nature communications

Article

https://doi.org/10.1038/s41467-026-71833-x

Quantitative stable isotope probing (qSIP)-informed metagenomics identifies viruses infecting chemoautotrophs

Received: 22 September 2025

Accepted: 27 March 2026

Published online: 14 April 2026

Check for updates

Elaine Luo[1,2] ✉, Ngoc D. Pham[1,2], Timothy J. Rogers[1,2], Md Moinuddin Sheam[1,2], Bayleigh E. Benner[3], Joseph J. Vallino[4], Gareth Trubl[5] & Julie A. Huber[6]

Aquatic environments absorb ~2.5 gigatonnes of atmospheric carbon each year[1], more than the carbon stored in the atmosphere, soils, and all biomass combined. Primary producers transform this dissolved inorganic carbon into biomass that can subsequently flow into other trophic levels, or be released back into the environment through viral lysis. While there is substantial knowledge about the diversity and activity of viruses infecting photoautotrophic primary producers and the ecosystem impact, little is known about viruses infecting chemoautotrophs, representing a gap in our understanding of key processes driving microbial carbon cycling. Here, we combine metagenomics with quantitative $^{12/13}$C stable isotopic probing (qSIP) mesocosm experiments in a marine-derived meromictic pond to quantify population-specific isotopic enrichment, identify key chemoautotrophic primary producers, and virus-host dynamics. Isotopically enriched carbon is tracked from the genomes of chemoautotrophs to putative viruses, showing that active populations of hydrogen/sulfur-oxidizing chemoautotrophs (*Thiomicrorhabdus, Hydrogenovibrio, Sulfurimonas, Sulfurovum*) are targeted by viruses. This work provides the foundation for revealing the diversity and activity of viruses infecting globally-widespread chemoautotrophs. Our study sheds light on trophic interactions that impact microbial carbon cycling in aphotic environments and builds toward biogeochemical models that incorporate viral impacts on chemoautotrophic microbial communities.

The diversity, mechanisms, and processes governing microbial primary production and the recycling of autotrophic biomass are fundamental to our planet's carbon cycle. These processes have implications for carbon sequestration, ocean biogeochemistry, and the overall balance of carbon dioxide in the atmosphere. Beneath the Earth's sunlit layer, primary production is driven by microbial chemoautotrophs that derive energy from the oxidation of reduced compounds, such as hydrogen and sulfur, to form the base of the food web. Growing evidence suggests that aquatic ecosystems fueled by chemoautotrophy are widely distributed on Earth, ranging from

[1]Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, NC, USA. [2]Center for Computational Intelligence to Predict Health and Environmental Risks, University of North Carolina at Charlotte, Charlotte, NC, USA. [3]Department of Biological and Physical Sciences, Johnson & Wales University, Providence, RI, USA. [4]Ecosystems Center, Marine Biological Laboratory, Woods Hole, MA, USA. [5]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA. [6]Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. ✉e-mail: elaine.luo@charlotte.edu

beneath ice shelves to coastal upwelling regions to oxygen minimum zones, deep-sea hydrothermal vents and cold seeps, groundwater, and meromictic ponds and lakes[2–7]. Studying microbial processes regulating chemoautotrophic primary production and trophic interactions regulating chemoautotrophs is fundamental to our understanding of microbial carbon cycling.

While the diversity, function, and activity of viruses targeting photoautotrophs have been well-described across aquatic ecosystems[8], we have little understanding of viruses infecting chemoautotrophs. Chemoautotrophy drives primary production beneath the earth's surface layer, where aphotic environments account for the majority of aquatic environments by volume. Viruses are a major source of cellular mortality and carbon cycling in aquatic environments[9–12]. Viral lysis is estimated to transform ~150 gigatonnes of carbon annually from biomass back into the environment, equivalent to ~25 times that of the ocean's biological carbon pump[13,14]. Despite recognition of the important role of viruses in aquatic habitats, there is a large gap in our understanding of viruses infecting chemoautotrophs, such as common sulfur and/or hydrogen oxidizing bacteria in the genera *Thiomicrorhabdus, Hydrogenovibrio, Sulfurimonas, Sulfurovum*. These genera contribute to primary production through sulfur and/or hydrogen oxidation, and are globally-distributed across sulfidic aquatic and terrestrial environments such as in oxygen minimum zones, deep-sea hydrothermal vents, meromictic lakes, sediments, springs, and oil caves[2,4,5,15–22]. In particular, *Sulfurimonas* accounted for a significant proportion (20–70%) of bacterial assemblages recovered from the Baltic Sea redoxicline and deep-sea hydrothermal vents, as well as most of the chemoautotrophic activity measured in these environments[20,21,23–25]. Unlike viruses infecting aquatic photoautotrophs, very few viruses infecting these key chemoautotrophs have been isolated in culture, limiting our ability to detect them in environmental sequences. In this study, we show that viruses are likely not merely passive players but active agents in targeting these productive chemoautotrophs, fundamentally shaping how we view carbon and nutrient cycling in redox-active ecosystems.

Although metagenomic studies have the potential to discover novel microbial diversity, linking viruses to their biogeochemical impacts remains an ongoing goal and challenge in environmental microbiology[12,26–28]. Putative viral sequences have been detected in the genomes of *Sulfurimonas* and in other chemoautotrophs across oxygen minimum zones, mesopelagic open ocean, and deep-sea hydrothermal vents[29–37]. Lysogeny was hypothesized to be the dominant mode of infection, based on metagenomic predictions on microbial communities at deep-sea hydrothermal vents and in the mesopelagic ocean[29–32], suggesting limited impact of viral lysis in these communities. These in silico predictions, however, cannot definitively link a virus to its host, cannot determine whether a virus or host is active, and cannot identify whether this interaction is ecologically or biogeochemically relevant. Whether and which viruses actively target chemoautotrophs in aquatic environments remains to be determined. These challenges highlight the need to develop new experimental approaches leveraging high-throughput metagenomics to identify viruses infecting environmentally-relevant chemoautotrophs.

Here, we combine qualitative stable isotope probing (qSIP) and metagenomics to track the flow of carbon across natural microbial populations. This approach provides critical mechanistic links between novel genomic diversity, function, and activity in ecosystem processes[38–41]. Since chemoautotrophs assimilate dissolved inorganic carbon (DIC) into their biomass during carbon fixation, we used ^{13}C-DIC as a stable isotope tracer to target chemoautotrophs and their viruses. We postulate that ^{13}C-DIC incorporation into genomes measures primary production, as it reflects the net carbon that has been assimilated into biomass after carbon fixation. During growth and replication, ^{13}C is incorporated into the genomes of active chemoautotrophs, resulting in ^{13}C-enriched DNA. Experiments on a few viral isolates indicate that viruses mostly use host nucleotides for their genomes during replication[42], an expectation that has been previously established using broader ^{13}C isotopic enrichment data[40]. As a result, we expect that viruses infecting productive chemoautotrophs assimilates ^{13}C-enriched host nucleotides into their own genomes. We utilized this established relationship[40] as a reference-independent approach to link viruses to chemoautotrophic hosts.

We conducted $^{12/13}$C-DIC mesocosm experiments using on samples collected from the chemocline of a marine-derived meromictic lake as a model ecosystem for chemoautotrophic communities (Figs. S1, S2). Meromictic ecosystems (bodies of water that do not mix) contain sharp redox gradients that are required to support chemoautotrophic microbial communities, and sulfur-oxidizing bacteria such as *Sulfurimonas* have been previously reported from these systems[16]. We sampled from a salt-derived coastal lake that contained stable stratification from freshwater and marine inputs that contained sharp physiochemical gradients in the top 10 m (Fig. S1) and an energy source (hydrogen sulfide) at 8–10 m[43,44] to support chemoautotrophy. Following incubation, microbial samples were collected on cell-enriched (> 0.22 µm, no pre-filter) and virus-enriched (0.02–0.22 µm) size fractions. Metagenomic DNA from both size fractions were density-fractionated (Fig. S3) to separate heavy (^{13}C-enriched) from light (^{13}C natural abundance) DNA to identify microbial populations that have assimilated ^{13}C into their genomes. Prior SIP-metagenomic studies have generally sequenced only the light and heavy fractions of DNA density gradients post-incubation, which enables a binary distinction between organisms that have and have not taken up the substrate with a predicted threshold. Here, we sequenced five density fractions from each sample (Fig. S3) to account for variability in DNA buoyant density (*e.g.*, due to differences in GC content) and perform qSIP to quantify isotopic enrichment for each microbial population.

We used differences in density between the ^{12}C control and ^{13}C enriched treatment to quantify isotopic enrichment using a previously established metric, the ^{13}C Atom Fraction Excess or Excess Atom Fraction (EAF)[45,46]. ^{13}C-EAF measures the fraction of ^{13}C that is in excess of natural abundances and measures the extent of isotopic assimilation. It reflects the in situ activity of environmental microbial populations for a targeted substrate[45,46]. By incubating environmental samples with $^{12/13}$C-DIC and quantifying the ^{13}C-EAF of microbial populations, we identified key active chemoautotrophs responsible for primary production in this ecosystem. We linked putative viral sequences to their hosts based on similarities in their isotopic enrichment and identified viruses that infect productive chemoautotrophs. Through targeted sampling of viral populations in both cell-enriched (> 0.22 µm) and virus-enriched (0.02–0.22 µm) samples, our observations suggest that these viruses actively cycle carbon through the production of new viral particles.

