genome to update the adaptive immunity system[20]. Conversely, bacteriophages can circumvent such defenses by employing diversity-generating retroelements, which introduce extensive mutations within specific regions of target genes though mutagenic retrohoming, thereby enhancing viral evasion of host immune recognition[21]. Furthermore, phages can evade CRISPR–Cas immunity by encoding anti-CRISPR proteins or by acquiring target sequence mutations and DNA modifications that compromise the base-pairing specificity between the crRNA and the protospacer[22,23]. Therefore, phages with point mutations in the targeted sequence may avoid recognition since the effector complex fails to bind. While this antagonism is acknowledged as a major evolutionary force[24], the population-level consequences of this adaptive competition remain largely unexplored. Although recent studies have revealed pervasive microbial–viral interactions within multiple habitats, including subsurface fractured shales[25] and anaerobic digestion plants[5], detailed mechanistic investigations are still limited. Controlled artificial ecosystems offer a unique opportunity to longitudinally track microbial and viral co-evolution under defined perturbations.

Here, we applied an integrated multi-omic approach to dissect microbial and viral dynamics in a thermophilic mixed-culture methanogenic community within a pilot-scale trickle bed reactor. Stepwise perturbations to the gaseous feed allowed us to assess microbial adaptability and restructuring over 353 days. Combining metagenomics, metatranscriptomics, condition-specific community genome-scale metabolic modeling (CoCo-GEMs)[26] and FBA, we reconstructed genome-level metabolic networks and tracked high-resolution SNV trajectories in both prokaryotes and viruses. Results reveal how system destabilization triggers viral bloom and offer mechanistic insights into the adaptive behavior of the CO_2-fixing anaerobic microbiota. Notably, the findings revealed that certain community members were able to recover from phage infection through targeted modifications of their defense systems, effectively countering the evolutionary strategies employed by phages to sustain their infectivity.

Results

CO_2-fixing microbiome dynamics in a pilot-scale reactor

A multifaceted strategy integrating genome-centric metagenomics, metatranscriptomics, and community-based metabolic modeling was employed to dissect the temporal shifts of a thermophilic microbial consortium cultivated in a pilot-scale trickle-bed reactor for CO_2 biomethanation. Over a 353-day operational period, GRT was systematically reduced from 4 h to 50 min through stepwise increases in gas inflow. Phase transitions were initiated once CH_4 purity stabilized above 95%, at which point microbial composition was profiled (Fig. 1A). A decline in CH_4 yield at 1hA_L, attributed to a trickling pump malfunction, was observed and counteracted through bioaugmentation by introducing fresh media. This intervention re-established biomethane purity over 90 days; however, such an event, marked by system destabilization followed by the establishment of a new equilibrium, represents a unique opportunity to explore the underlying microbial and viral interplay. The microbiome, reconstructed through a longitudinal sampling campaign, exhibited moderate complexity, comprising 113 medium-to-high quality MAGs, encompassing 12 archaeal and 101 bacterial taxa (Supplementary Data 1), collectively covering most of the community (around 80% average alignment rate). Beta diversity analysis using non-metric multidimensional scaling, an ordination method that reduces complex community dissimilarity data into a two-dimensional plot, revealed a marked shift in species composition concurrent with GRT decrease (Fig. 1B). In this representation, the distance between points reflects compositional differences based on Bray–Curtis dissimilarity. Planktonic samples separated along the primary axis, with a pronounced divergence between the initial (4h_L) and later-stage (50m_L) microbiomes. These changes may be attributed to widespread bacterial cell death followed by community restructuring, potentially resulting from a large-scale viral infection. Additionally, biofilm and planktonic populations clustered separately, reflecting niche-specific differences in composition and RA.

Dominant microbial fraction consisted of 20 MAGs stratified into 10 phyla, each with an RA greater than 1% in at least one sample and were retained for downstream analysis (Fig. 1C). The archaeal population was consistently dominated by *Methanothermobacter thermautotrophicus* BIN1 in both the biofilm matrix (23–31% RA) and in the planktonic community (37–58% RA). However, within the biofilm, versatile *Methanothrix* sp. BIN6 and hydrogenotrophic *Methanocella conradii* BIN8 exhibited a significantly higher RA compared to the liquid counterpart (RA > 2%), indicating potential niche partitioning. Among bacteria, Thermotogales sp. BIN79 and Hydrothermia sp. BIN34 were prevalent up to 1hA_L (RA > 5%). At lower GRTs, within the Bacillota phylum some species became progressively abundant (e.g., BIN72, BIN76, BIN100), replacing others (e.g., BIN15, BIN54, BIN96). Spatial stratification along the reactor column was also evident with distinct microbial consortia enriched at different depths of the biofilm. Such patterns are consistent with expected microenvironmental gradients in trickle-bed systems, where gas–liquid transfer rates, nutrient availability, and metabolite accumulation can vary with reactor depth. Given that feeding occurred from top to bottom, upper layers were likely exposed to more abundant substrates, while lower sections experienced comparatively depleted conditions. *Coprothermobacter proteolyticus* BIN23 and Hydrothermia sp. BIN34 were exclusively abundant in the middle-section biofilm, while *Saccharicenans* sp. BIN24 dominated at the bottom, all with RA > 2%, suggesting that vertical heterogeneity created ecological niches that favored different taxa (Fig. 1C).

https://doi.org/10.1038/s41467-026-73084-2

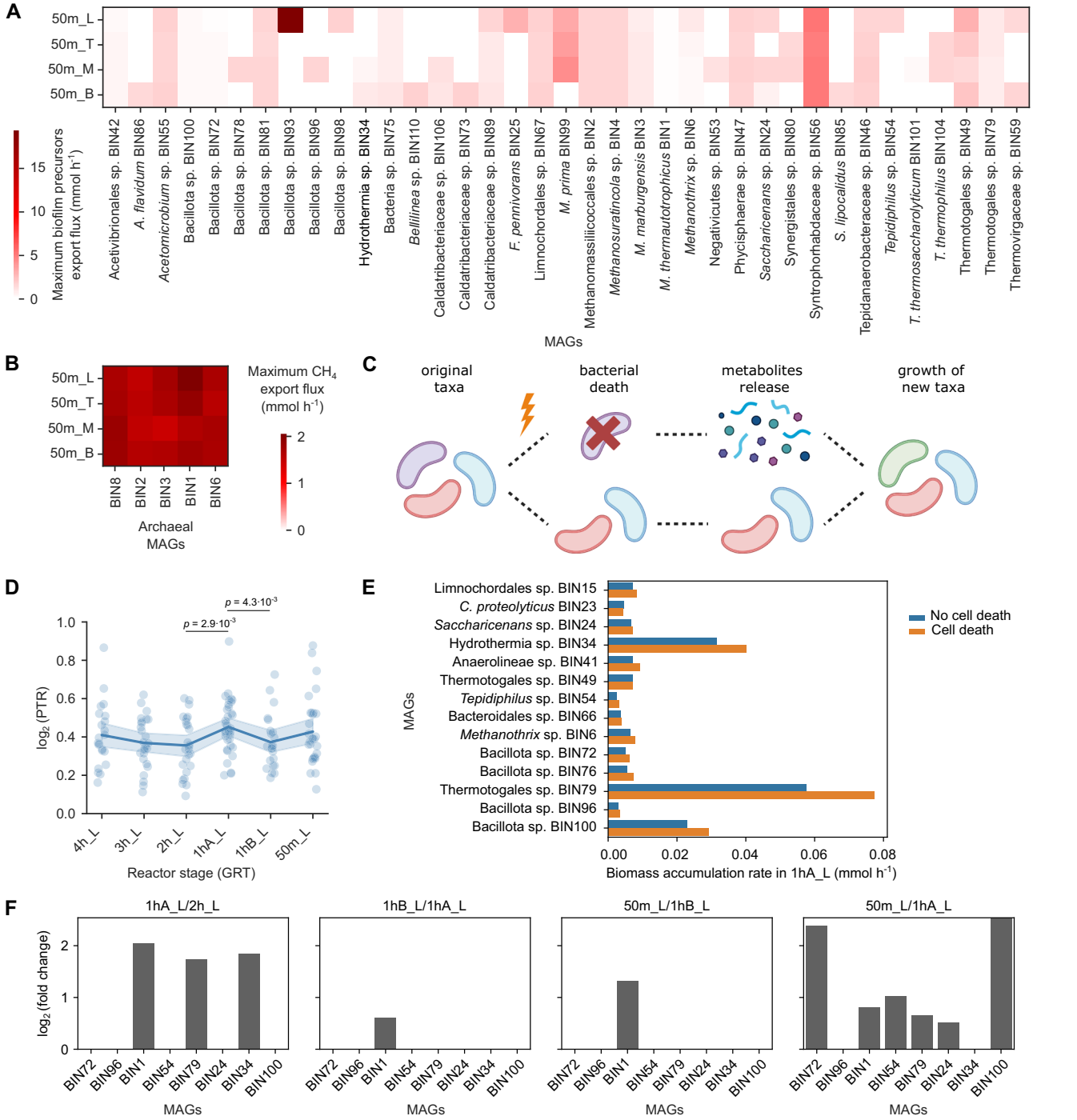

Fig. 2 | Metabolic modelling of microbiome fluctuations upon bioreactor malfunction and following bioaugmentation. A Maximal species-wise export rate of biofilm precursors (N-acetylneuraminate, N-acetyl-D-glucosamine, galactose) in the final time point. **B** Maximal archaeal CH_4 export rate in the final time point. **C** Schematic representation of the hypothesized microbial population dynamics in the 2h_L – 1hB_L interval. Created in BioRender. Treu, L. (2026) https://BioRender.com/tbpsgo5. **D** Microbial replication rate in different time points, expressed as $\log_2$(PTR). The line represents the mean and the corresponding band shows the 95% confidence interval. Statistical differences between 2h_L, 1hA_L, and 1hB_L are calculated by two-sided Wilcoxon signed-rank tests. **E** Comparison between biomass accumulation rates in 1hA_L if assuming a regular medium (blue) and a medium enriched by cell death releasing intracellular compounds of declining species (orange). **F** Differences in biomass accumulation rate across different sample pairs under the hypothesis of a medium enriched by cell death in 1hA_L and a medium enriched by bioaugmentation in 1hB_L.

