**Supplementary Methods**

*Geochemical measurements*

As previously reported [1], total nitrogen, total carbon, and total sulfur were determined using Elementar vario EL cube (Elementar Co., Germany). Ammonium was extracted using 2 M KCl and quantified with Hach Kit 26045445 (Hach, USA). Metals were extracted with 0.5 M HCl, then Fe(II) was measured via Ferrozine assay , while total iron and other metals were quantified by Inductively Coupled Plasma Mass Spectrometry (ICP-MS) after extraction with HCl. Geochemistry measurements per sample are reported in **Additional File 7**. Significant differences reported for geochemistry were analyzed using r stats package rstatix with function pairwise\_t\_test with parameters pool.sd=FALSE and p.adjust.method “bonferroni”. These results were confirmed with an analysis of variance (aov, significance = Pr <.05) with default parameters. A resazurin assay which measures the irreversible reduction of resazurin to resorufin was used as an indicator for aerobic metabolism [1,2].

*Organic matter characterization by FTICR-MS*

Briefly, three different solvents with varying polarities—water (H2O), methanol (CH3OH, ‘MeOH’) and chloroform (CHCl3)— were used to sequentially extract a large diversity of organic compounds from samples. Following extraction, ultra-high resolution mass spectrometry of the three different extracts from each sample was carried out using a 12 Tesla Bruker SolariX FTICR-MS located at the Environmental Molecular Sciences Laboratory (EMSL) in Richland, WA, USA. Here we focus on the water extractions in order to look at the water soluble, most microbially available organic matter fraction (**Additional File 8**). To describe the overall organic matter available in these sediments, we calculated the relative abundance of each class within samples as the number of peaks corresponding to each class/total assigned peaks (**Additional File 8, Additional File 3: Figure S5**).

**Supplementary text**

*C1 metabolism in HZ sediments*

For CO metabolisms in these sediments, we considered carboxydotrophs and carboxydovores. Carboxydotrophs grow chemolithoautotrophically with CO as the sole energy and carbon source, while carboxydovores require organic carbon for growth [3]. Based on genomic potential, we posit that Binatia, CSP1-3, and Micromonosporaceae are capable of carboxydotrophy, while Actino\_1 is a carboxydovore, using CO metabolism as supplemental energy or possible carbon source during starvation. We classified carboxydovores if a genome with a CO dehydrogenase also lacked a mechanism for carbon fixation (e.g. Calvin–Benson cycle), suggesting they can support aerobic respiration, but not carbon fixation, using CO [3]. Thus, Binatia have 80% of the Calvin–Benson cycle (no RUBISCO), CSP1-3 had 90% of the reverse TCA cycle, and Micromonosporaceae had 80% of the reverse TCA cycle, designating them as carboxydotrophs. In contrast, Actino\_1 had no mechanism for carbon fixation and thus was designated a carboxydovore.

In these sediments, we found no evidence for methane production or oxidation, indicating that these processes occurred minimally or elsewhere. Lack of methane production may be due to oxygen toxicity as river water intrudes and oxygenates sediments, or inability to compete for substrates. Further, the lack of methane oxidation may be due to low methane concentrations as a result of low production activity, consistent with prior reports from a nearby site that show very low to below detection methane fluxes Columbia River HZ sediments [4]. These findings contrast with recent work on methane fluxes that show river systems harbor nearly as much methane flux as rice paddies and wetlands [5], however it is possible that other sediments have greater organic C availability, different sediment textures affecting redox zonation, or that methane cycling is active in surface waters rather than in the HZ.

*Percent carbon and nitrogen predictions from microbial and viral abundances*

To garner additional support for our proposed linkage between viruses and sediment biogeochemistry via the microbial hosts, we showed the viral community were more strongly associated with percent carbon, and percent nitrogen in these samples than host communities (**Additional File 9**). Consistent with roles inferred from host genomes, the viral genome for Actino\_1 was better correlated than its host with overall C (virus: r2=0.72 p=4.75e-55, host r2=0.54 p=1.02e-25), and the viral genome for Nitro\_40CM-3\_1 was better correlated than its host to overall N (virus: r2=0.88 p=1.05e-55, host: r2=0.73 p=8.31e-40). To further address how viruses potentially impact ecosystem geochemistry, we conducted a sparse partial least square (sPLS) regression of both bacterial and viral abundances with regards to ecosystem geochemistry and calculated variable importance in projection (VIP) scores for each vMAG and MAG (**Additional File 9,** see sheet vMAG + MAG SPLS VIP). Overall, the viral community was a better predictor than microbial genome abundances of both percent carbon (virus p=0.0006 r2=0.78 cor=0.89, microbial genome p=0.0369, r2=0.44, cor=0.67) and percent nitrogen (virus p=0.0006 r2=0.77 cor=0.88; microbial genome p=0.1128 r2=0.28 cor=0.53) and a better or equal predictor with regards to the combined viral and microbial genome abundances (percent C: p=0.0006 r2=0.77 cor=0.88, percent N: p=0.0012 r2=0.75 cor=0.87) (**Additional File 3: Figure S12**).

