Microbial Genome-Resolved Metaproteomic Analyses Frame Intertwined Carbon and Nitrogen Cycles in River Hyporheic Sediments

Josué A. Rodríguez-Ramos
Colorado State University Department of Soil and Crop Sciences https://orcid.org/0000-0002-2049-2765

Mikayla A. Borton
Pacific Northwest National Laboratory Biological Sciences Division

Bridget B. McGivern
Colorado State University Department of Soil and Crop Sciences

Garrett J. Smith
Radboud University: Radboud Universiteit

Lindsey M. Solden
Colorado State University Department of Soil and Crop Sciences

Michael Shaffer
Colorado State University Department of Soil and Crop Sciences

Rebecca A. Daly
Colorado State University Department of Soil and Crop Sciences

Samuel O. Purvine
Pacific Northwest National Laboratory

Carrie D. Nicora
Pacific Northwest National Laboratory Biological Sciences Division

Elizabeth K. Eder
Pacific Northwest National Laboratory

Mary Lipton
Pacific Northwest National Laboratory

David W. Hoyt
Pacific Northwest National Laboratory

James C. Stegen
Pacific Northwest National Laboratory

Kelly C. Wrighton (kelly.wrighton@colostate.edu)
Colorado State University Department of Soil and Crop Sciences https://orcid.org/0000-0003-0434-4217

Research

Keywords: microbiome, viruses, metagenomics, river sediment, denitrification, nitrification, carboxydotrophs, Thaumarchaeota, peptidases, Binatia

Posted Date: August 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-746574/v1

Abstract

Background:

Rivers serve as a nexus for nutrient transfer between terrestrial and marine ecosystems and as such, have a significant impact on global carbon and nitrogen cycles. In river ecosystems, the sediments found within the hyporheic zone are microbial hotspots that can account for a significant portion of ecosystem respiration and have profound impacts on system biogeochemistry. Despite this, studies using genome-resolved analyses linking microbial and viral communities to nitrogen and carbon biogeochemistry are limited.

Results:

Here, we characterized the microbial and viral communities of Columbia River hyporheic zone sediments to reveal the metabolisms that actively cycle carbon and nitrogen. Using genome-resolved metagenomics, we created the Hyporheic Uncultured Microbial and Viral (HUM-V) database, containing a dereplicated database of 55 microbial Metagenome-Assembled Genomes (MAGs), representing 12 distinct phyla. We also sampled 111 viral Metagenome Assembled Genomes (vMAGs) from 26 distinct and novel genera. The HUM-V recruited metaproteomes from these same samples, providing the first inventory of microbial gene expression in hyporheic zone sediments. Combining this data with metabolite data, we generated a conceptual model where heterotrophic and autotrophic metabolisms co-occur to drive an integrated carbon and nitrogen cycle, revealing microbial sources and sinks for carbon dioxide and ammonium in these sediments. We uncovered the metabolic handoffs underpinning these processes including mutualistic nitrification by Thermoproteota (formerly Thaumarchaeota) and Nitrospirota, as well as identified possible cooperative and cheating behavior impacting nitrogen mineralization. Finally, by linking vMAGs to microbial genome hosts, we reveal possible viral controls on microbial nitrification and organic carbon degradation.

Conclusions:

Our multi-omics analyses provide new mechanistic insight into coupled carbon-nitrogen cycling in the hyporheic zone. This is a key step in developing predictive hydrobiogeochemical models that account for microbial cross-feeding and viral influences over potential and expressed microbial metabolisms. Furthermore, the publicly available HUM-V genome resource can be queried and expanded by researchers working in other ecosystems to assess the transferability of our results to other parts of the globe.

Background

The hyporheic zone (HZ) acts as a transitional space between river and groundwater compartments in the river corridor where the bidirectional supply of nutrients and organic carbon stimulate microbial activity [1–3]. Characterized as the permanently saturated interface between the river surface channel and underlying sediments, the HZ is considered a biogeochemical hotspot for microbial cycling of carbon

and nitrogen [1, 3]. These zones have been reported to support microbial heterotrophic respiration, denitrification, and nitrification, as well as the consumption and production of greenhouse gases (such as nitrous oxide and carbon dioxide) [4–7]. In addition to harboring diverse energy metabolisms, these sediments may act as an important sink of carbon and nitrogen in terms of microbial subsurface biomass [4, 8]. Overall estimates of ecosystem respiration have revealed that the HZ accounts for 40 to 90% of total river respiration [9, 10], highlighting that a substantial amount of respiration is associated with hyporheic microbial activities.