Despite substantial taxonomic restructuring, the activity repertoire of the microbial community remained remarkably stable throughout the experimental period (Fig. S1). In fact, early dominant Bacillota spp. (e.g., BIN15, BIN54, BIN96) and those enriched at lower GRTs (e.g., BIN72, BIN76, BIN100) shared conserved gene clusters associated with biofilm formation, including extracellular polymeric substances (EPS) secretion, such as stewartan (*cps*) and lipopolysaccharides (*lpt*), pilus assembly (*pilA*, *cpa*, *flp*), and adherence factors (*tad*). The presence and activity of these genes was also observed in Bacillota spp. that increased in RA at lower GRTs, including BIN72, BIN76, and BIN100, underscoring the functional continuity of biofilm-associated capabilities. In the same fashion, quorum sensing, and

https://doi.org/10.1038/s41467-026-73084-2

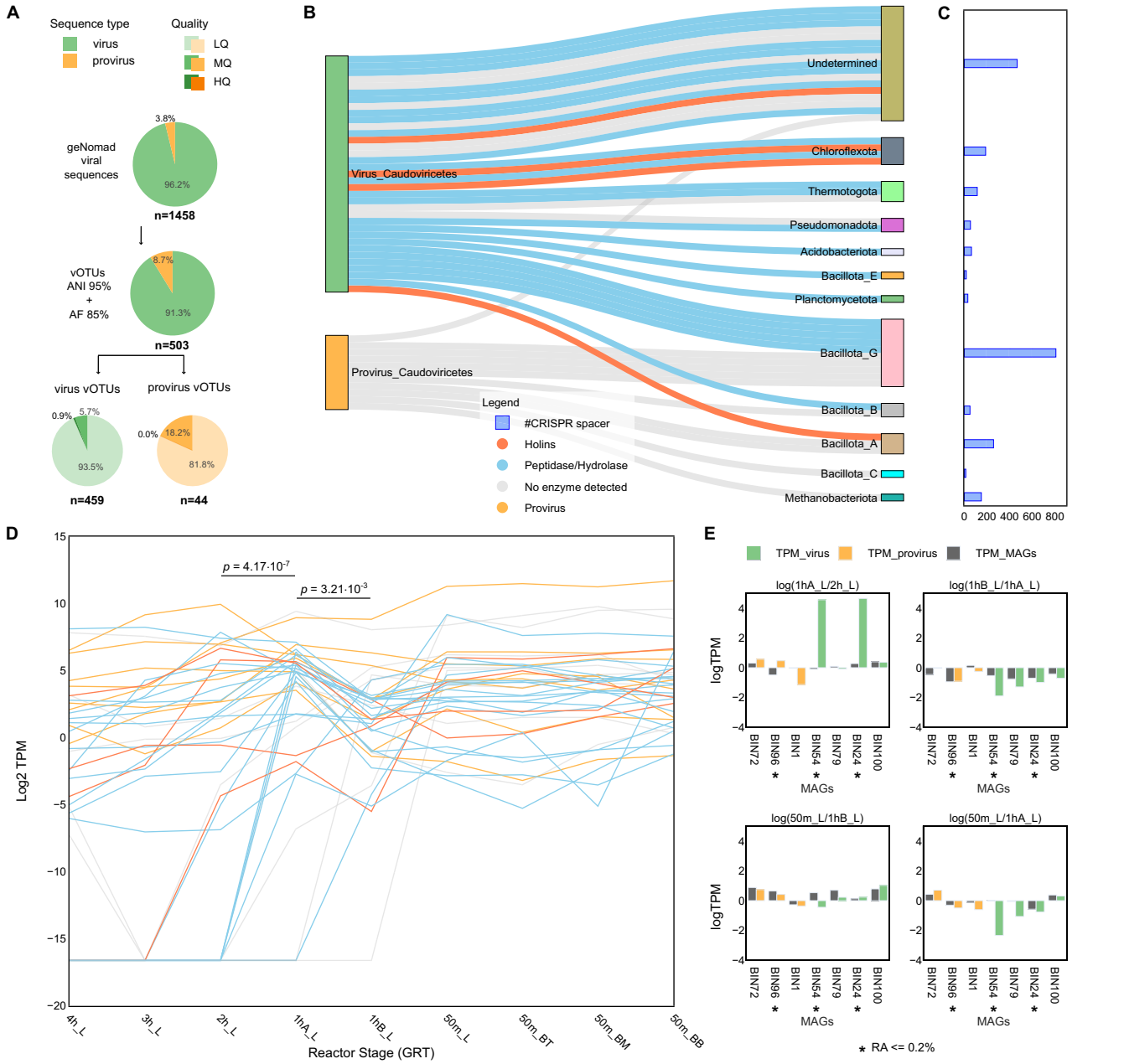

Fig. 3 | Virome characterization of an anaerobic CO_2-fixing microbiome. **A** vOTUs are classified according to their quality and their assignment as either viral (green) or proviral (yellow) genomes. Darker shades are applied according to the vOTU quality. **B** Network analysis of virus-host and provirus-host interactions. Hosts classification at phylum level is reported with a different color. The group named Undetermined includes all vOTUs without a host. Viral genomes are color-coded based on predicted features: light blue for endolysin-encoding, red for holin-encoding, yellow for integrated proviruses, and grey for genomes with no lytic enzymes. **C** Distribution of unique CRISPR spacers across host microbial phyla. **D** Temporal shifts of viral abundance represented as TPM (log_2-transformed), where each line represents the coverage profile of one vOTU. Viral genomes encoding endolysins and holins are highlighted in light blue and red, respectively. Genomes lacking lytic enzymes are shown in grey, and integrated proviruses are depicted in yellow. Statistical differences between 2h_L, 1hA_L, and 1hB_L are calculated by two-sided Wilcoxon signed-rank tests. **E** TPM values were calculated for both hosts (dark grey) and vOTUs (viral: green, proviral: yellow), and then the fold-change between time point 1hA_L, 1hB_L, and 50M_L were computed and log_2-transformed. Hosts perishing after the infection, specifically with RA < 0.2%, are marked with an asterisk (*).

carbon fixation genes remained widespread over the entire experiment, underlying a diffuse functional redundancy. Core functions such as biosynthesis and transport of signaling molecules (*liv*, ACLS) and activity of acetate oxidation routes (e.g., alternative Wood–Ljungdahl/Glycine Cleavage System pathway[27]), were possibly sustained both by persistent bacteria (e.g., Thermotogales sp. BIN79) and by microbes emerging from the rare biosphere post-perturbation (e.g., Bacteroidales sp. BIN66 and Thermotogales sp. BIN49).

Microbiome metabolic modelling suggests potential syntrophies and viral infection

The core metabolic potential and predicted metabolism of the microbiota across time points were assessed using GEMs and FBA. Additionally, community-level GEMs constrained by metatranscriptomics elucidated the activity of core functional modules across the vertical axis of the reactor (Figs. 2A, B and S1). CoCo-GEMs confirmed the relevance of Bacillota and Thermotogales lineages, with a

prominent role in the biosynthesis of the EPS matrix, which underpin biofilm structure and resilience. The *tad* operon, which includes genes associated with surface adhesion and a set of *cpa* pilus assembly genes, were actively transcribed (50–100 TPM). Simulations further suggested that genes encoding the pilus assembly proteins Flp/PilA and Rcp were expressed in Bacillota spp., with the highest transcripts per million (TPM) observed at the mid-height of the reactor. In contrast, Thermotogales spp. exhibited elevated activity of polysaccharide deacetylases (Pgd) and the Cps gene set, which are associated with capsular polysaccharide synthesis and sugar transfer, processes known to modulate microbial adhesion and EPS matrix properties[28,29]. Archaeal MAGs, predominantly *M. thermautotrophicus* BIN1, were also actively engaged in EPS component secretion. FBA simulations predicted active biosynthesis and secretion of N-acetyl-D-glucosamine (e.g., *pgi*, *glmS*, *glmM*, *glmU*) and galactose (e.g., *GALE*, *galT*) by both archaeal and bacterial members (Fig. S1). Globally, potential for producing basic molecules involved in biofilm formation was confirmed by community GEMs for many taxa making up the abundant as well as the rare microbiome (Fig. 2A).

M. thermautotrophicus BIN1, the most abundant and active community member, was the primary contributor to biomethanation. CoCo-GEMs simulations confirmed the archaeon to have the largest methanogenic potential per unit of mass in the liquid, potentially reflecting the metabolic investments in other activities, such as biofilm formation in the other samples (Fig. 2B). Notably, the acetoclastic pathway of *Methanothrix* sp. BIN6 was not found transcriptionally active, suggesting that it may prefer the thermodynamically more favorable hydrogenotrophic route as a survival strategy in this system (Fig. S1). In bacteria, the combined glycine cleavage system and Wood–Ljungdahl pathway, responsible for acetate oxidation[7], was active in Bacillota sp. BIN100, Thermotogales sp. BIN49 and BIN79, with high expression of genes involved in phosphate acetyltransferase-acetate kinase (M00579), reductive acetyl-CoA (M00377), and glycine cleavage (M00621) modules. Formate, a carbon source for hydrogenotrophic archaea[7], may play a key ecological role, with only BIN22 and BIN24 encoding formate transporters (*focA*, *oxlT*), potentially enabling syntrophic interactions with *M. thermautotrophicus* BIN1 and *M. conradii* BIN8, which actively express formate dehydrogenases (*fdh*) (Fig. S1).