Results and Discussion

Key microbial drivers of chemoautotrophic primary production

Our results indicate that, under experimental conditions, chemoautotrophic primary production was driven by a few rare but highly productive microbial genera. At the community level, we quantified the ^{13}C isotopic enrichment of > 0.45 µm bulk particulate organic carbon using gas chromatography mass spectrometry, showing that it increased from day 0 to 7 of incubation (Fig. S4). At the population level, we quantified isotopic enrichment using ^{13}C-EAF[45], for 28,788 cellular contigs and 187 high-confidence viral population genomes[47] (Tables S1, S2). Significant isotopic enrichment was detected in 3193 cellular contigs and 35 high-confidence viral population genomes (^{13}C-EAF > 0.049, Fig. 1). While most microbial populations demonstrated insignificant isotopic enrichment, a small subset of microbial populations were highly productive (Fig. 1). The genera with the highest primary production were sulfur and/or hydrogen-oxidizing bacteria: *Thiomicrorhabdus* and *Hydrogenovibrio* (phylum: Pseudomonadota),

https://doi.org/10.1038/s41467-026-71833-x

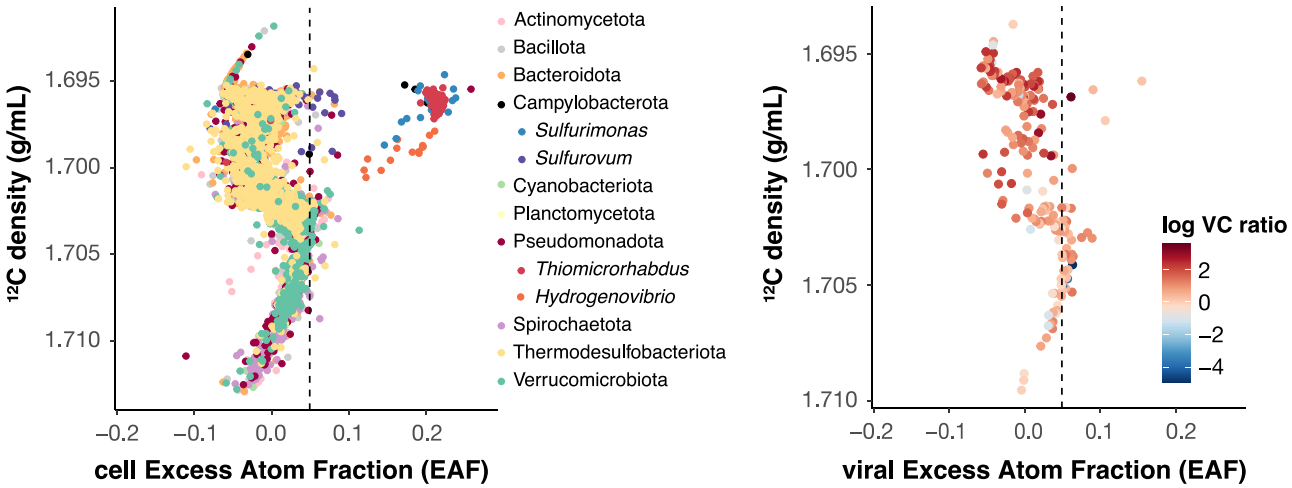

Fig. 1 | Isotopic enrichment of cellular and viral populations. Isotopic enrichment is determined by the contigs' ^{13}C Excess Atom Fraction (^{13}C-EAF), plotted relative to the contigs' average density in the ^{12}C control, of cellular contigs from metagenome-assembled genomes (MAGs, left) and viral population genomes (right). Cellular contigs are colored by their MAG phylum- or genus- (italic) level assignments from the Genome Taxonomy Database, and only phyla with >100 contigs are visualized. Contigs above the dashed line indicated significant carbon incorporation (^{13}C-EAF > 0.049). Viral populations are colored by the log-ratio of their relative abundance in the virus-enriched to cell-enriched sample (VC ratio) in the environmental sample, approximating their reproductive strategy. Prophages are expected to have near-zero extracellular presence (negative log VC ratio), while populations producing viral particles are expected to have detectable extracellular presence (higher log VC ratio). Source data are provided as a Source Data file.

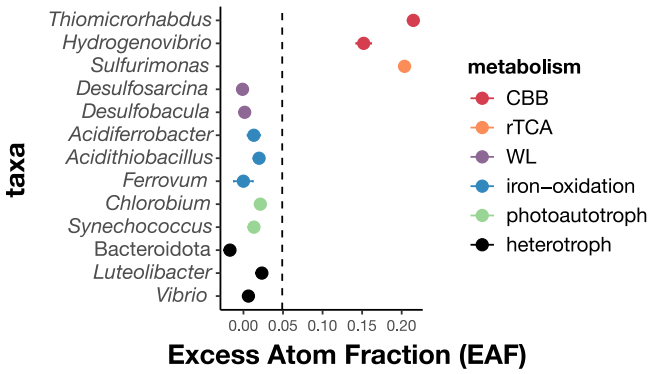

Fig. 2 | Isotopic enrichment as determined by the ^{13}C Excess Atom Fraction (EAF) of cellular MAGs. MAGs encoding the following carbon fixation pathways are shown: Calvin-Benson-Bassham cycle (CBB), reductive tricarboxylic acid cycle (rTCA), and Wood-Ljungdahl (WL). The isotopic enrichment of contigs from iron-oxidizing chemoautotrophs, photoautotrophs, and heterotrophs are shown below. The circle represents the mean ^{13}C-EAF of all contigs in that MAG or taxa and the line represents the standard error for all contigs within that taxonomic group (technical replicates), both color-coded by its metabolism. Genus-level taxonomic assignments are shown in italics, while the phylum-level taxonomic assignment is not. The dashed line represents significant isotopic enrichment (^{13}C-EAF > 0.049). Source data are provided as a Source Data file.

and *Sulfurimonas* and *Sulfurovum* (phylum: Campylobacterota), with ^{13}C-EAF averaging ~0.20 (i.e., ^{13}C accounts for 20% of total C in excess of natural abundance). The relative abundances of these four groups combined accounted for a small fraction (0.13%) of the total prokaryotic assemblages recovered from this environment, suggesting that the importance of these rare but key taxa would have been overlooked based on metagenomic sequencing alone.

Although *Thiomicrorhabdus* and *Sufurimonas* have been previously reported in the water column of meromictic ecosystems and hypothesized to contribute to sulfur cycling and carbon fixation[16,48–50], *Hydrogenovibrio* and *Sulfurovum* have not. Here, we show that they are not only present, but actively fixing carbon and driving to primary production. Differences in primary production were observed even within taxonomically closely-related populations within the same genus (Fig. S5). For example, 96%, 30%, 60%, and 9% of contigs identified as *Thiomicrorhabdus*, *Hydrogenovibrio*, *Sulfurimonas*, and *Sulfurovum*, respectively, showed high isotopic enrichment (^{13}C-EAF >= 0.2); while 0.6%, 0%, 24%, and 81%, respectively, did not show significant isotopic enrichment (^{13}C-EAF <= 0.049). These results indicate that primary production in this system is driven by a small subset of highly productive chemoautotrophic populations that likely serve as keystone species for microbial carbon cycling. Given that species richness was comparable pre-and post-incubation (Table S3), we hypothesize that these experimental results were generalizable to the sampled environment. Despite similar species richness, a change in species composition can influence the microbial processes observed. However, our findings are based on isotopic enrichment data (%^{13}C in genome), which is an independent from the population's relative abundance in a particular sample. As a result, we do not expect shifts in relative abundances (standing stock) from potential experimental bottle effects to skew our observed genomic isotopic enrichment data (rate of ^{13}C incorporation during incubation).