Despite the importance of microbial metabolism to river corridor biogeochemistry, the ability to partition metabolic handoffs between organisms, the linked use of carbon and nitrogen by individual organisms, and the mineralization of detritus have yet to be holistically interrogated. Metagenomic studies in river sediments have not fully inventoried carbon and nitrogen cycling metabolisms, instead focusing on specific aspects of the nitrogen cycle (e.g., genes in denitrification [11]). Moreover, most of these studies were not genome resolved, hindering the assignment of biogeochemical processes to specific microorganisms, and culture independent genomic reconstructions from river sediments are limited to a handful of studies [12, 13], all of which focused exclusively on nitrification. Thus, little is known about uncultured microbial communities in river sediments, with the enzymes, interconnected chemical reactions, and microbial metabolic lifestyles mediating carbon and nitrogen transformations in river sediments not currently resolvable from existing HZ microbiome datasets.

Here we address this knowledge gap, creating a genome-resolved inventory of the microbial and viral members in HZ sediments collected from the Columbia River in Washington State, USA (Fig. 1). This resource was used to recruit metaproteomic data, providing a first of its kind, comprehensive inventory of the active microbial organisms and their enzymatic machinery in river systems. We contextualized these biological findings, using chemical scaffolding provided from metabolomics and geochemistry. Reconstructing the expressed metabolic capabilities of numerous lineages enabled us to resolve microbial contributions to biogeochemistry in these sediments. Our proteome enabled road map outlined the metabolic circuitry coupling carbon and nitrogen biogeochemistry in these HZ sediments, providing a framework to develop hydrobiogeochemical models informed by biochemical mechanisms and ecological interactions.

Methods

Experimental Design

To investigate the microbial processes involved in biogeochemical cycling in HZ sediments, we leveraged previously collected geochemical, metagenomic, metaproteomic, and Fourier-transform ion cyclotron resonance mass spectrometry (FTICR-MS) data and combined it with new nuclear magnetic resonance (NMR) metabolomic characterization and additional metagenomic sequencing. Samples were collected from the hyporheic zone of the Columbia River (46°22’15.80"N, 119°16’31.52"W) in March 2015 as previously described [14]. Briefly, liquid nitrogen frozen sediment profiles (0-60cm) were collected along

two transects separated by approximately 170 meters (Fig. 1a). At each transect, three sediment cores up to 60 cm in depth were collected at 5-meter intervals perpendicular to the river flow. All cores were collected during conditions in which the sediments were fully saturated. Each core was sectioned into 10 cm segments from 0-60-centimeter depths and stored at -80ºC. Analyses were carried out at 10-centimeter increments, with the exception of one core that was pooled from 0–30 centimeters to have sufficient input masses. Collectively, this robust, paired multi-omic dataset was made up of previously reported metagenomes (n = 33, 3-4Gbp), metaproteomes (n = 33), FTICR-MS metabolomes (n = 33), and geochemical characterizations (n = 33), as well as new metagenomes (n = 10, 10-25Gbp) and NMR metabolomes (n = 17) (Fig. 1b). Genome recovery from these samples was improved by increasing the metagenomic sequencing depth per sample (from an average of 3.8 to 25.3 Gbp for selected samples) and employing a hybrid of co-assembly and single assembly methods (Fig. 1c).

DNA extraction and sequencing

As described previously [14], deoxyribonucleic acid (DNA) was extracted from the sediments using the MoBio PowerSoil kit (MoBio Laboratories, Inc., Carlsbad, CA) following manufacturer's instructions, with the addition of a 2-hour proteinase-K incubation at 55°C prior to bead-beating to facilitate cell lysis. Purified genomic DNA was sent to the Joint Genome Institute (JGI, n = 33) under Joint Genome Institute / Environmental Molecular Sciences Laboratory (EMSL) proposal 1781 or to the Genomics Shared Resource facility at The Ohio State University (OSU, n = 10), producing 43 metagenomes with average sequencing depth of 4 (JGI) and 25 Gbp (OSU) per sample, totaling 377Gbp. DNA submitted to JGI were prepared for sequencing using an Illumina Library creation kit (KAPA Biosystems), and then solid-phase reversible immobilization size selection. DNA submitted to OSU were prepared for sequencing with a Nextera XT library System followed by solid-phase reversible immobilization size selection. Libraries at both facilities were quantified to ensure input thresholds, and then sequenced using an Illumina HiSeq 2500 platform. Deeper sequencing was performed at OSU to enhance MAG recovery, resulting in an increase of 252 Gbp of additional sequencing for ten samples, increasing the sequencing depth per sample by at least 3-fold (15.37–49.24 Gbp per sample) **(**Fig. 1, **Additional File 1)**. **Additional File 1** details all sequencing information, including National Center for Biotechnology Information (NCBI) accession numbers.