To formulate a possible explanation for the observed shift in microbial abundances (1hA_L), graphically represented in Fig. 2C, we first considered microbial dynamics from the point of view of DNA replication rates expressed as $\log_2$(PTR). A spike in $\log_2$(PTR) was found in correspondence to the malfunction (Fig. 2D), suggesting a transient phase of intensified growth for several taxa. This, in conjunction with drastic abundance changes in the following time points (Fig. 1C), support the hypothesis of a major bacterial death event: under such a hypothetical event, cell lysis of some taxa would have made them decline while releasing cellular components in the medium and temporarily boosting growth of the survivors (Fig. 2C). To assess the plausibility of this scenario, GEMs were used to simulate the release of metabolites produced in the cytosolic compartment of species that declined in abundance between 2h_L and 1hA_L. Incorporating these compounds into the growth medium at 1hA_L revealed potential effects on the metabolism and growth of surviving community members. In this process, 157 metabolites were identified and propagated to the shared extracellular compartment, leading to a substantial reorganization of metabolic fluxes when comparing the scenarios with and without a medium enriched by the viral explosion (Supplementary Data 2 and Fig. S2). Among them were Tn antigen (N-Acetyl-alpha-D-galactosaminyl-polypeptide), L-idonate, and D-gluconate, whose increased availability likely facilitated community recovery, serving as nutrient sources, regulatory molecules, or metabolic intermediates that reshaped the ecosystem. Consistently with the increased replication rate, the predicted relative abundance (RA) of several organisms increased, as confirmed by the MAGs exhibiting a significant rise during the infection (Figs. 2E and S3). FBA simulation outcomes thus support the hypothesis of a bacterial death event triggered by bioreactor malfunction (1hA_L), where altered metabolite fluxes influenced biomass accumulation rate (Fig. 2F).

Even though experimental confirmation of cell lysis was not possible due to the unexpected nature of the community shift, the coincidence of the described patterns induced us to further investigate the ecological processes around them. Several mechanisms could potentially explain the hypothesized community restructuring (associated or not with cell lysis). One primary explanation is the onset of competitive interactions in which a part of the bacterial population is predated by one or more taxa. This explanation does not find support in our data, as only three MAGs showed a positive predatory index (Supplementary Data 3), namely *Thermanaerothrix* sp. BIN22, Rectinemataceae sp. BIN35, and Anaerolineae sp. BIN41, and their abundance trends indicate that they are unlikely to drive the observed restructuring (Fig. S4). Moreover, the experimental set-up makes it unlikely that factors like nutrient availability and introduction of new members would drive such large community restructuring. Hence, we considered the possible role of phages in inducing cell lysis as the most plausible explanation.

Viral bursts trigger species turnover from the rare biosphere

The presence and potential role of phages in community restructuring was supported by the identification of mobile genetic elements in the reconstructed metagenome. A total of 503 uncultivated virus genomes were recovered from assembled scaffolds and clustered into 38 medium- to high-quality vOTUs at the species-like level (Fig. 3A). Of these, eight vOTUs were identified as integrated proviruses, while the remaining 30 were predicted to be either virulent (57%) or temperate (43%) bacteriophages (Supplementary Data 4). Host prediction linked 19 *Caudoviricetes* vOTUs to members of 12 distinct bacterial and archaeal orders (Fig. 3A, B). On average, each vOTU was associated with a single host (median = 1.02), suggesting a predominance of unique host-virus relationships. Taxonomic profiling of putative hosts showed broad phylogenetic diversity (Supplementary Data 5), including key taxa involved in anaerobic digestion, such as Bacillota, Chloroflexota, and one provirus associated with *M. thermautotrophicus* BIN1. Notably, these hosts also contained the highest numbers of unique CRISPR spacers, consistent with frequent encounters with virulent phages (Fig. 3C). Functional annotation of viral genomes identified numerous features related to antimicrobial functions, including those implicated in biofilm disruption and cell wall degradation. Among these, 39 genes were associated with the lytic cycle were identified, including ATP-dependent Clp endopeptidase, N-acetylmuramoyl-L-alanine amidase and phage-specific holins, and a variety of proteases (Supplementary Data 6). These genes were predominantly found in viruses infecting hosts from the Bacillota, Chloroflexota and Pseudomonadota phyla (Fig. 3A), which may help explain the sharp drop in RA observed at the 1hA_L time point for Bacillota sp. BIN96, Anaerolineae sp. BIN41 and *Tediphilus* sp. BIN54 (Fig. 1C).

To directly correlate the observed microbial collapse with phage induction, viral read depth was normalized to viral genome length and total reads to reduce noise and tracked across time points (Fig. 3D). Microbes were categorized based on whether their abundance decline coincided with predicted viral predation (putative hosts) or likely arose from indirect ecological dependencies. Statistically significant changes in viral coverage were detected in 1hA_L and 1hB_L time points (Wilcoxon signed-rank test, $p = 4.17 \times 10^{-7}$ and $p = 3.21 \times 10^{-3}$, respectively). Differential responses to infection were evident: while Bacillota spp. BIN100 and BIN72 and Thermotogales sp. BIN79 exhibited resilience, other bacteria including BIN54 and BIN96, perished (Fig. 3E). The corresponding viruses infecting these taxa encoded multiple lytic

https://doi.org/10.1038/s41467-026-73084-2

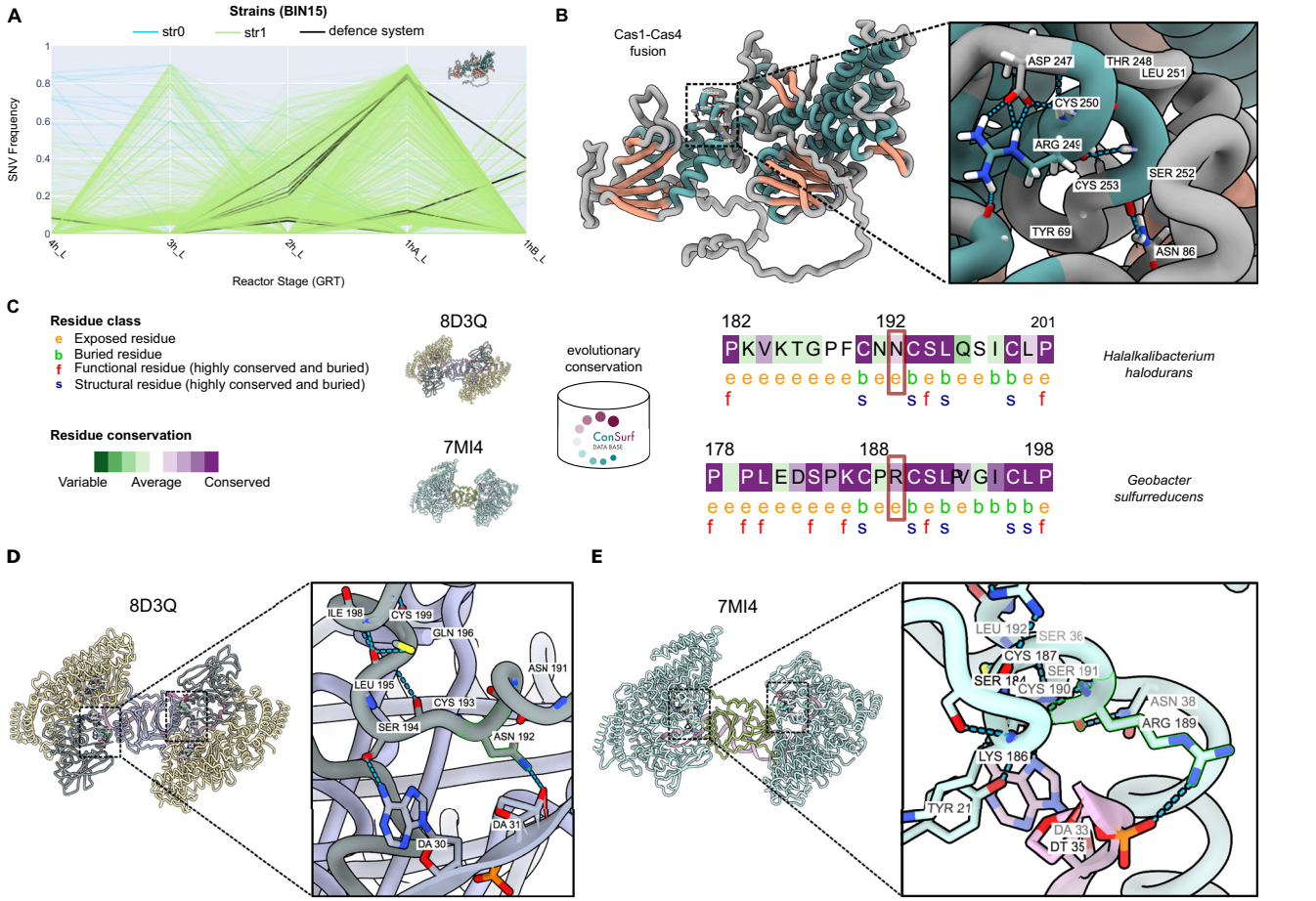

Fig. 4 | Strain level turnover and tridimensional analysis of Cas4/Cas1 fusion protein for Limnochordales sp. BIN15. A Frequency trajectories of nonsynonymous SNVs over time for Limnochordales sp. BIN15. Each strain was highlighted with a different color (str0: light blue, str1: light green), while variants on genes involved in prokaryotic defense systems are highlighted in black. **B** Structure prediction for Cas4/Cas1 fusion gene with fixed nonsynonymous SNVs identified in the dominant strain of Limnochordales sp. BIN15 at 1hA_L. Regions predicted as alpha-helix are highlighted in teal, while beta-sheet in orange. **C** Degree of amino acids conservation in the region neighboring the mutation obtained from the ConSurf-DB. Crystals of Cas4/Cas1 fusion protein associated with Cas2 and the PAM sequence of *Halalkalibacterium halodurans* (8D3Q) and *Geobacter sulfurreducens* (7MI4) were chosen as targets. **D** Crystal structure of 8D3Q. The scale of colors used goes from darkgreen (variable), white (average) and purple (conserved). **E** Crystal structure of 7MI4. Colors identify the Cas1, Cas2 and Cas4 subunits, as well as the PAM sequence. Hydrogen bonds are highlighted as shaded blue lines in the regions upstream and downstream the amino acid mutated in Limnochordales sp. BIN15. All residue positions are numbered relative to the full Cas4/Cas1 fusion protein unless otherwise indicated. For reference, G252 in the fusion corresponds to G190 in the Cas4 domain alone.

effectors such as cell-wall hydrolase (PF20307; PF07486), Clp endopeptidase (PF00574; PF19602) and peptides containing transglycosylase-like domain (PF06737), which were absent in the vOTUs associated with the resistant strains. Together, these observations are consistent with a potential role for phage-mediated turnover in destabilizing microbial populations at the 1hA_L time point, with downstream consequences for microbial structure and function. Several species unaffected by direct phage infection but failing to recover post-bioaugmentation intervention suggest a high degree of metabolic codependency with the phage-sensitive populations. For instance, *Tediphilus* sp. BIN54 and *Saccharicenans* sp. BIN24 were linked through strong citrate exchanges in CoCo-GEMs simulations, highlighting compound sharing to overcome metabolic deficiencies. Similarly, Hydrothermia sp. BIN34, which showed high ethanol export flux, may have depended on its consumption by BIN54 to prevent accumulation of toxic intermediates, underscoring the potential for indirect disruption via the collapse of metabolic partners.