The most abundant prokaryotic populations showed insignificant isotopic enrichment (Fig. 2). Contigs assigned to the genus *Chlorobium*, *Synechecoccus*, *Desulfosarcina*, *Desulfobacter*, and *Desulfonema* respectively, accounted for 4.5%, 4.0%, 3.4%, 1.1%, and 1.1% of the cellular assemblages recovered from the environmental sample. Although both *Chlorobium* and *Synechecoccus* are common primary producers across diverse aquatic ecosystems, respectively dominating anoxic and oxic depths of meromictic ecosystems[50–57], both showed undetectable ^{13}C enrichment in the dark experimental conditions (Fig. 2). We do not expect that these photoautotrophs contribute to primary production in the environment sampled, given that photosynthetically active radiation (PAR) at the time and depth of sampling, on a sunny day, was near-zero (0.04 µmol/m²/s). Even extremely low-light adapted *Chlorobium* strains isolated from the Black Sea could not grow at PAR levels of 0.25 µmol/m²/s[58]. Although the lowest reported minimum light requirements for phototrophic carbon fixation was 0.015 µmol/m²/s in a putative *Chlorobiacaea*, its rate of carbon fixation at these light levels estimated its doubling time at 26 years[59]. Furthermore, 0.04 µmol/m²/s is 140 times lower than the lowest reported minimum light requirements for net *Synechococcus* photosynthesis[60].

https://doi.org/10.1038/s41467-026-71833-x

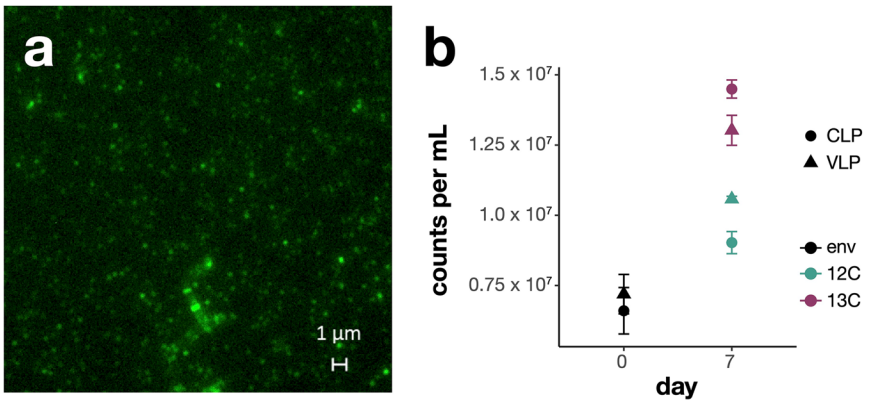

Fig. 3 | Visualization of cell-like particles (CLPs) and virus-like particles (VLPs). **a** Epifluorescence micrograph of 0.22 µm-prefiltered samples taken at the termination of incubation on day 7 from the ^{12}C sample indicates the presence of VLPs. **b** Biomass quantification from flow cytometry (mean +/− standard error of three technical replicates) shows an increase of both CLPs and VLPs in both treatments (12 C, 13 C) during the 7-day incubation period relative to the environmental control (env). Source data are provided as a Source Data file.

Taken together, we reason that the two most dominant autotrophs may have been vertically transported from upper layers and were irrelevant to primary production at the time and depth of sampling, a conclusion supported by their lack of isotopic enrichment. Our results show that dominant taxa, as recovered by metagenomic sequencing, may contribute little to key ecosystem processes in the habitat sampled. This finding has broad implications for environmental studies that rely heavily on metagenomic sequences to identify microbial function and infer their ecological and biogeochemical impacts.

Current limitations on linking viruses to active chemoautotrophs using in silico predictions

In silico predictions of viral taxonomy and function yielded limited information. Taxonomic identification for novel environmental viruses is hindered by the sparse representation of uncultivated viruses in reference databases. 31,448 non-redundant low-quality to complete putative viral populations were recovered with a sequence length of ≥ 5 kbp. Using a marker-protein-based annotation program, the majority (92%) of putative viral population genomes received only class-level annotations (Fig. S6) and were unidentifiable at the order level and beyond. None received a genus-level annotation. Using protein sequence similarity searches against reference databases, only 40 (0.1%) of putative viral population genomes received a genus-level annotation, all broadly related to eukaryotic *Ostreococcus* and *Micromonas* viruses. 277 (0.9%) of total putative viral populations were identified as temperate phages via marker genes and alignments to putative prophages in contigs from cell-enriched samples. Using metagenomic virus-host linkage methods, such as alignments to prophages in cellular assemblies and CRISPR spacers, only 294 (0.9%) were linked to cellular contigs (Fig. S6), and 205 (0.7%) of these yielded taxonomic annotations. Taxonomic identification of linked host contigs revealed that these putative viruses potentially target bacteria in the genus *Methylococcus*, *Spirochaeta*, *Desulfuromonas*, *Desulfobacter*, and *Chlorobium*, ordered from high to low isotopic enrichment of viral population genomes. In silico host predictions did not link any viruses to active primary producers in this environment (^{13}C-EAF > 0.049). These results highlight the current limitations of in silico approaches in identifying the taxonomy and function of novel viral diversity, which remains a major challenge in the field of environmental microbiology.

Identifying viruses infecting active chemoautotrophs using their isotopic signatures

Despite the absence of informative taxonomic or functional annotations, we linked viral sequences to key chemoautotrophs using similarities in the isotopic signatures of viral and host genomes. We show that similarities in isotopic signatures can be utilized to overcome current limitations as a reference-independent method to link active viral populations to their hosts. The isotopic enrichment of putative viral populations exhibited a strong linear correlation to the isotopic enrichment of their predicted hosts ($N = 15$, P-value = 1.55e-7), as demonstrated by high-confidence viral population genomes that yielded host predictions through prophage and CRISPR spacer mapping (Fig. S7). This pattern is consistent with a previous report[40] and reflects expectations that viruses mostly incorporate host nucleotides during replication[42]. Based on this expectation, we postulate that the putative viral populations with high isotopic enrichment in the cell-enriched size fraction (Fig. 1), and in particular those with the highest isotopic enrichment, targeted active chemoautotrophs in this environment (*Hydrogenovibrio*, *Thiomicrorhabdus*, *Sulfurimonas*, *Sulfurovum*).

Amongst sequences with significant isotopic enrichment, the isotopic enrichment of putative viral populations generally lagged behind cellular sequences. Since ^{13}C-EAF is a population-wide measurement, this result suggests that productive and uninfected chemoautotrophs may be incorporating ^{13}C into their nucleotides at a faster rate than infected chemoautotrophs, This interpretation is consistent with previous observations that viral infection reduces carbon fixation in photoautotrophic cyanobacteria[61]. Furthermore, putative viral populations with significant isotopic enrichment were rare (Fig. S7), accounting for 1.6% of the total viral assemblage recovered from the environmental sample, which is higher than the relative abundance of highly active chemoautotrophs (0.13%). This observation is consistent with a previous report in seawater suggesting that productive, rare taxa appear to be disproportionately targeted by viral lysis[62]. Taken together, we postulate that top-down control by viral infection is a mechanism that contributes to the low abundance of productive chemoautotrophs in this ecosystem.

Mechanism of infection

We postulate that viruses infecting active chemoautotrophs cycle carbon through the production of viral particles. Epifluorescence micrographs taken at day 7 showed the presence of virus-like particles at the end of our incubation period (Fig. 3a). Flow cytometry indicated that the abundances of cell-like and virus-like particles were higher at day 7 relative to those at day 0 (Fig. 3b), reflecting cellular replication and viral particle production during the 7-day incubation period. Both observations are consistent with viral particle production during our incubation period at the community level.

At the population level, we used the VC ratio to infer viral reproductive strategies. The VC ratio represents the log-ratio of the population's abundance in the Virus-enriched relative to the Cell-enriched

https://doi.org/10.1038/s41467-026-71833-x

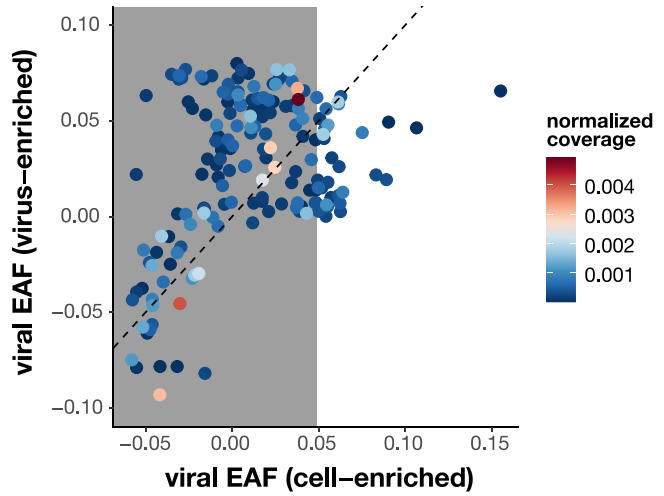

Fig. 4 | Isotopic enrichment as determined by the ^{13}C Excess Atom Fraction (EAF) of high-confidence population genomes in the virus-enriched and cell-enriched size fractions. Viral population genomes are colored by their relative abundance (normalized interquartile coverages) in the virus-enriched sample. The shaded area represents insignificant isotopic enrichment. The dashed line represents the 1:1 ratio. Source data are provided as a Source Data file.

sample. We expect that viral populations producing viral particles are observed in both size fractions, whereas passive prophages integrated into cellular genomes are found primarily in the cell-enriched size fraction. This expectation has been confirmed through previous work showing that putative prophages have lower VC ratio than putative lytic phages[29]. Using interquartile coverages, which yield more accurate coverage estimates due to removing hypervariable and conserved regions from biasing abundance calculations[63], we found that 99.8% of putative viral populations were present in the virus-enriched size fraction in the environmental control. Assuming that putative viral sequences recovered from the virus-enriched size fraction are from viral particles, this observation suggests that most putative viruses have the ability to produce viral particles in situ. Furthermore, all 35 high-confidence viral populations with significant isotopic enrichment (^{13}C-EAF > 0.049) in the cell-enriched size fraction were observed in the virus-enriched size fraction. Taken together, we postulate that viruses targeting productive chemoautotrophs were actively cycling carbon through viral particle production.