Metagenome assembly and binning

Raw reads were trimmed for length and quality using Sickle v1.33 (https://github.com/najoshi/sickle) and then subsequently assembled using Iterative De Bruijn Graph De Novo Assembler – Uneven Depth (IDBA-UD) 1.1.0 [15] with an initial kmer of 40 or metagenomic St. Petersburg genome Assembler (metaSPAdes) 3.13.0 [16] with default parameters. To further increase genomic recovery, for the ten samples that had shallow and deep sequencing, metagenomic reads were coassembled using IDBA-UD 1.1.0 with an initial kmer of 40. All assemblies, including co-assemblies, were then individually binned using Metabat2 [17] with default parameters to obtain Metagenome Assembled Genomes (MAGs).

For each bin, genome completion was estimated based on the presence of core gene sets (highly conserved genes that occur in single copy) for Bacteria (n = 31 genes) and Archaea (n = 104 genes) using Amphora2 [18]. Bins were discarded for further analysis if completion was < 70% or contamination was > 10% to select for only medium to high quality bins [19]. This resulted in 102 MAGs that were then dereplicated using dRep [20] with default parameters and resulted in a final set of 55 MAGs (> 99% ANI). To further assess bin quality, we used the Distilled and Refined Annotation of MAGs (DRAM) [21] annotation pipeline to identify ribosomal ribonucleic acids (rRNAs) and transfer ribonucleic acids (tRNAs). The 102 MAGs detailed here are deposited on NCBI under the BioProject ID PRJNA576070, with genome quality information reported in **Additional File 2.**

Phylogenetic and metabolic analysis of Metagenome Assembled Genomes (MAG)

Medium and high-quality MAGs were taxonomically classified using the Genome Taxonomy Database (GTDB) Toolkit v1.3.0 on September 2020 [22]. Novel taxonomy was identified as the first taxonomic level with no designation using GTDB taxonomy. For example, MAGs whose GTDB taxonomy string was not designated after the family level (e.g., g__) were identified as novel genera. Of the MAGs that had multiple representatives sharing taxonomy strings up to the Family level (Binatia and CSP1-3), we used average nucleotide identity (ANI) to determine whether they belonged to the same genus. MAG scaffolds were annotated using the DRAM pipeline [21]. The raw annotations for each genome are deposited in the Zenodo repository under doi 10.5281/zenodo.5128772 and can be accessed here: https://doi.org/10.5281/zenodo.5128772. **Additional File 2** shows the metabolic summary of genomes (product DRAM output) and output is also displayed in **Additional File 3: Figure S1**.

Target metabolic marker genes of interest recovered in bins were used to query the Integrated Microbial Genomes / Microbiomes Expert Review (IMG/M ER) (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi) and NCBI (https://www.ncbi.nlm.nih.gov) databases using BLASTp, or retrieved from the Protein Family (PFAM) database by protein family. Returned amino acid sequences were compiled with other known genes not retrieved via sequence homology, and de-replicated to make a reference sequence database. Sequences from the metagenomes were then aligned to the reference sequences using Multiple Sequence Comparison by Log-Expectation (MUSCLE) version 3.8.31 [23] or Multiple Alignment using Fast Fourier Transform (MAFFT) version 7.427 [24]. Alignments were manually curated to remove end and other gap regions. These alignments were then used to construct phylogenetic trees using FastTree version 2.1.11 [25] with default settings.

An additional phylogenetic analysis was performed on genes annotated as respiratory nitrate reductase (*nar*) and nitrite oxidoreductase (*nxr*) to resolve novel Binatia role in nitrogen cycling. Specifically, sequences from [26] were downloaded and combined with *nar* and *nxr* amino acid sequences from dereplicated bins, aligned using MUSCLE, version 3.8.31, and run through ProtPipeliner, a Python script developed in-house for generation of phylogenetic trees (https://github.com/TheWrightonLab). Phylogenetic trees are shown in **Additional File 3: Figure S2** and **Additional File 4**.