*Viral auxiliary metabolic genes*

We found evidence for possible bottom-up biogeochemical influences of viruses mediated by virus-encoded genes. We identified 14 putative AMGs with a DRAM-v auxiliary score of 1-3, meaning the predicted metabolic gene was flanked by a viral or viral-like gene on either side, providing confidence that this was a high-quality viral region (**Additional File 6**). The most abundant/prevalent AMGs include GH39 (hemicellulose cleaving), K00957 (sulfate adenylyltransferase), PL1 (pectin cleaving), and GH140 (glycan degradation) (**Additional File 3: Figure S11a**). One of our viral genomes that was putatively linked to Steroidobacteraceae encoded a pectin lyase gene (PL1). This PL1 could cleave pectin, generating pectin oligosaccharides, which the host could potentially utilize via host encoded glycoside hydrolases (GH4 and GH2) to yield galactose monomers that it uses for energy metabolism (**Additional File 3: Figure S11ab**). While theoretical in nature, we include this example as it indicates how virally encoded genes could expand the substrate ranges for their hosts, expanding the host niches and further influencing HZ biogeochemistry. Beyond expanding viral taxonomic sampling to uncover many new species of viruses not previously included in existing databases, we show that viral community members are actively expressing genes, instead of being only passive members of the HZ system. Our field derived multi-omics analyses indicate that viruses could potentially modulate microbial C and N cycling through predation and augmentation of host metabolism via AMGs.

**References:**

1. Graham EB, Crump AR, Kennedy DW, Arntzen E, Fansler S, Purvine SO, et al. Multi ’omics comparison reveals metabolome biochemistry, not microbiome composition or gene expression, corresponds to elevated biogeochemical function in the hyporheic zone. Sci Total Environ. 2018;642:742–53.

2. Haggerty R, Martí E, Argerich A, von Schiller D, Grimm NB. Resazurin as a “smart” tracer for quantifying metabolically active transient storage in stream ecosystems. J Geophys Res [Internet]. American Geophysical Union (AGU); 2009;114. Available from: http://doi.wiley.com/10.1029/2008JG000942

3. Cordero PRF, Bayly K, Man Leung P, Huang C, Islam ZF, Schittenhelm RB, et al. Atmospheric carbon monoxide oxidation is a widespread mechanism supporting microbial survival. ISME J. 2019;13:2868–81.

4. Villa JA, Smith GJ, Ju Y, Renteria L, Angle JC, Arntzen E, et al. Methane and nitrous oxide porewater concentrations and surface fluxes of a regulated river. Sci Total Environ. 2020;715:136920.

5. Rosentreter JA, Borges AV, Deemer BR, Holgerson MA, Liu S, Song C, et al. Half of global methane emissions come from highly variable aquatic ecosystem sources. Nat Geosci. Nature Publishing Group; 2021;14:225–30.

**Timeline

Description automatically generated with medium confidenceSupplementary figures:**

**Figure S1. DRAM annotation of MAGs.** Heatmap showing the DRAM product output for medium and high-quality genomes (n=102) from HUM-V. The interactive version of this heatmap is available here: <https://zenodo.org/record/5124964>

**Diagram

Description automatically generated with medium confidence**

**Figure S2. Phylogenetic trees of Nxr, Nap, Nar, and Nos amino acid sequences in HUM-V. a)** Phylogenetic tree of molybdopterin oxidoreductase protein family (PF00384) with sequences recovered in MAGs generated using FastTree on 8,313 amino acid positions aligned using MAFFT. Tips of MAGs are colored according to their phylum, and statistical support and phylogenetic distances are indicated. Functional groups of interest, e.g., dissimilatory nitrate and nitrite reducing enzymes, are displayed. **b)** Phylogenetic tree of nitrous oxide reductase protein clusters (available on NCBI) and sequences recovered in MAGs generated using FastTree on 620 amino acid positions aligned using MAFFT. Tips of MAGs are colored according to their phylum, and statistical support and phylogenetic distances are indicated. The two clades of this enzyme and their functional differences are briefly described. The *nxr*-like *nar* group includes sequences that share high sequence homology with *nxr* but current evidence suggest they perform the same biochemical reactions as *nar* as shown by Daims et al., 2016: A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria.

Timeline

Description automatically generated with medium confidence

**Figure S3**. **Full genome-resolved MAG database shows low abundant members recruit a considerable portion of metaproteomic peptides. A)** Left side of butterfly plot showing dereplicated MAG (55 shown) summed relative abundances across all samples. Each color represents a MAG phylum as assigned by the GTDB. MAGs that contain a partial or complete 16S rRNA sequence are in italicized and in bold. **B)** Right side of butterfly plot showing normalized peptide abundance per MAG, as well as non-unique specialized peptides (e.g., those that have non-unique mapping but all map to the same taxonomy and function - see methods) and non-unique peptides (e.g., peptides that have non-unique mapping and cannot be assigned to the same genus or function - see methods).