We assume that isotopically enriched putative viral sequences recovered from the virus-enriched size fraction (0.02–0.22 µm) were recovered from viral particles. However, it is possible that ^{13}C-enriched prophages could end up in the virus-enriched size fraction, in tandem with host cellular debris. To rule out this possibility, we mapped reads from all virus-enriched samples taken at day 7 to cellular contigs to detect cellular debris at the end of the incubation. More than half (1613/3193) of cellular contigs with significant isotopic enrichment (^{13}C-EAF > 0.049) were undetectable (zero interquartile coverage) across all virus-enriched samples. This result is inconsistent with the idea that cellular debris, along with its associated prophages, dominated virus-enriched samples and suggests that putative viral sequences captured in the virus-enriched size fraction were from viral particles rather than prophages. Furthermore, if isotopically-enriched viral contigs were indeed prophages from isotopically-enriched cellular debris, we would expect to find them at equal abundances in the virus-enriched size fraction. However, the average interquartile coverages for isotopically-enriched cellular contigs was ~15 times lower than that of isotopically-enriched viral contigs in the virus-enriched size fraction. As a result, we reason that isotopically-enriched viral sequences recovered from the virus-enriched size fraction largely represent sequences captured in viral particles.

Viral particles experience decay and estimated rates of turnover range from 0.036 to 30 days across aquatic environments, averaging between 1.6 and 6.1 days[64]. Taken together, we reason that isotopically-enriched viral sequences recovered in the virus-enriched size fractions represent viral particles that were recently produced from productive chemoautotrophs. These results reveal multiple lines of evidence supporting the idea that viral populations actively infect chemoautotrophs through the production of new viral particles, positioning them as key contributors to carbon cycling in these environments.

Estimating population-specific viral turnover rates

Given that we can reasonably assume that isotopically enriched viral sequences recovered from the virus-enriched size fraction (0.02–0.22 µm) were largely recovered from virus-like particles, we utilized isotopic enrichment data from both size fractions to estimate viral turnover rates at the population and the community level. We expect that if a viral population completely turned over in the viral particle pool during our incubation period, its isotopic enrichment in the virus-enriched size fraction would reflect its isotopic enrichment in the cell-enriched samples (i.e., 1:1 ratio). However, the isotopic enrichment of putative viruses in the virus-enriched samples were generally below 1:1 compared with their isotopic enrichment in the cell-enriched samples (Fig. 4, Table S4), suggesting that the standing stock of virus-like particles did not completely turn over during the 7-day incubation period.

For each putative viral population genome, we calculated its approximate of turnover time, τ, during the incubation period using the ratio of its isotopic enrichment in the virus-enriched size fraction, f_{vir}, relative to that in the cell-enriched size fraction, f_{cell}, using the formula, $\tau = -t/(\ln(1 - f_{vir}/f_{cell}))$. The most active putative viral population showed a ^{13}C-EAF of 0.066 in the virus-enriched size fraction and 0.155 in the cell-enriched size fraction, giving a f_{vir}/f_{cell} ratio of 0.426 and a turnover time of 12.6 days using the above equation. The second most active putative viral population was estimated to have a f_{vir}/f_{cell} ratio of 0.43, a turnover time of 12.5 days. The third most active putative viral population had a f_{vir}/f_{cell} ratio of 0.55, a turnover time of 8.8 days. The above approximation for τ assumes the value of f_{cell} reaches a steady state very quickly, so we also developed a non-steady-state model to estimate τ from the rates of autotrophic production and viral pool size determined from model fits to the particulate organic carbon, ^{13}C-particulate organic carbon (Fig. S4), and ^{13}C-EAF data (Fig. 1) (see Supplemental). The non-steady-state model predicted turnover times ranging from 4.9 to 7.2 days.

Based on these calculations, we estimate the turnover rates of putative viruses targeting chemoautotrophs in this environment at 5 to 13 days. This range of turnover rate estimations likely reflect lineage-specific variability in viral production and decay driven by biological characteristics and environmental conditions. These estimates are consistent with a general expectations of lower viral decay rates, which contribute to lower turnover rates, in the absence of light-induced degradation[65], and with previous estimates of turnover rates ranging from 0.036 to 30 days in aquatic habitats[64]. Whereas previous calculations of viral turnover were restricted to bulk, community-wide estimates[64], our population-specific estimates of viral turnover rates can be utilized to model microbial dynamics in complex communities.

Cross-feeding and multi-trophic interactions

As a typical consideration for tracer-based experiments, cross-feeding can allow for ^{13}C-enriched nucleotides into other non-targeted trophic groups. In our case, cross-feeding (heterotrophic assimilation of ^{13}C-organic carbon, decomposition of viruses[66], and/or anaplerotic reactions) could have allowed heterotrophic cells and their viruses to incorporate ^{13}C into their genomes. Common aquatic heterotrophs did not show significant isotopic enrichment (Fig. 2), indicating

undetectable ^{13}C-nucleotide contamination resulting from cross-feeding in our incubation timeframe of 7 days.

We observed no evidence of eukaryotic grazing of chemoautotrophs, as no environmental sequences were recovered from known eukaryotic protists. 0.00092% of reads from the environmental sample were identified as eukaryotic, all of which were classified as fungi. The majority (0.00075%) were classified *Malassezia restricta*, which has been observed across marine water column samples, deep-sea hydrothermal vents, sediments, and anoxic environments[67]. Given the rarity of other detected eukaryotic sequences, we postulate that viral lysis is the predominant top-down control on chemoautotrophs in this environment.

Metabolic pathways and primary productivity

A metagenome-assembled-genome (MAG)-level analysis of 158 medium to high-quality MAGs (Fig. S8) indicated that at incubation temperatures of 20 °C, the primary mechanisms of carbon fixation were the reductive tricarboxylic acid (rTCA) and Calvin-Benson-Bassham (CBB) pathways. While the WL pathway was recovered from sulfate-reducers (*Desulfosarcina*, *Desulfobacula*), it did not appear to contribute to carbon fixation (Fig. 2), consistent with previous reports of sulfate-reducing bacteria utilizing this pathway in reverse for anaerobic oxidation of organic matter[68]. The isotopic enrichment of MAGs with predicted rTCA pathways (*Sulfurimonas*) were similar to those with CBB (*Thiomicrorhabdus*, *Hydrogenovibrio*). Assuming similar rates of cellular turnover amongst these genera, these results suggest that carbon fixation efficiency amongst populations with rTCA was similar to those with CBB in this environment.

A gene-level analysis, normalized to a prokaryotic single-copy marker gene, indicated that genes involved in thiosulfate oxidation were enriched in cellular populations with significant isotopic enrichment (Fig. S9). Taxonomic annotations of contigs containing these genes confirmed that they were unique to genera with high isotopic enrichment (*Thiomicrohabdus*, *Hydrogenovibrio*, *Sulfurimonas*, and *Sulfurovum*). Hydrogen oxidation genes such as hydrogenases were not enriched in populations with significant isotopic enrichment (Fig. S9). Furthermore, contigs containing genes associated with hydrogen oxidation displayed lower EAF than those containing genes associated with sulfur-oxidation (Fig. S10). Our results show that, although these active genera have been reported to oxidize both sulfur and hydrogen, sulfur metabolism may be the predominant mechanism driving chemoautotrophic primary production in this ecosystem.

Conclusions and future directions

Although viruses infecting photoautotrophs have received attention across aquatic habitats, relatively little is known about viruses infecting chemoautotrophs. In this study, we identified viruses infecting chemoautotrophs (*Thiomicrorhabdus*, *Hydrogenovibrio*, *Sulfurimonas*, *Sulfurovum*), some of which are widely distributed and active primary producers across marine and terrestrial environments[2,4,5,15–24]. We observed evidence consistent with the idea that viruses actively infect chemoautotrophs through the production of viral particles. Our findings fill a gap in our understanding of virus-host dynamics and microbial processes governing carbon cycling in aphotic aquatic ecosystems. Given the broad distributions of these chemoautotrophs[2,4,5,15–22], our results highlight a previously overlooked interaction relevant to microbial carbon cycling that we postulate may be widespread across aphotic environments.

Linking microbial diversity (via genomes) to key ecosystem rates and processes (*e.g.*, carbon cycling) remains a major challenge in environmental microbiology due to sample complexity, particularly for organisms that are poorly represented in reference databases. Here, we addressed this challenge by combining quantitative $^{12/3}$C-DIC stable isotope probing (qSIP) with metagenomics to track the flow of carbon from the environment into cellular and viral genomes. We conducted experiments targeting the aphotic chemocline of a marine-derived meromictic coastal lake as a model system to identify key processes governing carbon cycling in chemoautotrophic communities. We found that primary production in this ecosystem is driven by highly active, low-abundance genera that could have likely been overlooked using metagenomic approaches alone. Despite the absence of informative taxonomic or functional annotations, we linked viruses to these key chemoautotrophs using similarities in the isotopic signatures of viral and host genomes.