For polyphenol and organic polymer degradation, we used functional annotation in addition to predicted secretion to assess functional potential. To determine if the predicted genes encoded a secreted protein, we used pSortb [27] and SignalP [28] to predict location; if those methods did not detect a signal peptide, the amino acid sequence was queried to SecretomeP and a SecP score > 0.5 [29] was used as a threshold to report non-canonical secretion signals. Metabolic information for each MAG discussed in this manuscript are available in **Additional File 2** and **Additional File 5**.

Viral Analyses

Metagenomic assemblies (n = 43) were screened for DNA viral sequences using VirSorter v1.0.3 with the ViromeDB database option [30], retaining viral contigs ranked 1, 2, 4 or 5 with greater than 10kb in genome length as stated by the Minimum Information about an Uncultivated Virus Genome (MIUViG) standards [31]. To determine an approximate species level taxonomy for viral scaffolds, they were clustered into viral metagenome assembled genomes (vMAGs) at 95% ANI across 85% of the shortest contig using ClusterGenomes 5.1 (https://github.com/simroux/ClusterGenome) [31]. After clustering, vMAGs were manually confirmed to be viral by assessing the total of viral genes with regards to non-viral genes in the genome, where genomes containing more than 18% of non-viral genes were discarded (J flag, DRAM [21]).This resulted in 111 vMAGs that were deposited on NCBI under the BioProject ID PRJNA576070 **Additional File 6.**

To determine taxonomic affiliation, vMAGs were clustered to viruses belonging to standard viral reference taxonomy databases NCBI Bacterial and Archaeal Viral RefSeq V85 with the International Committee on Taxonomy of Viruses (ICTV) and NCBI Taxonomy using the network-based protein classification software vContact2 v0.9.8 [32, 33]. Default methods were used. To determine geographic distribution of viruses in freshwater ecosystems, we also included viruses mined from publicly available freshwater metagenomes in vContact2 analyses: 1) East River, CO (PRJNA579838) 2) A previous study from the Columbia River, WA (PRJNA375338) 3) Prairie Potholes, ND (PRJNA365086) and 4) the Amazon River (PRJNA237344).

Viral contigs were annotated with DRAM-v [21], with annotations for each viral genome reported in **Additional File 6**. Genes that were identified by DRAM-v as being possible auxiliary metabolic genes (categories 1–3) were subjected to protein modeling using Protein Homology / AnalogY Recognition Engine (PHYRE2) in order to improve the accuracy of annotation [34]. To identify likely vMAG hosts, oligonucleotide frequencies between virus (n = 111) and non-dereplicated hosts (n = 102) were analyzed using VirHostMatcher using a threshold of d2* measurements of < 0.25 [35]. Per genome results of VirHostMatcher predictions are reported in **Additional File 6**. The lowest d2* value for each viral contig < 0.25 was used. Viruses and hosts could not be linked by matching clustered regularly interspaced short palindromic repeats (CRISPR) spacers in host genomes to vMAG genomes.

Genome relative abundance calculations

Relative abundance for each MAG and vMAG was estimated using in-house scripts available at https://github.com/TheWrightonLab [36, 37]. Briefly, all metagenomic reads were concatenated for each sample, rarified to 3Gbp, and multi-mapped to 55 unique MAGs via Bowtie2 [38]. For MAGs, a minimum

scaffold coverage of 75% and depth of 3x required for read recruitment at 7 mismatches. For vMAGs, reads were mapped using Bowtie2 [38] at a maximum mismatch of 15, a minimum contig coverage of 75% and a minimum depth coverage of 2x. Relative abundances for each MAG and vMAG were calculated as their coverage proportion from the sum of the whole coverage of all bins for each set of metagenomic reads. To identify correlations between MAGs and vMAGs to geochemistry and predictive capability, we also mapped the subset of deep sequencing (n = 10) and rarefied to the lowest deep metagenome available (4.8Gbp). Correlations and sparse Partial Least Squares Regression (sPLS) predictions (PLS R package [39]) were done using mapping data pertaining to only the 10 deeply sequenced metagenomes. Genome relative abundances per sample for MAGs and vMAGs are reported in **Additional File 2** and **Additional File 6**.