Among the identified proviruses, one was integrated in the genomes of highly abundant Bacillota spp., which showed a transient reduction in RA at the 1hA_L time point. Despite an increase in proviral coverage, both MAGs (BIN72, BIN76) successfully recovered following bioaugmentation, suggesting that the infection remained latent and did not progress to lysis or significantly impair host viability (Figs. 3E and S5). In contrast, a second provirus associated with Bacillota sp. BIN96 appeared to contribute to host collapse, suggesting activation of the lytic cycle and its detrimental impact on the host population (Fig. 3E).

CRISPR Cas-related SNVs support survival during phage exposure

The phage-mediated turnover, coupled with environmental perturbations caused by the reactor malfunction, arguably imposed strong selective pressure that contributed to reshaping the microbial community. To investigate adaptive dynamics, within-species variation was assessed in 20 dominant MAGs by analyzing SNVs patterns over time. After filtering for high-confidence calls, 213,809 SNVs were detected, 32% of which were nonsynonymous and thus potentially associated with protein-level changes. SNV density varied widely across MAGs and time points (Fig. 1C). To assess whether mutational patterns differed across functional gene categories, we compared SNV densities between defense-associated genes and all other genes. Several MAGs exhibited clear enrichment of SNVs in defense genes, while other

https://doi.org/10.1038/s41467-026-73084-2

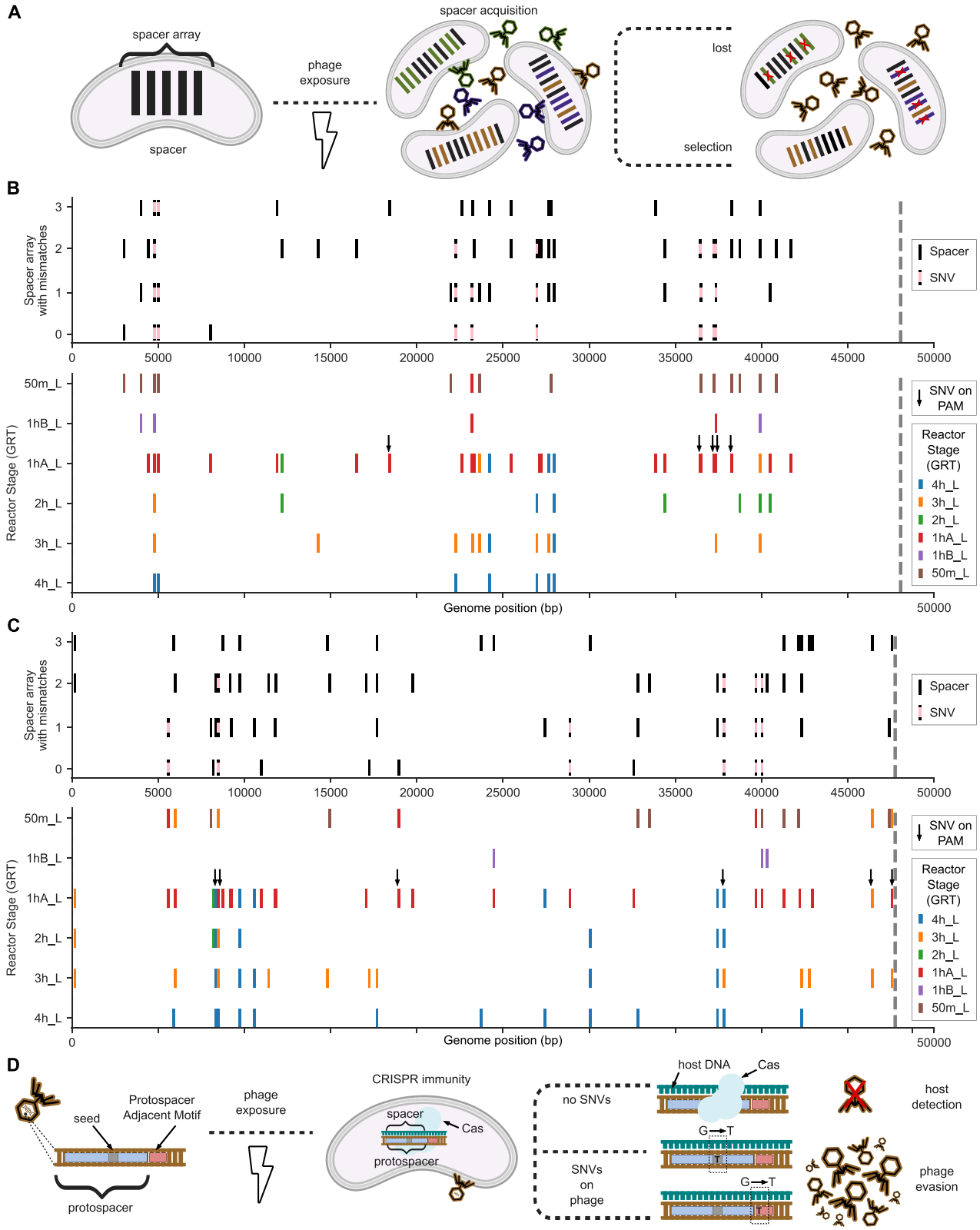

Fig. 5 | Longitudinal dynamics of spacer acquisition and mutation. A Schematic representation describing the accumulation of CRISPR spacers. Results of temporal spacers acquisition are shown for phages. **B** PIL_671 infecting Bacillota sp. BIN72 and **C** PIL_676 infecting Bacillota sp. BIN100. For each phage, the complete spacer array and its subdivision into mismatch categories (0, 1, 2, and 3) are displayed, with spacers showing a progressive increase in mismatches over time highlighted in pink. Spacer acquisition is visualized across time points, with distinct colors (blue: 4h_L, orange: 3h_L, green: 2h_L, red: 1hA_L, purple: 1hB_L, brown: 50m_L) representing each experimental condition and an arrow indicating the presence of SNVs in the 10 bp downstream spacer. Spacers retained across multiple conditions are shown in the color corresponding to their initial acquisition. **D** Schematic representation illustrating putative phage evasion strategy via SNVs on seeds or PAM. Created in BioRender. Treu, L. (2026) https://BioRender.com/tbpsgo5.

species showed markedly elevated SNV densities in specific loci (Fig. S6). Additional support for the hypothesis of ongoing adaptive evolution emerged from genome-wide and gene-level estimates of the average ratio of non-synonymous and synonymous substitutions (Supplementary Data 7). Genes with high *dN/dS* ratios were predominantly involved in biosynthetic pathways (e.g., secondary metabolites, cofactors, amino acids) and substrate-specific metabolic functions, including central carbon metabolism. Notably, subtype I-C *cas4/cas1* fusion gene in Limnochordales sp. BIN15 has a dN/dS = 1.5 at 1hA_L, reinforcing the notion of positive selection acting on host defense mechanisms during viral perturbation.

To identify signatures of selection and strain-level turnover, nonsynonymous SNVs were clustered and phased into haplotypes to infer microbial strain composition. A total of 16,162 SNVs with intermediate allele frequencies ($0.15 < f < 0.85$) across all time points were retained, focusing on those showing marked frequency shifts over time (e.g., from <20% to >70%). Up to four coexisting strains were inferred per MAG, each characterized by distinct SNV frequency trajectories (Figs. 4A and S7). These patterns revealed three primary modes of within-species adaptation: (i) evolutionary modification, where novel SNVs emerged and arose in frequency within the local environment, as observed in Hydrothermia sp. BIN34 and Limnochordales sp. BIN15. A cluster of SNVs nearly reached fixation, only to experience a subsequent decline in frequency later (Figs. 4A and S7C). (ii) strain replacement, where a pre-existing, genetically divergent lineage overtook the population. For example, a genomic sweep occurred in Bacillota sp. BIN100 (Fig. S7B) at 1hA_L time point. Here, SNV trajectories indicate that a distantly related strain was already present at substantial frequencies (~5%) long before its significant fitness advantage began to manifest. iii) strain coexistence, where multiple lineages persisted with stable dominance hierarchies, as observed in *Methanothermobacter* spp. (Fig. S7E).

In several taxa, the dominant strains appeared to gain a phenotypic advantage that enhances resistance to viral infection. Accordingly, anti-phage defense systems were systematically identified across MAGs and examined for signatures of selection (Supplementary Data 8). In Limnochordales sp. BIN15, four nonsynonymous SNVs (A7T, P18A, V19A, and G252S) rose from low ($f < 0.1$) to high allele frequencies ($f > 0.5$) in the sole gene encoding a Cas4/Cas1 fusion protein from a type I-C CRISPR-Cas system (Fig. 4A). These mutations, mostly located in random coil regions, later revert to ancestral alleles, suggesting transient rather than sustained selection. The G252S substitution in the Cas4/Cas1 fusion protein, corresponding to G190S when numbered within the Cas4 domain alone, lies within the Cas4 domain (Figs. 4B and S8). Although not located in active sites or Fe–S clusters, structural comparison with resolved Cas1/Cas4/Cas2 complexes indicates that this residue engages in non-covalent interactions with PAM nucleotides of prespacers. To assess potential functional impacts, two experimentally resolved Cas1/Cas4/Cas2 complexes bound to PAM-containing prespacers (PDB: 7MI4[30], 8D3Q[31]) were compared to the BIN15 Cas4/Cas1 protein sequence (Fig. S8). The Gly→Ser mutation at position 190 (Cas4 domain) in Limnochordales sp. BIN15, which is Gly→Asn in *Halalkalibacterium halodurans* (pos = 192) and Gly→Arg in *Geobacter sulfurreducens* (pos = 189), is proximal to a highly conserved Cys-Ser-Thr triad (Fig. 4C). Specifically, focusing on the formation of hydrogen bonds highlights how this position is engaging in non-covalent interactions with the PAM nucleotides of the prespacer (adenine at positions 31 and 33, labeled DA31 and DA33 in the crystal structures) in both *G. sulfurreducens* and *H. halodurans*. Aside from CRISPR-Cas, the only other detected defense system identified in this population was BREX, which showed no signs of selection. Thus, the observed transient SNVs in *cas4/cas1* represent the sole detectable adaptive signal in Limnochordales sp. BIN15, potentially reflecting phage-driven dynamics. While it remains unclear whether such SNV enhance or impair Cas4 function, the temporary rise in frequency suggests plausible influence on host–phage interactions during the infection. Future work will be needed to test whether such variants can alter PAM recognition or spacer acquisition efficiency.