We demonstrated the ability to estimate population-specific rates of viral turnover using isotopic enrichment data and anticipate that this data will be foundational for building and validating ecosystem models. This data will also be useful in benchmarking in silico metagenomic programs to enable more accurate and high-throughput predictions of virus-host linkages, as well as cross-validating with wet lab approaches, such as single-cell and hi-C sequencing, to link viruses to their hosts. Looking ahead, new developments leveraging reference-independent approaches can further predict the biogeochemical impacts of microbial populations and discover pathways underpinning key ecosystem processes. We reason that this approach is particularly useful in understudied environments to provide critical links between microbial diversity and key ecosystem processes across multiple biological scales (trophic interactions, populations, pathways, and genes).

Methods

Sampling site

Water samples for microbial analysis were collected at 10 m from Siders Pond, a year-round marine-derived coastal meromictic lake in Falmouth, Massachusetts (41.549006, −70.622039°). It is a 15 m-deep kettle hole lake typical of historically glaciated coastal regions. Its stratification is due to inputs of both freshwater and seawater, and the lack of seasonal mixing leads to a shallow chemocline that separates the photic, oxygenated upper freshwater from the aphotic, anoxic bottom saline water (Fig. S1). Anoxic conditions (< 0.2 mg/L) was observed below 8.5 m (Fig. S1, Table S6). An increase in hydrogen sulfide in the bottom saline water was previously observed at 8–10 m[43,44]. We chose a sampling depth of 10 m to target chemoautotrophic sulfur oxidizers that may be utilizing this standing stock of sulfide that was previously observed at these depths.

$^{12/13}$C-DIC incubation experiment

Water from 10 m deep in Siders Pond was collected on Nov 11th, 2021 using a handheld pump attached to a YSI probe measuring photosynthetically active radiation, dissolved oxygen, and salinity (Fig. S1). 1 L of this environmental sample was filtered using a peristaltic pump through 0.22 µm filters (Millipore-Sigma Sterivex SVGP01015) and then through a 25 mm 0.02 µm filter (Whatman Anotop WHA68092102) to respectively collect the cell-enriched and virus-enriched size fraction as environmental controls. Six samples were incubated at 20 °C in the dark in 1 L glass sealed bottles that were pre-purged by removing air, filling with nitrogen gas, and purged again prior to filling with sample fluid. Each 1 L bottle was dosed with 7.33 mL of 400 mM ^{12}C-sodium bicarbonate (^{12}C control) or ^{13}C-sodium bicarbonate (^{13}C treatment) for a final concentration of 2.9 mM. The concentration of DIC at 10 m in Siders Pond was measured at 4.11 mM on 6-Oct-2021 by students in MBL's Semester in Environmental Science program, which would yield a 41.3% ^{13}C isotopic enrichment of DIC in the ^{13}C treatment. The incubations were terminated by sequential filtration through 0.22 µm and 0.02 µm filters to respectively collect cell-enriched and virus-enriched samples at the following timepoints: 1 day (^{12}C + ^{13}C), 3 days (^{13}C), 5 days (^{13}C), and 7 days (^{12}C + ^{13}C).

Isotopic enrichment during incubation timeframe

Prior to incubation termination, 50 mL of sample was filtered onto 0.45 µm glass fiber filters, washed with 50 mL of 1x phosphorus-

buffered saline, dried in a 50 °C oven overnight in individual sterile petri plates, and analyzed on gas chromatography mass spectrometry for quantification of isotopic enrichment. We expect microbial assemblages captured on the > 0.22 µm size fraction to be similar to those captured in the > 0.45 µm size fraction, so we utilized the 0.45 µm glass fiber because it can be combusted for gas chromatography mass spectrometry (GC-MS) analysis. Filters were analyzed at the Marine Biological Laboratory Stable Isotope Laboratory for δ13C using a Europa 20-20 continuous-flow isotope ratio mass spectrometer interfaced with a Europa ANCA-SL elemental analyzer (Sercon Ltd., Cheshire, UK). From these results, the day 7 incubation ($^{12}C/^{13}C$ pair) was chosen for downstream molecular analyses. 792–15217 ng of DNA was recovered in the cell-enriched and virus-enriched samples, enabling successful density fractionation of metagenomic DNA across treatments and size fractions. All samples showed consistent, distinct peaks in DNA density between ^{12}C and ^{13}C treatments that indicated successful density fractionation (Fig. S3).

Quantification and visualization of cell-like and virus-like particles

Prior to incubation termination, duplicate 15 mL samples were fixed using electron-microscopy grade glutaraldehyde at a final concentration of 2% for 30 min at room temperature, then stored at 4 °C. Cell and virus-like particles were quantified using flow cytometry (Fig. S11). Samples were diluted 1:250 in Tris-EDTA (TE) buffer to a final volume of 500 µl. An ultrafiltrate blank was also prepared for each sample by filtering the samples through 0.02 µm Whatman anotop syringe filter (Sigma-Aldrich, USA) and diluted in TE buffer to obtain final 500 µl volume. The samples and blanks were dyed using 5 µl SYBR Gold (Fisher Scientific, USA) working stock and incubated at 80 °C for 10 min in the dark. All samples and blanks were prepared in three replicates. The samples and blanks were analyzed through Cytoflex (Beckman coulter, USA) at medium flow rate of 30 µl/min for one minute of acquisition. Virus-like particles were quantified using the violet side scatter configuration, while cells were measured by the default detector settings. The gain acquisition settings for virus-like particles were as follows: FSC 3000, SSC 1680, Violet SSC 1, B525 481, B585 202, V525 47. Also, for cell settings the gain acquisition settings were FSC 3000, SSC 1680, B525 481, B585 202, V450 34, V525 47. Separate gates were set for each sample to count cells and virus-like particles, selected to exclude background noise based on the comparison with the ultrafiltrate blank. All data were subsequently analyzed by the CytExpert software (Beckman Coulter, USA) to determine the counts of cell-like particles and virus-like particles in each sample. Epifluorescence micrographs of virus-like particles were prepared by vortexing fixed samples, filtering through a 0.22 um syringe, and strained with SYBR gold at a concentration of 1x for 30 min. 4 mL of sample was filtered using vacuum filtration onto a 25 mm 0.02 µm filter (Whatman Anodisc) and imaged on Zeiss Imager A2 at 1000x total magnification.

DNA extraction and density-gradient centrifugation

DNA was extracted from the virus-enriched (0.02 µm filters) and cell-enriched (0.22 µm filters) samples from the environmental control and 7-day $^{12}C/^{13}C$ incubation pair using the Masterpure extraction kit (Lucigen MC85200), yielding 1.8–8.4 µg and 20–25 µg of DNA respectively from the virus-enriched and cell-enriched samples. A major challenge of qSIP-metagenomics is generating sufficient DNA biomass to enable successful density-fractionation and sequencing of multiple density fractions per sample. This consideration is particularly challenging for low-biomass samples, such as the viral-particle size fraction that we targeted in this study. Here, we show that density-fractionation of virus-enriched samples (0.02–0.22 µm) and quantification of viral activity can be successfully achieved with 1 L of aquatic samples.

Both virus-enriched and cell-enriched samples from the ^{12}C and ^{13}C treatments were density-fractionated as previously described in ref. 69 with the following modifications. 500 ng of input DNA was utilized as input DNA. The input DNA was diluted using gradient buffer to 700 µL total, added to 4.8 mL of 1.77 g/mL cesium chloride solution, and individually loaded into 5.1 mL centrifuge tubes (Beckman Coulter Quick-Seal). Virus-enriched DNA was ultracentrifuged at 44,000 g for 5 days (Beckman Coulter Optima XE), and cell-enriched DNA was separately ultracentrifuged at 44,000 g for 3 days, both in a VTi 65.2 vertical rotor. Each sample was separated into 12 density fractions of ~400uL each. The density of each fraction was measured using a hand-held refractometer calibrated with MilliQ water to nD-tC at 20 °C = 1.3330, and converted using the equation (nD-tC at 20 °C) *10.9276–13.593 as previously described in ref. 70. DNA from each fraction was precipitated using PEG solution[69] and 75 mg of Glycoblue (ThermoFisher AM9516). DNA recovered from each density fraction was quantified using Picogreen (ThermoFisher P11496) and qPCR (Fig. S3). DNA recovery efficiencies, as estimated by the total DNA recovered across all density fractions quantified by picogreen divided by the input DNA, ranged from 42 to 100%. Variabilities in bulk DNA recovery (absolute DNA abundance) does not impact metagenomic sequencing (generates relative abundance data). 3–6 of the heaviest and lightest fractions, which contain the lowest amount of DNA, were pooled to a minimum of 25 ng DNA, resulting in 5 density fractions per sample for sequencing (Fig. S3). DNA libraries were prepared using the Illumina KAPA HyperPrep kit and sequenced on the Illumina Novaseq X plus 10B platform.