Metaproteome generation and peptide mapping

Sediment samples were prepared for metaproteome analysis as previously reported in Graham et al. 2018 [14] and the protocol outlined by Nicora et al [40]. For protein identification, spectra were searched against two files that included (i) 55 dereplicated MAG and (ii) 111 clustered vMAGs amino acid sequences. Exact sequence duplicates were removed, and 16 commonly observed contaminants (e.g., tryptic fragments, human keratins, and serum albumin precursors) were included. The tandem mass spectrometry (MS/MS) spectra from all liquid chromatography tandem mass spectrometry

(LC-MS/MS) datasets were converted to ASCII text (.dta format) using MSConvert (http://proteowizard.sourceforge.net/tools/msconvert.html) which more precisely assigns the charge and parent mass values to an MS/MS spectrum. The data files were then interrogated via target-decoy approach [41] using MSGF+ [42] using a ± 20 ppm parent mass tolerance, partially tryptic digestion enzyme settings, and a variable posttranslational modification of oxidized Methionine. All MS/MS search results for each dataset were collated into tab separated ASCII text files listing the best scoring identification for each spectrum. Collated search results were further combined into a single result file. These results were imported into a Microsoft SQL Server database. Results were filtered to N1% false detection rate (FDR) using an MSGF + supplied Q-Value that assesses reversed sequence decoy identifications for a given MSGF score across each dataset. Using the protein references as a grouping term, unique peptides belonging to each protein were counted, as were all peptide spectrum matches (PSMs) belonging to all peptides for that protein (i.e., a protein level observation count value). PSM observation counts were reported for each sample that was analyzed. Crosstabulation tables were created to enumerate protein level PSM observations for each sample, allowing low-precision quantitative comparisons to be made.

Microbial metaproteomes were converted to normalized spectral abundance frequency (NSAF) values and subsequently divided into unique, non-unique specialized, and non-unique categories, while viral metaproteomes were analyzed using peptide counts only from unique hits due to low recruitment [36]. Peptide recruitment for each MAG amino acid sequence per sample is reported in **Additional File 5**. Hits were divided into 3 categories: (1) uniques (peptide hits to a single protein), (2) non-unique specialized (peptide hits to multiple amino acid sequences that all had same annotation and MAG taxonomy), (3)

non-unique (peptide hits to multiple amino acid sequences with different annotation or from MAGs with different taxonomy) [43]. This designation was necessary as several hits could not be resolved to the MAG level due to functional conservation across closely related genomes in the HUM-V database. Data in Fig. 2 showcases (1) and (2) categories, with the entire dataset shown in **Additional File 3: Figure S3**. Including the non-unique specialized hits assigned an additional 14% of the proteome (grey bar, Fig. 2b) and confirmed we did not underrepresent the gene expression from genomically well-sampled strains (e.g., Nitrospiraceae). Metaproteome hits for MAGs were used for further metabolic analyses if they were detected in at least three samples. Annotations for the entire metaproteomic dataset are shown in **Additional File 3: Figure S4**.

Geochemical measurements, FTICR-MS characterization of organic matter, and NMR detected metabolites.

As previously reported [14], total nitrogen, total carbon, and total sulfur were determined using Elementar vario EL cube (Elementar Co., Germany), with details in the Supplementary Information (**Additional File 7**). To characterize organic matter, we used FTICR-MS to analyze sediments as previously reported [14], with details in the Supplementary Information (**Additional File 3: Figure S5, Additional File 8**).

To identify the metabolites available to microorganisms in this river system, we performed ^{1}H Nuclear Magnetic Resonance (NMR) spectroscopy on sediment pore water. Sediment samples were mixed with 200, 300, or 600 µL of MilliQ water depending on the sediment mass (**Additional File 7**) and centrifuged to remove the sediment. Supernatant (180 µL) was then diluted by 10% (vol/vol) with 5 mM 2,2-dimethyl-2-silapentane-5-sulfonate-d_6 as an internal standard. All NMR spectra were collected using a Varian Direct Drive 600-MHz NMR spectrometer equipped with a 5-mm triple resonance salt-tolerant cold probe. Chemical shifts were referenced to the 1H or 13C methyl signal in DSS-d6 at 0 ppm. The 1D ^{1}H NMR spectra of all samples were processed, assigned, and analyzed using Chenomx NMR Suite 8.3 with quantification based on spectral intensities relative to the internal standard as described previously [36, 44]. Candidate metabolites present in each of the complex mixtures were determined by matching the chemical shift, J-coupling, and intensity information of experimental NMR signals against the NMR signals of standard metabolites in the Chenomx library. Compounds were assigned a rank and assign confidence to metabolites (RANCM) value according to the amount of spectral information used to identify the compound (**Additional File 7**) [45].