**Diagram

Description automatically generated**

**Figure S4. DRAM annotation of proteome.** Heatmap shows the DRAM product output for the metaproteome data pertaining to HUM-V. The interactive version of this heatmap is available here: <https://zenodo.org/record/5124964>

**Diagram

Description automatically generated**

**Figure S5. Relative abundance of FTICR-MS classes across samples.** The relative abundance of biochemical classes identified in FTICR-MS across samples is not statistically structured by depth. Bar plots show the average relative abundance for each class, with error bars representing one standard deviation (n=33). Individual data points are plotted for each sample with point size increasing with depth. Peaks were classified as described in **Methods**. Raw data is provided in **Additional File 8.**

**Diagram

Description automatically generated**

**Figure S6. MAG and vMAG metagenomic and metaproteomic data show no significant differences by depth or site. a**)Procrustes ordination of MAG and vMAG NMDS ordinations using relative abundance genome data across 33 shallower sequenced samples. **b)** Euler diagram showing the number of total proteins recruiting peptides in each transect, where the overlap represents the proportion of these proteins recruiting peptides in both transects. Only proteins recruiting 2 or more total peptides were included (n=898) to allow for recruitment of at least 1 peptide for a given protein to both transects or both depths and reduce false positives. **c)** NMDS of MAG peptide relative abundances at a per protein (left) and per genome (right) resolution. Colors represent depth. **d)** NMDS of vMAG presence / absence peptide recruitment on a per genome basis. Colors represent depth as in **c**. Significance between depths and sites were inferred from anosim and mrpp tests in R, where a p-value <0.05 was considered significant.

**Diagram

Description automatically generated**

**Figure S7. Geochemistry measurements revealed significant differences for ammonium across grouped depths (surface and deep).** Box and whisker plots report concentrations. Left graph shows grouped surface (0-30cm) and deep (30-60cm) measurements for each. Right graph shows measurement across different depths (0-60cm). P-values for significant differences across graphs are shown by red asterisks (\*) for each respective graph, where p-value<0.05 was considered significant.

**Chart, bar chart

Description automatically generated**

**Figure S8. NMR detected metabolites and ammonium indicate prevalence of nitrogen-containing compounds, saccharides, organic acids, and alcohols.** Bar graphs showing specific NMR metabolites and ammonium with the percent of total samples that they were found in. Shading denotes the different categories of detected compounds. For detailed information on how these were collected see methods section. Only metabolites detected in more than 4 samples are shown.

**Diagram

Description automatically generated with medium confidence**

**Figure S9. Average nucleotide identity between MAGs within Nitrospiraceae and Thermoproteota show different genera across respective phyla.** Heatmaps showing the average nucleotide identity (ANI) for closely related strains of dereplicated MAGs. For Nitrospiraceae: 3 different genera divided as follows: NS7: 15 MAGs; NITRO\_2-02: 3 MAGs; NITRO\_2-02\_11: 1 MAG; and NITRO\_40CM-3: 1 MAG.For Thermoproteota (archaea): 3 dRepped bins - minimum 98.5% ANI between each other indicating that they are the same species.

Graphical user interface, application

Description automatically generated

**Figure S10. Virus-Host associations show viruses could infect key players in river sediment carbon and nitrogen cycling.** **a)** Genome cartoons colored by their phyla. Gray dotted circles represent vMAGs that putatively infect each genome. **b)** A breakdown of each vMAG name for each of the infected host genomes. Colors match genome host phyla assignment.

**Chart

Description automatically generated**

**Figure S11. Viral encoded auxiliary metabolic genes have the potential expand host carbon metabolism.** **a)** Bar chart showing the different types of AMGs identified in viral genomes with putative hosts, with the predicted substrate for each AMG denoted by colored boxes on the y-axis. **b)** Viruses encode AMGs which can expand the metabolic repertoire of their microbial hosts. Conceptual models of a putative glycoside hydrolase (GH) and a pectin lyase (PL) are shown. **c)** For one putative host, Steroidobacteraceae, with a linkage to a viral genome that encodes a PL1, a conceptual model of an integrated metabolism is shown. Specifically, integration of vMAG.63 genome into the host genome may provide the capacity to cleave the pectin backbone (PL1), followed by oligo cleavage of pectin, resulting in galactose monomers for host cell metabolism.

**Graphical user interface, chart, application

Description automatically generated**

**Figure S12. sPLS regressions show vMAG abundance is better predictor of carbon and nitrogen relative to MAGs.** sPLS plots showing predicted vs measured values for percent carbon (c\_per) and percent nitrogen (n\_per) for our samples using n=10 deep sequences. Shown sPLS are **a)** vMAGs only, **b)** MAGs only, and **c)** vMAGs and MAGs combined. Values corresponding to cor.test function output in R for the predicted and measured values are shown in each box (t, degrees of freedom, p-value, confidence interval, and correlation).