Read processing and metagenomic assembly

Raw reads were trimmed using two passes through BBDuk[71] v39.01 to remove Illumina adapters and phiX with the following parameters: ktrim = r k = 21 mink = 11 hdist = 2 tbo tpe for adapter removal during the first pass and k = 27 hdist = 1 qtrim = rl trimq = 17 cardinality = t mingc = 0.05 maxgc = 0.95 for phiX removal during the second pass. Reads from cell-enriched (0.22 µm) samples were pooled across density fractions and assembled in three groups (environmental control, ^{12}C, and ^{13}C) using MEGAHIT[72]. Virsorter2 v.2.2.4[73] was used to identify putative viral contigs in cell-enriched assemblies. QC'ed reads were then mapped using Bowtie2 2.5.1[74] to these putative viral contigs in their corresponding assemblies to identify viral reads within cell-enriched samples. Viral reads from the three cell-enriched assemblies were pooled with the corresponding reads from the three virus-enriched (0.02 µm) samples (environmental control, ^{12}C, and ^{13}C) and co-assembled in metaSPAdes with a minimum contig length of 1.5 kb. QUAST[75] v5.2.0 was used to assess assembly quality.

Viral population genomes

Putative viral contigs were identified from the viral assemblies using Virsorter2 v2.2.4[73] retaining 48,925 putative viral contigs of > 5 kb from all categories across the three assemblies. Putative viral contigs were then dereplicated by clustering at > 95% ANI across > 50% of the shorter contig using anicalc.py and aniclust.py scripts[76], resulting in 31,488 putative viral population genomes. CheckV[76] v1.0.1 was used to assess completeness. High-confidence viral population genomes were identified as previously described in ref. 47, and remaining contigs that did not pass through the workflow were retained only if they contained one or more viral structural genes. Temperate phages were identified as previously described in ref. 77. The VC ratio for each viral population, the ratio for its relative abundance in the virus-enriched fraction relative to that in the cell-enriched fraction, was calculated as previously described in ref. 29.

Viral taxonomic and functional annotation

Viral taxonomic assignments were performed using two complementary approaches: (i) marker-based classification via geNomad[78];

(ii) protein similarity searches against reference viral databases, including NCBI NR and IMG/V4. For the marker-based assignment, the geNomad[78] v1.7.0 pipeline (genomad end-to-end) was used to classify sequences using taxonomically informative protein profiles. To assign taxonomy based on similarity, open reading frames were predicted using Prodigal[79] v.2.6.3 with the parameter "-p meta". The predicted ORFs were then compared to viral proteins in NCBI NR[80] (retrieved in 2025-02-14) and IMG/V4[81] using used DIAMOND[82] v.2.9. Each viral polation genome was assigned to the lowest common taxonomic rank supported by at least 50% of its annotated proteins at > 60% AAI. Metacerubus[83] was used for functional annotation, which includes the KEGG[84], COG[85], VOG[86], PHROG[87], and PFAM[88] databases. Putative prophages were identified using marker genes as previously described in refs. 29,89.

Virus-host linkage

Viral contigs were linked to their hosts through CRISPR spacer mapping using the CRISPRCASFinder[90] and CRASS[91] programs based on 100% alignment across 100% of the spacer, as well as mapping to prophages at > 95% ANI across >1 kb to a cellular contig that is at least twice the length of the viral contig. The taxonomy of the linked hosts were identified through Kaiju v1.10.1[92].

Metagenomic-assembled genomes (MAGs)

Assemblies from the cell-enriched (> 0.22 µm) samples were run through the Binning and Bin_refinement modules of MetaWRAP[93] to construct MAGs. Within the Binning module, MaxBin2[94], MetaBAT2[95] and CONCOCT[96] were used for curating the initial bin sets. The Bin_refinement module was used to refine the bin sets into a consensus bin set for each assembly for a total of 227 bins at ≥ 50% complete and < 10% contamination (Table S5). MAGs were then dereplicated using dRep[97] into one consensus bin set of 158 MAGs (Fig. S8). CheckM[98] v1.1.3 was used for both the Bin_refinement module and dRep to evaluate MAG completeness and contamination based on prokaryotic lineage-specific marker genes. The phylogenetic tree of 158 non-redundant MAGs was constructed based on 120 bacterial marker genes and 53 archaeal marker genes in GTDB-tk. Bacterial and archaeal multiple sequence alignment files obtained GTDB-tk analysis were merged using mafft-profile, a subprogram of MAFFT v7.525[99]. ModelFinder (-m MFP) from IQ-TREE v2.3.4[100] was used to find the best substitution models for the protein alignments. The maximum-likelihood phylogenomic tree was then built using ultrafast bootstrap in IQ-TREE. The tree was visualized using using iTOL (https://itol.embl.de/).

Cellular functional and taxonomic annotation

Kaiju[92] (default setting) was used for taxonomic identification of cellular contigs. Genome Taxonomy Database Toolkit[101] v. 2.1.1 (GTDB) was used for taxonomic identification of MAGs. DRAM[102] and Metacerubus[83] were used for the functional annotation of MAGs, which includes the KEGG[84] and COG[85] databases. The keggLink function from the KEGGREST[103] package on R was used to pull all KO numbers for each carbon fixation pathway from the KEGG database. MAGs were predicted to contain a carbon fixation pathway if they contained all key enzymes as well as ≥ 50% of genes in the complete KEGG pathway, consistent with a previous report[6]. Normalized gene counts, approximating copies per genome, was calculated by dividing the number of genes in that functional category by the number of universal single-marker genes (COG0012/KO6942).

Quantifying taxon-specific isotope incorporation

Bowtie2[74] and samtools[104] were used in conjunction with Anvi'o[63] v.7.1 to calculate interquartile coverage for each contig, which reduces the impact of conserved or hypervariable regions in respectively over/underestimating coverage calculations. R[105] 'decostand' function of the vegan[106] package v2.6-10 was used to convert interquartile coverages for each contig into relative abundances. The excess atom fraction of ^{13}C (^{13}C-EAF) within each contig was calculated as previously described in ref. 45 with the following modifications. The total genomic copy per µL of contig i within density fraction k was calculated by multiplying the total number of genome copies (f; determined by either qPCR ratio for the cell-enriched fractions or DNA yield ratio for the virus-enriched fractions) of density fraction (k) by the relative abundance (R) of contig i within density fraction k. Additionally, instead of calculating the GC content using the regression formula, we calculated GC content of each contig by using the program seqkit[107] with the following parameters: 'seqkit fx2tab –name –only-id –gc'. To improve accuracy in our calculations, we retained only contigs that were present in three or more density fractions in both treatments. To identify populations that demonstrated significant carbon incorporation, we plotted the calculated ^{13}C-EAF values against the ranked ^{13}C-EAF values (numerically ranked from lowest to highest position). A segmented linear regression was then ran against the resulting spline function to identify a break point at which ^{13}C incorporation can be inferred[38]. The function segmented from the R segmented package[108] was used to identify the breakpoint, which was identified at rank 9067.20 with a standard error of 3.28. Three times the standard error was added to this breakpoint to identify the ^{13}C-EAF threshold of 0.049. Assuming genomic ^{13}C at natural abundances of 0.011 (1.1%), a population with a ^{13}C-EAF of 0 indicates that 1.1% of its genomic carbon is ^{13}C, whereas a population with a ^{13}C-EAF of 0.049 indicates that 6% of its genomic carbon is ^{13}C. Microbial populations with ^{13}C-EAF calculations above this threshold were defined as having significant isotopic enrichment.

Identification of eukaryotic sequences

Eukdetect[109] v1.3 "run all" and RiboTagger[110] v0.8.0 were used to identify eukaryotic reads. Eukdetect identifies eukaryotic sequences based on a database of 521,824 microbial eukaryote marker genes, while RiboTagger identifies eukaryotic DNA by aligning reads to 18S rRNA database (v4, v5, v6, and v7 regions). For quantitative comparison among different metagenomics datasets, the relative abundance of eukaryotic species was calculated by normalizing the number of reads mapped to the identified species to million sequencing reads (reads per million) in individual metagenomes.

Reporting summary

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

Data availability

Raw sequence data for this project is available on NCBI under Bioproject PRJNA1390687. Source data are provided with this paper. All other data is provided with the paper or appended as supplementary data. Source data are provided with this paper.

References

1. Friedlingstein, P. et al. Global Carbon Budget 2019. *Earth Syst. Sci. Data* **11**, 1783–1838 (2019).
2. Ricci, F. & Greening, C. Chemosynthesis: a neglected foundation of marine ecology and biogeochemistry. *Trends Microbiol* **32**, 631–639 (2024).
3. Hadas, O., Pinkas, R. & Erez, J. High chemoautotrophic primary production in Lake Kinneret, Israel: A neglected link in the carbon cycle of the lake. *Limnol. Oceanogr.* **46**, 1968–1976 (2001).
4. Ulloa, O., Canfield, D. E., DeLong, E. F., Letelier, R. M. & Stewart, F. J. Microbial oceanography of anoxic oxygen minimum zones. *Proc. Natl. Acad. Sci.* **109**, 15996–16003 (2012).
5. Domack, E. et al. A chemotrophic ecosystem found beneath Antarctic Ice Shelf. *Eos Trans. Am. Geophys. Union* **86**, 269–272 (2005).