For many metabolites, including aspartate, asparagine, sucrose, acetate, methanol, and glucose, we utilized 2D NMR to corroborate the 1D data, providing more confidence to an assignment. The two-dimensional 1H-1H total correlation spectroscopy (TOCSY) spectra were collected using the Varian TOCSY pulse sequence with a TOCSY mixing time of 80 ms (MLEV-17). Spectral widths were 12 ppm in both directions with 256 increments acquired in the indirect dimension and 64 transients per increment. The relaxation delay was 1.5 s during which presaturation of the water signal was applied and the acquisition time was 143 ms during which 2048 total points were acquired. The 2D 1H-13C heteronuclear

single-quantum correlation spectroscopy (HSQC) spectra were acquired using the Varian gHSQCAD pulse sequence with a 1JCH of 146 Hz. Spectral widths were 12 ppm and 160 ppm for the direct and indirect dimensions, respectively, with 256 increments acquired in the indirect dimension and 128 transients per increment. The relaxation delay was 1.5 s during which presaturation of the water signal was applied. The acquisition time was 143 ms in which 13C composite pulse decoupling (wurst140) was applied and 2048 total points were acquired. NMR-identified metabolites discussed in the text were present in 30% of the samples.

Results And Discussion

The HUM-V genome database enabled metaproteomic characterization of river sediment microbiomes

Here we created the Hyporheic Uncultured Microbial and Viral (HUM-V) genomic catalog from Columbia River HZ sediments. We leveraged this resource for metaproteomic peptide recruitment, enabling identification of the community members and their gene expression in these sediments.

We reconstructed 655 bacterial and archaeal metagenome assembled genomes (MAGs); 102 were medium or high-quality genomes based on the Genome Consortium Standards [19] (**Additional File 2**). These genomes were dereplicated into 55 genomic representatives to form the bacterial and archaeal portion of the HUM-V microbial genome database. These dereplicated HUM-V MAGs were distributed across 9 Bacterial and 2 Archaeal phyla. In terms of new genomic discoveries, 1 genome represented a new order within the Actinobacteriota, and 12 genomes represented 6 new genera from archaeal and bacterial phyla including members of the Thermoplasmatota, Acidobacteria, Actinobacteriota, CSP1-3, Proteobacteria, and Desulfobacterota (Fig. 1b, Fig. 2a).

From the same metagenomic assemblies we reconstructed and reported viral metagenome assembled genomes (vMAGs), making this one of only a handful of genome-resolved studies that include viral genomes derived from rivers [46–48], and to our knowledge, the first study to complement these with bacterial and archaeal genomes. We reconstructed 2,482 vMAGs that dereplicated into 111 dereplicated viral populations > 10kb in size (**Additional File 6**). Given their sparse sampling from river corridors, only 5 of the HUM-V viral genomes had taxonomic assignments using established viral taxonomies from standard reference databases. To better understand if the remaining 95% (n = 105) of viral genomes were completely novel or had been previously detected in similar ecosystems, we repeated the analyses, this time adding 1,861 viral genomes we reconstructed or pulled from public metagenomes from four freshwater sites in North and South America (Fig. 2c, **Additional File 6**). Of the 105 remaining viral genomes in HUM-V, 15% (n = 17) clustered with these freshwater derived sequences, indicating a portion of this viral community is shared across diverse geographic and freshwater systems. Of the remaining viral genomes, 23% (n = 26) clustered only with genomes recovered in this data set, indicating multiple samplings of the same virus spatially at this site, while 57% (n = 63) of the viral genomes we sampled

were singletons (i.e., only sampled from these sediments once). These results hint at the possible cosmopolitan and endemic viral lineages that warrant further exploration.