Similarly, in Bacillota sp. BIN72, a gene encoding another Cas system exhibited signs of positive selection, with three SNVs sweeping through the population: Q156E in a random coil region and A146G-V226A in α-helices H4 and H6, respectively (Fig. S9A, B). Relative to its closest known homolog (RJP19136) reveals that these mutations are localized within the Cas1 domain, potentially favoring spacer acquisition, contributing to improved viral defense and survival. In other MAGs (i.e., Bacillota sp. BIN100), SNVs also accumulated in other defense-related genes, though these variants never surpassed the reference allele in frequency, indicating neutral drift or transient selective advantage (Fig. S7).

Phage exposure hampers spacer acquisition over time

Reconstructed phage genomes were profiled for their genetic heterogeneity, yielding 4472 high-confidence SNVs in 30 vOTUs. The raw SNVs counts were generally below 1000 at all timepoints, except for 1hA_L, where the count exceeded 2500 (Fig. S10A). This increase was statistically significant (p-value = 9.77×10^{-4}) when comparing the SNVs frequency across vOTUs before (2h_L) and during (1hA_L) the infection event (Fig. S10B, "Methods"). This pattern mirrors observations at the MAG level, indicating that both virome and microbiome experience a diversification process, possibly under selective pressure, during the bioreactor perturbation. Supporting this conclusion, a subsequent decrease in SNV counts was observed after the infection (1hB_L) when compared to 1hA_L, with a p-value of 3.10×10^{-2} from the Wilcoxon Signed-Rank Test (Fig. S10B, "Methods"). To investigate the underlying evolutionary arms race between phages and their hosts, we further examined CRISPR spacer acquisition and turnover in three representative cases (Fig. 5A).

The phage PIL_407, predicted to infect Limnochordales sp. BIN15, exhibited limited evidence of selection, with only one SNV with high allele frequency ($f > 0.5$) detected at 1hA_L. This variant was in a non-coding intergenic region (position 32,271), though it may influence transcriptional regulatory processes, or be associated with evasion from CRISPR spacer targeting. During infection, a transient spike in spacer acquisition in the corresponding host was observed, increasing from 14 to 34 CRISPR spacers at the infection peak, followed by rapid loss (Supplementary Data 9). However, no evidence of selection at the spacer level was detected, as no SNVs were localized within protospacer regions, nor at the PAM or seed. Additionally, there were no spacers with increasing mismatch count over time.

Conversely, PIL_671, which infects Bacillota sp. BIN72, showed markedly different dynamics. The SNV density was over four times that observed in PIL 407, with a total of 169 SNVs across the experimental period. A pronounced mutation hotspot (37–40 kbp) reached 30.7 SNVs/kbp, which was nearly ninefold higher than the genome-wide average density (3.52 SNVs/kbp, Fig. S11A). Variants within this region affected genes encoding two transcriptional regulators and the antitoxin EndoAI. Additional highly variable regions (>10 SNVs/kbp) encoded putative regulators and an HNH endonuclease. Mapping of CRISPR spacers to the phage genome revealed 194 unique hits, 10.82% of which overlapped with SNV-rich regions. However, no spacers were found targeting the intergenic region 39559–39582, where six SNVs with $f > 0.5$ have been identified at 1hA_L. As with PIL_407, rapid spacer acquisition occurred during infection, yet many spacers failed to reach fixation (Fig. 5B). This transient pattern was consistently observed across other activated phages (Supplementary Data 9), supporting the notion that increased exposure to viral particles triggers CRISPR activation without necessarily leading to long-term spacer retention. Moreover, the presence of multiple SNVs within eighth spacers, which progressively acquired mismatches from 4h_L to 1hA_L, provides clear evidence of an ongoing arms race between phage and host (Fig. 5B).

These observations are consistent with the previously proposed adaptive response of Bacillota sp. BIN72, involving modifications to the *cas1* gene (Fig. S9B). In parallel, phage PIL_671 appears to acquire variation within targeted protospacer regions, as well as multiple SNVs in the downstream PAM sequence (Fig. 5B), likely representing a strategy to evade CRISPR-mediated recognition. The concurrent rise in SNV counts in both host and phage genomes at 1hA_L further underscores the presence of reciprocal selective pressure towards population diversification (Figs. S9A and S10B).

A third example is phage PIL_676, targeting Bacillota sp. BIN100. This phage showed the highest degree of genetic heterogeneity (12.13 SNVs/kbp), with multiple hotspots distributed along the genome (Fig. S11B). Several of these regions encompass genes involved in capsid assembly and early infection processes, including a peptidase (PF03420) and a recombinase (PF09956) containing domains implicated in virion assembly, as well as a phosphoesterase (PF01710) associated with efficient DNA transposition during host attachment and genome penetration. Additionally, a variable region included an endolysin containing a transglycosylase SLT domain (PF01464), suggesting that selective pressure acts on functions related to infectivity. Despite this diversity, all detected SNVs were synonymous, pointing to possible selection acting at the RNA or regulatory level rather than on protein sequence. These mutations are likely driven by the robust antiviral defense arsenal of Bacillota sp. BIN100 (e.g., SoFic, RM type II, AbiE, Gao Upx and type I-C CRISPR-Cas), which allows the host to withstand infection while exerting strong selective constraints on the phage. Moreover, similarly to PIL_671, multiple protospacers in PIL_676 accumulate SNVs either within the protospacer sequence itself or in the associated PAM (Fig. 5C), indicative of convergent evolutionary trajectories toward CRISPR evasion. Notably, these dynamics were observed exclusively for these two phages, whereas in the rest of the viral population selective pressure acted on both coding and non-coding regions lacking protospacers. The timing and repeated emergence of PAM-disrupting SNVs across targeted protospacers strongly support CRISPR-driven selection hypothesis, with phages actively trying to evade host detection (Fig. 5D).

Discussion

Microbial consortia are central to bioconversion processes such as CO_2 biomethanation; however, their functional stability is highly susceptible to system perturbations, such as viral predation. In this study, we examine a thermophilic pilot-scale bioreactor subjected to a stepwise increase in H_2/CO_2 flow rates, integrating metagenomics, metatranscriptomics, virus mining, and FBA to capture both microbial and viral responses. Our results reveal a coordinated sequence of events consistent with a virus–host evolutionary arms race: lytic phage outbreaks coincide with host population collapse, followed by transient adaptive modifications in bacterial defense systems and, in some cases, recovery of resilient taxa. This framework highlights how phage-mediated selective pressures can shape community composition, functional stability, and resilience in engineered anaerobic ecosystems.

Consistent with previous findings[32], progressive reduction of the GRT imposed a strong selective pressure on the microbiome, driving shifts in species composition despite stable biomethanation. The transition to 1 h GRT coincided with a system malfunction, prompting bioaugmentation intervention to restore bioreactor function and return the microbiome to baseline conditions, as successfully implemented in similar studies[33]. The observed microbial dynamics could be consistent with multiple mechanisms, including generalized phage-mediated turnover and "kill-the-winner" hypothesis, without asserting direct causality. The absence of either competitive interaction, nutrient-driven restructuring, and the introduction of novel species implicates viral predation, amplified by operational perturbations, as the primary driver of transient community restructuring. Metagenomic analysis revealed a pronounced spike in viral coverage at the one-hour GRT timepoint, and a subsequent drop as host populations were depleted. The timing of these events aligned closely with host population decline, consistent with phage-mediated lysis. The release of viral particles likely impacted both infected hosts and their syntrophic partners, amplifying the ecological consequences of lytic outbreaks. FBA further supported this interpretation, indicating that metabolites released through lysis could temporarily sustain opportunistic taxa, linking viral predation to ecosystem-level effects in a deterministic framework.

Surviving bacterial populations exhibited adaptive responses consistent with an ongoing evolutionary arms race by transient evolutionary modifications in their anti-phage defense systems. In Limnochordales sp. BIN15, four nonsynonymous variants arose in the gene encoding a Cas4/Cas1 fusion protein of a type I-C CRISPR-Cas system. These mutations reached intermediate–high frequency during the infection timepoint before reverting to the ancestral state, consistent with transient selection rather than fixation. The G252S substitution occurred within the Cas4 domain, proximal to a conserved Cys–Ser–Thr motif involved in prespacer processing. While not located in canonical Fe–S cluster[34] or Mn^{2+}-binding sites[35] where mutations are typically deleterious, structural comparison with resolved Cas1/Cas4/Cas2 complexes indicates that the corresponding residue interacts with PAM nucleotides of prespacers in other species[30,31]. These observations suggest that the transient variants could have influenced CRISPR activity, although whether such changes enhance or impair spacer acquisition remains unclear. A comparable pattern was observed in Bacillota sp. BIN72, where Cas1, also involved in spacer acquisition, harbored nonsynonymous changes that may have influenced prespacer processing, though no structural evidence was available to confirm their effect. Together, these cases highlight how selective pressure during viral predation can transiently shape host defense proteins, even outside canonical active or cofactor-binding sites, underscoring the diversity of evolutionary trajectories in microbial immune systems.