6. Overholt, W. A. et al. Carbon fixation rates in groundwater similar to those in oligotrophic marine systems. *Nat. Geosci.* **15**, 561–567 (2022).
7. Farías, L., Fernández, C., Faúndez, J., Cornejo, M. & Alcaman, M. E. Chemolithoautotrophic production mediating the cycling of the greenhouse gases N_2 O and CH_4 in an upwelling ecosystem. *Biogeosciences* **6**, 3053–3069 (2009).
8. Suttle, C. A. Cyanophages and Their Role in the Ecology of Cyanobacteria. In *The Ecology of Cyanobacteria* (eds Whitton, B. A. & Potts, P.) 563–589 (Kluwer Academic Publishers, 2002).
9. Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541–548 (1999).
10. Wommack, K. E. & Colwell, R. R. Virioplankton: Viruses in Aquatic Ecosystems. *Microbiol. Mol. Biol. Rev.* **64**, 69–114 (2000).
11. Wilhelm, S. & Suttle, C. A. Viruses and Nutrient Cycles in the Sea. *BioScience* **49**, 781–788 (1999).
12. Hevroni, G., Vincent, F., Ku, C., Sheyn, U. & Vardi, A. Daily turnover of active giant virus infection during algal blooms revealed by single-cell transcriptomics. *Sci. Adv.* **9**, eadf7971 (2023).
13. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
14. Lara, E. et al. Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Sci. Adv.* **3**, e1602565 (2017).
15. Han, Y. & Perner, M. The globally widespread genus Sulfurimonas: versatile energy metabolisms and adaptations to redox clines. *Front. Microbiol.* **6**, 989 (2015).
16. Biderre-Petit, C. et al. A pan-genomic approach reveals novel *Sulfurimonas* clade in the ferruginous meromictic Lake Pavin. *Mol. Ecol. Resour.* **24**, e13923 (2024).
17. Sass, K. & Perner, M. Characterization of Two Hydrogen-Oxidizing Hydrogenovibrio Strains From Kermadec Volcanic Island Arc Hydrothermal Vents. *Front. Mar. Sci.* **7**, 295 (2020).
18. Scott, K. M. et al. Diversity in CO_2 -Concentrating Mechanisms among Chemolithoautotrophs from the Genera. *Hydrogenovibrio* **85**, e02096–18 (2019).
19. Klepac-Ceraj, V. et al. Microbial diversity under extreme euxinia: Mahoney Lake, Canada. *Geobiology* **10**, 223–235 (2012).
20. Grote, J., Jost, G., Labrenz, M., Herndl, G. J. & Jürgens, K. *Epsilonproteobacteria* Represent the Major Portion of Chemoautotrophic Bacteria in Sulfidic Waters of Pelagic Redoxclines of the Baltic and Black Seas. *Appl. Environ. Microbiol.* **74**, 7546–7551 (2008).
21. Glaubitz, S. et al. 13C-isotope analyses reveal that chemolithoautotrophic Gamma- and Epsilonproteobacteria feed a microbial food web in a pelagic redoxcline of the central Baltic Sea. *Environ. Microbiol.* **11**, 326–337 (2009).
22. Labrenz, M. et al. Sulfurimonas gotlandica sp. nov., a chemoautotrophic and psychrotolerant epsilonproteobacterium isolated from a pelagic redoxcline, and an emended description of the genus Sulfurimonas. *Int. J. Syst. Evol. Microbiol* **63**, 4141–4148 (2013).
23. Perner, M. et al. In situ chemistry and microbial community compositions in five deep-sea hydrothermal fluid samples from Irina II in the Logatchev field. *Environ. Microbiol.* **15**, 1551–1560 (2013).
24. Akerman, N. H., Butterfield, D. A. & Huber, J. A. Phylogenetic diversity and functional gene patterns of sulfur-oxidizing subseafloor Epsilonproteobacteria in diffuse hydrothermal vent fluids. *Front. Microbiol.* **4**, 185 (2013).
25. Perner, M. et al. Linking geology, fluid chemistry, and microbial activity of basalt- and ultramafic-hosted deep-sea hydrothermal vent environments. *Geobiology* **11**, 340–355 (2013).
26. Laber, C. P. et al. Coccolithovirus facilitation of carbon export in the North Atlantic. *Nat. Microbiol.* **3**, 537–547 (2018).
27. Tran, P. Q. & Anantharaman, K. Biogeochemistry Goes Viral: towards a Multifaceted Approach To Study Viruses and Biogeochemical Cycling. *mSystems* **6** (2021).
28. Locke, H. et al. Chapter Two - Marine viruses and climate change: Virioplankton, the carbon cycle, and our future ocean. In *Advances in Virus Research*, Viruses and Climate Change (ed. Roossinck, M. J.) 67–146 (Academic Press, 2022).
29. Luo, E. et al. Double-stranded DNA virioplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. *ISME J.* **14**, 1304–1315 (2020).
30. Labonté, J. M. et al. Single cell genomics-based analysis of gene content and expression of prophages in a diffuse-flow deep-sea hydrothermal system. *Front. Microbiol.* **10**, 1262 (2019).
31. Rastelli, E. et al. High potential for temperate viruses to drive carbon cycling in chemoautotrophy-dominated shallow-water hydrothermal vents. *Environ. Microbiol.* **19**, 4432–4446 (2017).
32. Jian, H. et al. Diversity and distribution of viruses inhabiting the deepest ocean on Earth. *ISME J.* **15**, 3094–3110 (2021).
33. Kieft, K. et al. Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *Nat. Commun.* **12**, 3503 (2021).
34. Roux, S. et al. Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and metagenomics. *eLife* **3**, e03125 (2014).
35. Castelán-Sánchez, H. G. et al. Extremophile deep-sea viral communities from hydrothermal vents: Structural and functional analysis. *Mar. Genomics* **46**, 16–28 (2019).
36. Cheng, R. et al. Virus diversity and interactions with hosts in deep-sea hydrothermal vents. *Microbiome* **10**, 235 (2022).
37. Li, X., Cheng, R., Zhang, C. & Shao, Z. Genomic diversity of phages infecting the globally widespread genus Sulfurimonas. *Commun. Biol.* **7**, 1428 (2024).
38. Starr, E. P. et al. Stable-Isotope-Informed, Genome-Resolved Metagenomics Uncovers Potential Cross-Kingdom Interactions in Rhizosphere Soil. *mSphere* **6**, e00085–21 (2021).
39. Trubl, G. et al. Active virus-host interactions at sub-freezing temperatures in Arctic peat soil. *Microbiome* **9**, 208 (2021).
40. Greenlon, A. et al. Quantitative Stable-Isotope Probing (qSIP) with Metagenomics Links Microbial Physiology and Activity to Soil Moisture in Mediterranean-Climate Grassland Ecosystems. *mSystems* **7**, e00417–e00422 (2022).
41. Lee, S. et al. Methane-derived carbon flows into host–virus networks at different trophic levels in soil. *Proc. Natl. Acad. Sci.* **118**, e2105124118 (2021).
42. Wikner, J., Vallino, J. J., Steward, G. F., Smith, D. C. & Azam, F. Nucleic acids from the host bacterium as a major source of nucleotides for three marine bacteriophages. *FEMS Microbiol. Ecol.* **12**, 237–248 (1993).
43. Ostrander, C. M. et al. Thallium isotope cycling between waters, particles, and sediments across a redox gradient. *Geochim. Cosmochim. Acta* **348**, 397–409 (2023).
44. Vallino, J. J. & Huber, J. A. Using maximum entropy production to describe microbial biogeochemistry over time and space in a meromictic pond. *Front. Environ. Sci.* **6**, 100 (2018).
45. Hungate, B. A. et al. Quantitative Microbial Ecology through Stable Isotope Probing. *Appl. Environ. Microbiol.* **81**, 7570–7581 (2015).
46. Vyshenska, D. et al. A standardized quantitative analysis strategy for stable isotope probing metagenomics. *mSystems* e01280-22. https://doi.org/10.1128/msystems.01280-22. (2023).
47. Guo, J., Vik, D., Pratama, A. A., Roux, S. & Sullivan, M. Viral sequence identification SOP with VirSorter2 V.3. https://doi.org/10.17504/protocols.io.bwm5pc86.(2021).
48. Watanabe, T., Kubo, K., Kamei, Y., Kojima, H. & Fukui, M. Dissimilatory microbial sulfur and methane metabolism in the water