CRISPR spacer acquisition patterns across the microbiome further reflected this selective landscape. New spacers accumulated significantly during the peak viral load, yet many failed to reach fixation and were subsequently lost. The timescale of spacer turnover observed here, spanning approximately 80 days, falls into an intermediate behavior. Studies on model bacteria in controlled systems reported a spacer turnover of 4–24 h[36–38], while in natural environments (i.e., soil, human gut) the event appears to be much slower[39–41], taking even years. Many acquired spacers were not retained, likely due to genetic drift, population replacement, or other neutral processes affecting allele fate. Thus, while spacer turnover is consistent with a coevolutionary arms race, it may also reflect stochastic fluctuations rather than solely adaptive responses, tempering, but not diminishing, the mechanistic insights from our data. Our observations align with previous studies, which showed that coevolving phages maintain host resistance diversity through negative-frequency-dependent selection and repeated acquisition of new immunity[42]. Other experimental and modeling studies indicate that the spread of CRISPR-Cas systems depends on phage abundance, the frequency of alternative resistance mechanisms, and population dynamics[43], while fitness costs of CRISPR can drive spacer loss or replacement[44]. Together, these findings suggest that CRISPR-mediated processes support microbial resilience, shape community structure, and operate alongside stochastic and ecological forces, positioning our results within broader host–phage coevolutionary principles.

While this spacer turnover was consistent across hosts, phage genetic heterogeneity varied markedly. The presence of selective pressure was diffused across phages during the infection, but the mechanism through which phages tried to escape bacterial immunity was highly virus-host system-specific. For instance, one phage exhibited relatively low mutation density, with variants primarily affecting

genes involved in host cell lysis. In contrast, other phages acquired a substantially higher number of SNVs per kilobase, many of which mapped to genes involved in transcriptional regulation or host penetration. Notably, in two phages we observed SNVs accumulating directly within protospacer regions or their associated PAM sequences, progressively increasing mismatch levels over the course of infection. The recurrence of PAM- and protospacer-associated mutations across independently targeted loci argues against purely neutral evolution and instead supports CRISPR-driven selection. The diversity of observed phage strategies suggests that no single ecological or evolutionary model fully captures the ecological complexity. Elements of kill-the-winner dynamics[45], royal family turnover[46], and transient coexistence[47] appear to operate simultaneously, with their relative contributions shaped by lineage-specific parameters such as initial genetic diversity, the frequency of resistant versus susceptible hosts, and the architecture of antiviral defense systems. This perspective aligns with recent work demonstrating that intrinsic population structure and evolutionary potential strongly influence host-phage arms races[48,49], in addition to external environmental parameters. Importantly, the scale of this study, encompassing a diverse, multi-species microbiome and its associated virome under continuous operation, is much larger. While in vitro assays often isolate single mechanisms, our results show that in complex engineered systems, viral predation triggers a spectrum of responses that vary across lineages and timescales. Despite this heterogeneity, a unifying feature emerges: most host–phage pairs respond to environmental stress with a transient intensification of active infection, followed by divergent evolutionary and ecological outcomes. These outcomes are likely governed primarily by lineage- or population-specific parameters, such as the initial genetic diversity and relative frequencies of resistant versus susceptible hosts[42], rather than by external environmental factors alone.

Overall, our findings showcase how viral predation can act as a major driver of microbial dynamics in engineered CO_2-fixing ecosystems. The observed transient host adaptations, spacer turnover, and phage genetic diversification collectively illustrate an ongoing arms race operating at both ecological and molecular scales. By linking temporal changes in viral load, host decline, and immune adaptation, we show that phage–host interactions can override environmental perturbations and reshape community structure within operational timescales. These insights not only advance understanding of microbial resilience in bioreactors but also provide a conceptual foundation for future studies aiming to predict, manipulate, or harness phage–host interactions for biotechnological applications.

Methods

Experimental setup and biochemical monitoring

The study employed a pilot-scale trickle bed reactor consisting of a 100 L packed-bed cultivation vessel, gas and liquid delivery systems, and a gas-metering system (Supplementary Data 10). The bioreactor was filled with high-density polyethylene raschig rings (dimensions: 10×10 mm, surface area 3.3×10^{-3} m^2/g) and operated at a stable thermophilic temperature (55 ± 1 °C). The nutrient medium was stored in a stainless-steel sump with a working volume of 100 L. To ensure homogeneity and prevent settling of particulates, intermittent recirculation of the medium was performed using a dedicated peristaltic pump (Sci-Q 300, Watson Marlow, United Kingdom), enabling proper mixing within the sump. Delivery of the nutrient medium to the reactor was carried out by a separate peristaltic pump (Sci-Q 300, Watson Marlow, United Kingdom) at a flow rate of 65 L/h. The input gases (H_2 and CO_2) from gas cylinders were co-fed into the trickle-bed reactor along with the liquid stream. Their flow rates were regulated using gas mass flow controllers (GFC Mass Flow Controller, Aalborg, USA). Output gas production was monitored daily using drum-type gas meters (Ritter, Germany). Gas phase composition was analyzed using a gas chromatograph (GC-2014 Pro, Shimadzu, Japan) equipped with a thermal conductivity detector as described by ref. 32. Liquid samples were collected twice weekly for pH, measured with an HI2020 EDGE pH meter (HANNA Instruments, USA), and volatile fatty acid concentration, determined using gas chromatography (GC-2010 Pro, Shimadzu, Japan) with a flame ionization detector as described by ref. 50. A thermophilic hydrogenotrophic enrichment culture was used to inoculate the pilot-scale TBR. The inoculum originated from a previously maintained laboratory-scale culture cultivated at 55 °C using H_2/CO_2 as substrates. Prior to reactor operation, a conditioning step was performed to ensure complete wetting of the packing material. This involved continuously trickling 30 L of sieved (< 2 mm) and degassed digestate, used as a nutrient solution, over the packing material. Sieving was essential to remove coarse solids, preventing clogging and ensuring uniform liquid distribution. After this conditioning phase, the enriched inoculum was evenly applied to the upper surface of the TBR, promoting homogeneous colonization and robust biofilm formation. The experimental design consisted of multiple stages (Fig. 1A), involving continuous operation with progressively reduced GRT over 353 days, from 4 h to 50 min. Transition to the next stages was applied only when stable CH_4 output (> 95%) was reached. The experimental stages included: 4h_L, 3h_L, 2h_L, 1hA_L, 1hB_L, and 50m_L (Supplementary Data 11). Time points 1hA_L and 1hB_L correspond to the sampling points immediately before and long after (60 days) the bioaugmentation event. This intervention followed a malfunction of the liquid trickling pump, which interrupted reactor operation for ~72 h. During this period, gas and liquid feeding were halted. To restore optimal performance, the TBR column was replenished with fresh inoculum from the same enrichment culture, and operation resumed under identical conditions.

Nucleic acids extraction and sequencing

Metagenomic samples were collected from the liquid phase of the cultivation vessel on the last day of each operation change, i.e., the day prior to each GRT reduction. Liquid samples are classified as follows: 4h_L, 3h_L, 2h_L, 1hA_L, 1hB_L, 50m_L. Biofilm samples collected at GRT of 50 min are classified as follows: 50m_BT (top), 50m_BM (middle) and 50m_BB (bottom). DNA extraction was carried out employing the DNeasy® PowerSoil® Kit (QIAGEN, Hilden, Germany), following the manufacturer's recommended procedure. The sampling of biofilm was performed at the end of the whole experimental period to allow the full development of the extracellular matrix. Total RNA was extracted using the RNeasy PowerSoil Total RNA Kit (QIAGEN, Hilden, Germany), following the manufacturer's protocols. Aliquots for DNA and RNA sequencing were withdrawn from each sample. Prior to extraction, the RNA aliquots underwent DNAse treatment in accordance with the manufacturer's protocol. For each extracted sample, quality and concentration were evaluated using Nanodrop2000 and Qubit fluorometers (ThermoFisher Scientific, Waltham, MA, USA), respectively. Furthermore, RNA quality was assessed using an Agilent 2100 Bioanalyzer with RNA 6000 Nano reagents and RNA Nano Chips (Agilent Technologies, Santa Clara, CA, USA). Ribosomal RNA was removed using the FastSelect -5S/16S/23S Kit (QIAGEN GmbH, Germany). The library was prepared using the Nextera DNA Flex Library Prep Kit and the llumina Ligation Stranded mRNA Prep. Then, 150 + 150 PE sequencing was conducted on an Illumina NovaSeq 6000 platform (Illumina Inc. in San Diego, CA.) at the DiBio-seq facility (University of Padova, Italy).

Metagenomic and metatranscriptomics reconstruction of the microbiome

Paired DNA reads (150 bp) were quality-filtered with trimmomatic[51] v0.39, co-assembled using SPAdes[52] v3.15.5 (parameters:"–meta -k 21,33,55,77,99,119") and mapped to the assembly with Bowtie2[53] v2.4.5 (parameters: "default"). Binning was conducted using a combination of

tools, including MetaBAT1[54], MetaBAT2[55] v2.12.1, MaxBin2[56] v2.2.6, and Vamb[57] v4.1.3 and refined with Binette[58] v1.0.1, using all software with default parameters. MAGs were assessed for quality using CheckM2[59] v1.0.1 and dereplicated with dRep[60] v3.4.0 (parameters:"-comp 50 -con 10 -sa 0.95 -nc 0.5"), with only those of medium or high quality according to MIMAGS specifications further processed[61]. Taxonomy was assigned using GTDB-Tk[62] v2.1.0 (database version R214, parameters:"default"). Prodigal[63] v2.6.3 (parameters:"-p meta") and EggNOG-mapper[64] v2.1.10 (parameters:"-m diamond –dmnd_iterate no") were utilized for protein encoding genes prediction and annotation. Prokaryotic reads were aligned on reconstructed MAGs with Bowtie2[53] v2.4.5, then RA and TPM of the MAGs were estimated with CoverM[65] v0.6.1. The Kyoto Encyclopedia of Genes and Genomes (KEGG) was used as reference for enzyme classification, orthology, and metabolic pathways reconstruction. RNA-seq reads were processed as the DNA reads for quality and mapping, while gene fragment counts were computed with HTSeq[66] v2.0.5 in stranded mode (parameters:"-a 0 –mode intersection-nonempty –nonunique fraction -s reverse -t CDS -i ID"). Metadata on recovered MAGs are summarized in Supplementary Data 1. Predatory index for dominant microbiome fraction was computed from gene copy numbers of predatory-specific (mevalonate pathways) and non-predatory-specific (DOXP pathway) genes, as reported in previous literature[67].

SNVs analysis and tridimensional structure prediction of anti-phage systems

InStrain[68] v1.8.0 (parameters:"-min_mapq 2, –min_read_ani 0.98, and –min_genome_coverage 1") was used to characterize the SNVs profile of high-quality MAGs and phage genomes. The InStrain profile module requires a multifasta file containing the MAGs, individual BAM files for each sample generated via Bowtie2[53] v2.4.5 (parameters:"default"), a scaffold-to-bin file, and a gene annotation file produced with Prodigal[63] v2.6.3 (parameters:"-p meta"). Variant calling results were subsequently filtered to remove low-confidence SNVs: first, SNVs within 150 base pairs of the 3′ and 5′ ends of each scaffold were removed due to reduced coverage and reliability; second, variants for which the difference between the SNV coverage and the average scaffold coverage exceeded the interval [−100; +100] were discarded. Following this selection, variants were grouped based on their frequency over time using scipy.stats (Python), with hierarchical clustering algorithm. The Ward distance metric was applied to the profile of frequency of SNVs and used to calculate variant similarity. The ratio of nonsynonymous to synonymous substitutions (dN/dS) was calculated based on the SNVs being classified as either synonymous (S) or nonsynonymous (N). The same procedure applied to MAGs was reiterated for phages, with the exception that the output of geNomad[69] v1.8.0 (parameters:"default") was used as gene annotation file. Integrated bacteriophage genomes were excluded from this analysis.

Anti-phage systems detection was performed with DefenceFinder[70] v1.3.0 (parameters:"-a"). The presence and completeness of each system are reported for all MAGs in Supplementary Data 8. Protein structure prediction for genes encoding anti-phage systems was conducted with ColabFold[71] v1.5.5 (parameters:"-num-relax 5 -model_type auto -num_recycles auto"). The predicted structures were compared to the best BLASTp[72] hit with a known tridimensional structure (closest homolog) using the iCn3D built-in portal. ChimeraX[73] v1.9 was employed for PDB visualization of both predicted structures and crystals of deposited Cas1/Cas4 fusion proteins interacting with protospacer-adjacent motif (PAM) sequence. The conservation level of amino acids in the region surrounding the mutation was obtained from the ConSurf-DB[74] database. Residue numbering of nonsynonymous mutations was based on the full-length Cas4/Cas1 fusion protein from Limnochordales sp. BIN15. For comparative structural analyses with experimentally resolved Cas1/Cas4/Cas2 complexes (PDB: 7MI4, 8D3Q), numbering within the isolated Cas4 domain is also provided. For example, the G252S mutation in the fusion corresponds to G190S when indexed relative to the Cas4 domain alone.

Viruses identification, clustering, host predictions and spacers dynamics over time

Viral contigs were identified using geNomad[69] v1.8.0 (parameters:"–enable-score-calibration –conservative", database version–latest) and processed with CheckV[75] v1.0.1 (database version 1.5) for genome quality. Viral genome selection from contigs was based on a minimum length of 2 kb (for sequences with direct or inverted terminal repeats) or 4 kb (for the remaining sequences) for *de-novo* assembled contigs. Viral genomes were clustered into viral Operational Taxonomic Units (vOTUs) according to the MIUViG guidelines[76] (95% average nucleotide identity, 85% aligned fraction) using the supporting scripts anicalc.py and aniclust.py of CheckV. Contigs identified as proviral by either geNomad or CheckV were also clustered into vOTUs and represent the set of integrated proviruses. Lifestyle of phage vOTUs whose genome was assembled in contigs entirely viral was predicted with BACPHLIP[77]. Host (bacterial and archaeal) and virus interactions were computed using the iPHoP[78] v1.3.3 (parameters:"default") integrated pipeline and visualized using plotly v2.29.1. When multiple predictions were available for a query virus, the host with the highest confidence score was selected as the representative interaction. Reads were mapped to vOTUs with Bowtie2[53] v2.4.5, and abundances were expressed as TPM with CoverM[65] v0.6.1. TPM normalization accounts for contig length and total sample reads, including both viral and prokaryotic.

The intermediate output of geNomad ("virus_genes.tsv"), containing the annotation results for all predicted viral genes, was used to screen vOTUs. Specific attention was given to manually curating key enzyme classes involved in distinct stages of the viral lytic cycle, including glycosidases, endopeptidases, amidases, holins, and lysins/endolysins. Information on gene function was retrieved using Pfam (v37.2), KEGG orthology (v112.0), and TIGRFAM (v15.0) databases. A spacer collection was constructed by extracting the spacers from the metagenomic reads using SpacerExtractor v0.9.1. These spacers were then aligned to MAGs using Bowtie v1.38, generating a high-confidence spacer database. Subsequently, vOTUs were matched to the spacer database via blastn[72] v2.9.0 (parameters: "-max_target_seqs = 1000 -word_size = 8 -dust = no") and only considering alignments with at least 25 bp and <2 mismatches, covering ≥95% of the spacer length. After these filtering criteria, a total of 1541 CRISPR spacers matched 38 vOTUs. Conversely, when mapping spacers onto phage genomes, the mismatch filter was removed to trace the temporal evolution of spacer acquisition, thereby including spacers with up to three mismatches. With the updated cut-offs a total of 3329 CRISPR spacers matched 30 phage vOTUs.

Genome-scale metabolic modelling

Genome-scale metabolic models (GEMs) were generated from MAGs with a RA higher than 0.1% in at least one metagenomic sample ($n = 101$) and representing on average 98% of the community. For samples with paired metatranscriptomic data, a 0.1% threshold of combined MAG and transcriptome RA was applied as detailed before[59], for a total average combined abundance of 94%. Draft GEMs were reconstructed and gap-filled using gapseq v1.2[79] starting from the open reading frames (Prodigal) and searching homologous sequences (-l all). Growth medium prediction and GEM gap-filling were individually performed for every taxon excluding oxygen (-c "cpd00007:0") and assuming a H_2-rich environment (-e highH2). If needed, consumption of H_2 and CO_2 or H_2, CO_2 and acetate was allowed to achieve a positive growth rate.

Construction of community models and simulation of associated metabolism were performed using MICOM v0.37.0[80]. Microbial

growth was estimated with cooperative tradeoff FBA analysis based on measured input gas consumption rates, volatile fatty acid concentration and approximate medium composition (Supplementary Data 2). In the longitudinal analysis over decreasing GRTs, these data provided the sole constraints on top of those defined within FBA and CH_4 production rate was calculated by minimizing the sum of total flux (pfba = True). Phage bloom and bioaugmentation processes were incorporated into the simulations by modifying the medium composition. For the phage infection, bacterial reduction before and after the viral explosion was modeled as cell lysis. The proportion of lysed intracellular and extracellular compounds, calculated based on the reduction rate of each bacterium, was added to the medium for sample 1hA_L. In the bioaugmentation scenario the inoculum used was stored and maintained from the 2h_L culture, and then applied to 1hB_L. Therefore, the microbial community in the bioaugmentation culture was identical to that of 2h_L, and a proportion of extracellular components was added to the medium of 1hB_L to reflect both influx and efflux dynamics. This adjustment resulted in the rapid consumption of specific components by the community, thereby shifting the metabolic balance. Finally, CoCo was utilized to constrain the intracellular reaction fluxes of individual taxa along with medium and biochemical data[26]. The parameters, including the FBA tradeoff, were selected by maximizing the correlation between obtained growth rates and replication rates estimated with CoPTR, resulting in a significant correlation (Pearson's $r = 0.26$, p-value $= 5 \cdot 10^{-3}$) for the predicted peak-to-trough ratios (PTR). To estimate the metabolic potential in the different components of the cultivation vessels, we implemented a flux variability analysis while setting the growth rates to those obtained by FBA.

Reporting summary

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

Data availability

The raw sequencing data generated in this study have been deposited in the NCBI database under accession code PRJNA1272763. Source data are provided with this paper.

Code availability

Code used to analyze the data are available at [https://github.com/MetabioinfomicsLab/SNVs_paper].