column of a shallow meromictic lake. *Syst. Appl. Microbiol.* **45**, 126320 (2022).
49. Updegraff, T. et al. Thiomicrorhabdus heinhorstiae sp. nov. and Thiomicrorhabdus cannonii sp. nov.: novel sulphur-oxidizing chemolithoautotrophs isolated from the chemocline of Hospital Hole, an anchialine sinkhole in Spring Hill, Florida, USA. *Int. J. Syst. Evol. Microbiol.* **72**, 3 (2022).
50. Savvichev, A. S. et al. Microbial Processes and Microbial Communities in the Water Column of the Polar Meromictic Lake Bol'shie Khruslomeny at the White Sea Coast. *Front. Microbiol.* **11**, 1945 (2020).
51. Gregersen, L. H. et al. Dominance of a clonal green sulfur bacterial population in a stratified lake: Clonal dominance in microbial lake community. *FEMS Microbiol. Ecol.* **70**, 30–41 (2009).
52. Lauro, F. M. et al. An integrative study of a meromictic lake ecosystem in Antarctica. *ISME J.* **5**, 879–895 (2011).
53. Tran, P. Q. et al. Depth-discrete metagenomics reveals the roles of microbes in biogeochemical cycling in the tropical freshwater Lake Tanganyika. *ISME J.* **15**, 1971–1986 (2021).
54. Block, K. R., O'Brien, J. M., Edwards, W. J. & Marnocha, C. L. Vertical structure of the bacterial diversity in meromictic Fayetteville Green Lake. *Microbiol. Open* **10**, e1228 (2021).
55. Phillips, A. A. et al. Microbial succession and dynamics in meromictic Mono Lake, California. *Geobiology* **19**, 376–393 (2021).
56. Saini, J. S. et al. Bacterial, phytoplankton, and viral distributions and their biogeochemical contexts in meromictic lake cadagno offer insights into the proterozoic ocean microbial loop. *mBio* **13**, e00052–22 (2022).
57. Cabello-Yeves, P. J., Picazo, A., Roda-Garcia, J. J., Rodriguez-Valera, F. & Camacho, A. Vertical niche occupation and potential metabolic interplay of microbial consortia in a deeply stratified meromictic model lake. *Limnol. Oceanogr.* **68**, 2492–2511 (2023).
58. Overmann, J., Cypionka, H. & Pfenning, N. An extremely low-light adapted phototrophic sulfur bacterium from the Black Sea. *Limnol. Oceanogr.* **37**, 150–155 (1992).
59. Manske, A. K., Glaeser, J., Kuypers, M. M. M. & Overmann, J. Physiology and phylogeny of green sulfur bacteria forming a monospecific phototrophic assemblage at a depth of 100 meters in the black sea. *Appl. Environ. Microbiol.* **71**, 8049–8060 (2005).
60. Bao, N. & Gao, K. Interactive Effects of Elevated CO2 Concentration and Light on the Picophytoplankton Synechococcus. *Front. Mar. Sci.* **8**, 634189 (2021).
61. Puxty, R. J., Millard, A. D., Evans, D. J. & Scanlan, D. J. Viruses inhibit CO_2 fixation in the most abundant phototrophs on earth. *Curr. Biol.* **26**, 1585–1589 (2016).
62. Zhong, K. X., Wirth, J. F., Chan, A. M. & Suttle, C. A. Mortality by ribosomal sequencing (MoRS) provides a window into taxon-specific cell lysis. *ISME J.* **17**, 105–116 (2023).
63. Eren, A. M. et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
64. Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28**, 127–181 (2004).
65. Suttle, C. A. & Chen, F. Mechanisms and rates of decay of marine viruses in seawater. *Appl. Environ. Microbiol.* **58**, 3721–3729 (1992).
66. Martínez Martínez, J., Talmy, D., Kimbrel, J. A., Weber, P. K. & Mayali, X. Coastal bacteria and protists assimilate viral carbon and nitrogen. *ISME J.* **18**, wrae231 (2024).
67. Amend, A. From dandruff to deep-sea vents: Malassezia-like fungi are ecologically hyper-diverse. *PLoS Pathog.* **10**, e1004277 (2014).
68. Kleindienst, S. et al. Diverse sulfate-reducing bacteria of the *Desulfosarcina/Desulfococcus* clade are the key alkane degraders at marine seeps. *ISME J.* **8**, 2029–2044 (2014).
69. Dunford, E. A. & Neufeld, J. D. DNA Stable-Isotope Probing (DNA-SIP). *J. Vis. Exp.* 2027. https://doi.org/10.3791/2027.(2010).
70. Buckley, D. H., Huangyutitham, V., Hsu, S.-F. & Nelson, T. A. Stable Isotope Probing with^{15} N Achieved by Disentangling the Effects of Genome G+C Content and Isotope Enrichment on DNA Density. *Appl. Environ. Microbiol.* **73**, 3189–3195 (2007).
71. Bushnell, B. BBMap: A. Fast, accurate, splice-aware aligner in *Report Number: LBNL-7065E*, (17 AD).
72. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
73. Guo, J. et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).
74. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
75. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
76. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021).
77. Luo, E., Leu, A. O., Eppley, J. M., Karl, D. M. & DeLong, E. F. Diversity and origins of bacterial and archaeal viruses on sinking particles reaching the abyssal ocean. *ISME J.* **16**, 1627–1635 (2022).
78. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* **42**, 1303–1312 (2024).
79. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma.* **11**, 119 (2010).
80. Sayers, E. W. et al. GenBank 2024 Update. *Nucleic Acids Res.* **52**, D134–D137 (2024).
81. Camargo, A. P. et al. IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Res.* **51**, D733–D743 (2023).
82. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
83. Figueroa, J. L. III, Dhungel, E., Bellanger, M., Brouwer, C. R. & White, R. A. III MetaCerberus: distributed highly parallelized HMM-based processing for robust functional annotation across the tree of life. *Bioinformatics* **40**, btae119 (2024).
84. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
85. Galperin, M. Y. et al. COG database update 2024. *Nucleic Acids Res.* **53**, D356–D363 (2025).
86. Grazziotin, A. L., Koonin, E. V. & Kristensen, D. M. Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic Acids Res.* **45**, D491–D498 (2017).
87. Terzian, P. et al. PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genomics Bioinforma.* **3**, lqab067 (2021).
88. Mistry, J. et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
89. Luo, E., Aylward, F. O., Mende, D. R. & DeLong, E. F. Bacteriophage Distributions and Temporal Variability in the Ocean's Interior. *mBio* **8**, e01903–e01917 (2017).
90. Couvin, D. et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251 (2018).
91. Skennerton, C. T., Imelfort, M. & Tyson, G. W. Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. *Nucleic Acids Res.* **41**, e105–e105 (2013).

92. Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **7**, 11257 (2016).
93. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
94. 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).
95. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
96. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
97. 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).
98. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
99. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
100. Minh, B. Q. et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
101. Parks, D. H. et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785–D794 (2022).
102. Shaffer, M. et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res.* **48**, 8883–8900 (2020).
103. Tenenbaum, D. & Maintainer, B. KEGGREST: Client-side REST access to the Kyoto Encyclopedia of Genes and Genomes (KEGG). *R Package Version 1480* (2025).
104. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
105. R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna Austria* (2023).
106. Okasen, J. et al. vegan: Community Ecology Package. *R Package Version 26-10* (2024).
107. Shen, W. et al. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE* **11**, e0163962 (2016).
108. Muggeo, V. M. R. Interval estimation for the breakpoint in segmented regression: a smoothed score-based approach. *Aust. N. Z. J. Stat.* **59**, 311–322 (2017).
109. Lind, A. L. & Pollard, K. S. Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing. *Microbiome* **9**, 58 (2021).
110. Xie, C., Goi, C. L. W., Huson, D. H., Little, P. F. R. & Williams, R. B. H. RiboTagger: fast and unbiased 16S/18S profiling using whole community shotgun metagenomic or metatranscriptome surveys. *BMC Bioinforma.* **17**, 508 (2016).

Acknowledgements

We would like to acknowledge Gretta Serres, Alex Worden, Matthew Johnson, Sabrina Elkassas, Cynthia Becker, and Amy Apprill for discussions and/or methodology relevant to this manuscript. The data generation, analyses, and manuscript were supported by the University of North Carolina at Charlotte (startup funds to E.L.), the Hypothesis Fund (to E.L.), and the National Science Foundation (OCE-2513189 to E.L.). Wet lab analyses were supported by the Woods Hole Oceanographic Institution (Weston Howland Jr. Postdoctoral Fellowship to E.L.), the National Oceanic and Atmospheric Administration (NA19OAR4320072 subaward 0007525/102212019 to J.A.H.), and the National Science Foundation (OCE-1947776 to J.A.H.). B.E.B. was supported by the National Science Foundation (PRFB2010963). JJV was supported by the Simons Foundation (549941FY22). G.T. was supported by the US Department of Energy (SCW1632) and was conducted under Contract DE-AC52-07NA27344.

Author contributions

E.L. conceptualized, supervised, and acquired funding for the study with input from J.A.H.; E.L. and B.E.B. collected the samples with help from JV and JAH. EL conducted experiments and wet lab analyses with input from G.T. and J.A.H.; N.P., T.J.R., M.M.S., J.J.V., and E.L. conducted the data analyses. E.L. wrote the manuscript with input from all co-authors. Correspondence and requests for materials should be addressed to elaine.luo@charlotte.edu.

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/s41467-026-71833-x.

Correspondence and requests for materials should be addressed to Elaine Luo.

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