References

1. Pavlopoulos, G. A. et al. Unraveling the functional dark matter through global metagenomics. *Nature* **622**, 594–602 (2023).
2. De Vrieze, J. The next frontier of the anaerobic digestion microbiome: from ecology to process control. *Environ. Sci. Ecotechnol.* **3**, 100032 (2020).
3. Tsapekos, P. et al. Pilot-scale biomethanation in a trickle bed reactor: process performance and microbiome functional reconstruction. *Energy Convers. Manag.* **244**, 114491 (2021).
4. Tsapekos, P., Alvarado-Morales, M. & Angelidaki, I. H2 competition between homoacetogenic bacteria and methanogenic archaea during biomethanation from a combined experimental-modelling approach. *J. Environ. Chem. Eng.* **10**, 107281 (2022).
5. Centurion, V. B. et al. A unified compendium of prokaryotic and viral genomes from over 300 anaerobic digestion microbiomes. *Environ. Microb.* **19**, 1 (2024).
6. De Bernardini, N. et al. Integrating metagenomic binning with flux balance analysis to unravel syntrophies in anaerobic CO_2 methanation. *Microbiome* **10**, 117 (2022).
7. Zampieri, G. et al. Core cooperative metabolism in low-complexity CO2-fixing anaerobic microbiota. *ISME J.* **19**, wraf017 (2025).
8. Ebrahimian, F. et al. Effect of pressure on biomethanation process and spatial stratification of microbial communities in trickle bed reactors under decreasing gas retention time. *Bioresour. Technol.* **361**, 127701 (2022).
9. Chen, D. W. & Garud, N. R. Rapid evolution and strain turnover in the infant gut microbiome. *Genome Res.* **32**, 1124–1136 (2022).
10. Vasquez, K. S. et al. High-resolution lineage tracking of within-host evolution and strain transmission in a human gut symbiont across ecological scales. Preprint at *bioRxiv* https://doi.org/10.1101/2024.02.17.580834 (2024).
11. Zhao, S. et al. Adaptive evolution within gut microbiomes of healthy people. *Cell Host Microbe* **25**, 656–667.e8 (2019).
12. Fay, J. C. & Wu, C.-I. Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).
13. Charlesworth, B., Morgan, M. T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).
14. Correa, A. M. S. et al. Revisiting the rules of life for viruses of microorganisms. *Nat. Rev. Microbiol.* **19**, 501–513 (2021).
15. Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R. C. & San Millán, Á. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nat. Rev. Microbiol.* **19**, 347–359 (2021).
16. Fernández, L., Rodríguez, A. & García, P. Phage or foe: an insight into the impact of viral predation on microbial communities. *ISME J.* **12**, 1171–1179 (2018).
17. Rossi, A. et al. Analysis of the anaerobic digestion metagenome under environmental stresses stimulating prophage induction. *Microbiome* **10**, 125 (2022).
18. Georjon, H. & Bernheim, A. The highly diverse antiphage defence systems of bacteria. *Nat. Rev. Microbiol.* **21**, 686–700 (2023).
19. Crespo-Bellido, A. & Duffy, S. The how of counter-defense: viral evolution to combat host immunity. *Curr. Opin. Microbiol.* **74**, 102320 (2023).
20. Hille, F. et al. The biology of CRISPR-Cas: backward and forward. *Cell* **172**, 1239–1259 (2018).
21. Wu, L. et al. Diversity-generating retroelements: natural variation, classification and evolution inferred from a large-scale genomic survey. *Nucleic Acids Res.* **46**, 11–24 (2018).
22. Pawluk, A., Davidson, A. R. & Maxwell, K. L. Anti-CRISPR: discovery, mechanism and function. *Nat. Rev. Microbiol.* **16**, 12–17 (2018).
23. Malone, L. M., Birkholz, N. & Fineran, P. C. Conquering CRISPR: how phages overcome bacterial adaptive immunity. *Curr. Opin. Biotechnol.* **68**, 30–36 (2021).
24. Meyer, J. R. et al. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **335**, 428–432 (2012).
25. Amundson, K. K., Roux, S., Shelton, J. L. & Wilkins, M. J. Long-term CRISPR locus dynamics and stable host-virus co-existence in subsurface fractured shales. *Curr. Biol.* **33**, 3125–3135.e4 (2023).
26. Zampieri, G., Campanaro, S., Angione, C. & Treu, L. Metatranscriptomics-guided genome-scale metabolic modeling of microbial communities. *Cell Rep. Methods* **3**, 100383 (2023).
27. Nobu, M. K. et al. Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME J.* **9**, 1710–1722 (2015).
28. Balomenou, S. et al. Distinct functions of polysaccharide deacetylases in cell shape, neutral polysaccharide synthesis and virulence of Acillus anthracis. *Mol. Microbiol.* **87**, 867–883 (2013).
29. Whitfield, C., Wear, S. S. & Sande, C. Assembly of bacterial capsular polysaccharides and exopolysaccharides. *Annu. Rev. Microbiol.* **74**, 521–543 (2020).
30. Hu, C. et al. Mechanism for Cas4-assisted directional spacer acquisition in CRISPR–Cas. *Nature* **598**, 515–520 (2021).
31. Dhingra, Y., Suresh, S. K., Juneja, P. & Sashital, D. G. PAM binding ensures orientational integration during Cas4-Cas1-Cas2-mediated CRISPR adaptation. *Mol. Cell* **82**, 4353–4367.e6 (2022).

32. Chatzis, A. et al. Comparative study on packing materials for improved biological methanation in trickle Bed reactors. *Bioresour. Technol.* **385**, 129456 (2023).
33. Chen, L. et al. A critical review on bioaugmentation assisted anaerobic digestion for methane production: Performances, microbiome-functionalities and challenges. *J. Environ. Manag.* **380**, 125127 (2025).
34. Zhang, J., Kasciukovic, T. & White, M. F. The CRISPR associated protein Cas4 is a 5′ to 3′ DNA exonuclease with an iron-sulfur cluster. *PLoS ONE* **7**, e47232 (2012).
35. Khrustaleva, T. A. Secondary structure preferences of Mn2+ binding sites in bacterial proteins. *Adv. Bioinforma.* **2014**, 501841 (2014).
36. Heler, R., Wright, A. V., Vucelja, M., Doudna, J. A. & Marraffini, L. A. Spacer acquisition rates determine the immunological diversity of the type II CRISPR-Cas immune response. *Cell Host Microbe* **25**, 242–249.e3 (2019).
37. Levin, B. R., Moineau, S., Bushman, M. & Barrangou, R. The population and evolutionary dynamics of phage and bacteria with CRISPR–mediated immunity. *PLoS Genet.* **9**, e1003312 (2013).
38. Datsenko, K. A. et al. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**, 945 (2012).
39. Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**, 1047–1050 (2008).
40. Laanto, E., Hoikkala, V., Ravantti, J. & Sundberg, L.-R. Long-term genomic coevolution of host-parasite interaction in the natural environment. *Nat. Commun.* **8**, 111 (2017).
41. Zhang, A.-N. et al. CRISPR-Cas spacer acquisition is a rare event in human gut microbiome. *Cell Genomics* **5**, 10072 (2025).
42. Guillemet, M. et al. Competition and coevolution drive the evolution and the diversification of CRISPR immunity. *Nat. Ecol. Evol.* **6**, 1480–1488 (2022).
43. Elliott, J. F. K. et al. Conditions for the spread of CRISPR-Cas immune systems into bacterial populations. *ISME J.* **18**, wrae108 (2024).
44. Meaden, S. et al. Phage gene expression and host responses lead to infection-dependent costs of CRISPR immunity. *ISME J.* **15**, 534–544 (2021).
45. Thingstad, T. F. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol. Oceanogr.* **45**, 1320–1328 (2000).
46. Breitbart, M., Bonnain, C., Malki, K. & Sawaya, N. A. Phage puppet masters of the marine microbial realm. *Nat. Microbiol.* **3**, 754–766 (2018).
47. López-Beltrán, A., Botelho, J. & Iranzo, J. Dynamics of CRISPR-mediated virus–host interactions in the human gut microbiome. *ISME J.* **18**, wrae134 (2024).
48. Piel, D. et al. Phage–host coevolution in natural populations. *Nat. Microbiol.* **7**, 1075–1086 (2022).
49. Shaer Tamar, E. & Kishony, R. Multistep diversification in spatio-temporal bacterial-phage coevolution. *Nat. Commun.* **13**, 7971 (2022).
50. Anagnostopoulou, C. et al. Valorization of household food wastes to lactic acid production: a response surface methodology approach to optimize fermentation process. *Chemosphere* **296**, 133871 (2022).
51. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
52. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes de novo assembler. *Curr. Protoc. Bioinforma.* **70**, e102 (2020).
53. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
54. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ.* **3**, e1165 (2015).
55. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ.* **7**, e7359 (2019).
56. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
57. Nissen, J. N. et al. Improved metagenome binning and assembly using deep variational autoencoders. *Nat. Biotechnol.* **39**, 555–560 (2021).
58. Mainguy, J. & Hoede, C. Binette: a fast and accurate bin refinement tool to construct high quality metagenome assembled genomes. *J. Open Source Softw.* **9**, 6782 (2024).
59. Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat. Methods* **20**, 1203–1212 (2023).
60. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
61. Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
62. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022).
63. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma.* **11**, 119 (2010).
64. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
65. Aroney, S. T. N. et al. CoverM: read alignment statistics for metagenomics. *Bioinformatics* **41**, btaf147 (2025).
66. Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E. & Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943–2945 (2022).
67. Pasternak, Z. et al. By their genes ye shall know them: genomic signatures of predatory bacteria. *ISME J.* **7**, 756–769 (2013).
68. Olm, M. R. et al. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat. Biotechnol.* **39**, 727–736 (2021).
69. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol*. https://doi.org/10.1038/s41587-023-01953-y (2023).
70. Tesson, F. et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.* **13**, 2561 (2022).
71. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
72. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinforma.* **10**, 421 (2009).
73. Meng, E. C. et al. UCSF ChimeraX: tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
74. Ben Chorin, A. et al. ConSurf-DB: an accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. *Protein Sci.* **29**, 258–267 (2020).
75. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021).
76. Roux, S. et al. Minimum information about an uncultivated virus genome (MIUViG). *Nat. Biotechnol.* **37**, 29–37 (2019).
77. Hockenberry, A. J. & Wilke, C. O. BACPHLIP: predicting bacteriophage lifestyle from conserved protein domains. *PeerJ* **9**, e11396 (2021).

78. Roux, S. et al. iPHoP: an integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol.* **21**, e3002083 (2023).
79. Zimmermann, J., Kaleta, C. & Waschina, S. gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biol.* **22**, 81 (2021).
80. Diener, C., Gibbons, S. M. & Resendis-Antonio, O. MICOM: metagenome-scale modeling to infer metabolic interactions in the gut microbiota. *mSystems* https://doi.org/10.1128/msystems.00606-19 (2020).

Acknowledgements

Gabriele Ghiotto acknowledges the project LIFE20 CCM/GR/001642–LIFE CO2 to CH4 of the European Union LIFE + program for financially supporting his salary. Acknowledgments also to the European Union's Horizon 2020 research and innovation program under grant agreement No 101084405 (CRONUS) for the resources provided. The work conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231 (N.C.K., A.P.C., S.R.).

Author contributions

G.G.: Conceptualization, software, formal analysis, data curation, writing–original draft, writing–review & editing, visualization. G.Z.: Writing–original draft, software, data curation. E.O.: Software, data curation. A.C.: Investigation, data curation. P.G.K.: Resources, supervision, funding acquisition. A.P.C.: Supervision, Software. S.R.: Writing–review & editing, supervision, software. S.C.: Writing–review & editing, conceptualization, resources, supervision, funding acquisition. N.C.K.: Conceptualization, resources, supervision. L.T.: Conceptualization, resources, supervision, project administration, funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-73084-2.

Correspondence and requests for materials should be addressed to S. Campanaro.

Peer review information *Nature Communications* thanks Antun Skanata and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.