HUM-V recruited viral and microbial peptides from our HZ sediment metaproteomic dataset (n = 33 lateral and depth resolved samples) (Fig. 2b**d, Additional File 5**). Across all sediment samples, microbial genomes recruited 13,102 total peptides to ~ 1,300 proteins in HUM-V, with 68% of these proteins uniquely assigned to a single microbial genome. For viruses and microbes alike, the most abundant genomes were not necessarily the most actively expressing proteins. The most abundantly ranked microbial members included the Nitrospiraceae genus NS7, Binatia, and Nitrososphaeraceae genus TA-21 (Fig. 2b), yet only the Nitrososphaeraceae had high proteomic recruitment (15%). Similarly, some low abundance members (e.g., members of the Actinobacteria) accounted for a majority of the uniquely assigned proteome relative abundance (49%) (Fig. 2b). Like our microbial dataset, 66% of the viral genomes encoded genes that uniquely recruited peptides (Fig. 2d). This exceeded prior viral metaproteome recruitment from other environmental systems (e.g., wastewater, saliva, rumen (0.4–15%, [49–51]), thus we infer a relatively large portion of the viral community was active at the time of sampling. While microbial and viral activity did not appear to be structured by transect, sediment depth, or geochemical conditions, these two assemblages were coordinated to one another (**Additional File 3: Figure S6**). Explaining this lack of geochemical or spatial structuring, it is possible that the microbial heterogeneity in these samples occurred over a finer spatial resolution (pore or biofilm scale) than the bulk 10 cm depths sampled or that these HZ sediment microbiomes are metabolically robust to the small, but significant changes in chemistry measured across spatial gradients (**Additional File 3: Figure S6, Additional File 3: Figure S7, Additional File 7**).

Microbial cross feeding of organic carbon is likely sustained by aerobic respiration

It is well recognized that microbial carbon oxidation in HZ sediments largely contributes to river respiration, yet the microbial food webs underpinning this process have yet to be documented. Consistent with resazurin (raz) data (see Additional File 3: supplementary methods) that indicated these sediments were oxygenated and supported aerobic microbial respiration (**Additional File 3: Figure S7**) [52], all but one of the microbial genomes recovered from this site encoded aerobic respiration machinery, including a complete electron transport chain and a cytochrome oxidase (**Additional File 3: Figure S1**). Proteomic evidence for aerobic respiration (cytochrome c oxidase *aa3*) was detected from nearly all samples, but only assigned to few members of the Nitrososphaeraceae. However, given limitations with detecting membrane cytochromes [53], we consider it likely this metabolism was more active than was captured in proteomic data, as we failed to find any evidence for other anaerobic metabolisms (e.g., methanogenesis).

While the overall carbon content of these sediments was low (< 10 mg/g) (**Additional File 7**), our FTICR-MS analysis indicated that plant litter could be an important substrate, as lignin-like compounds were the most abundant biochemical class detected (**Additional File 3: Figure S5**, **Additional File 8**). In support of

this, from our metagenomes, 38% of the HUM-V genomes encoded genes for potentially degrading phenolic/aromatic monomers, while 10% could degrade the larger, more recalcitrant polymers (**Additional File 2**). Gene expression of carbohydrate-active enzymes (CAZymes) also supported the degradation of plant polymers like starch and cellulose via extracellular glucoamylase (GH15) and endo-glucanase (GH5) from an actinobacterial genome (Microm_1) and the Nitrososphaeraceae (Nitroso_2), respectively (Fig. 3). In summary, many types of chemical and biological data reveal that heterotrophic, aerobic metabolism in these low carbon sediments is likely maintained by inputs from decomposition.

Given the capacity for plant polymer decomposition (e.g., lignin, cellulose, and starch) across HUM-V genomes, we next tracked the microbial fate of the degradation products of these metabolisms, including sugar monomers, short chain fatty acids, and carbon dioxide (Fig. 3, **Additional File 2, Additional File 5**). Metabolites detected by NMR included sugars (e.g., glucose, sucrose, and trehalose), which could be the result of depolymerization of plant derived polymers, and we confirmed the CAZYmes to use these substrates were also expressed *in situ*. Additionally, NMR also detected organic acids (acetate, butyrate, lactate, pyruvate, propionate) and alcohols (ethanol, methanol, isopropanol), with proteomics supporting the usage of acetate and methanol by Anaeromyxobacter MAG (Anaerom_1) and archaeal Woeseia (Woese_1), respectively. Here our metabolite and proteomic data demonstrated that plant biomass degradation supports sequential metabolic handoffs that lead to carbon dioxide production.

Carbon dioxide production and consumption is widely encoded by HUM-V microorganisms

In addition to carbon dioxide being generated from the heterotrophic metabolisms described above, our proteomics revealed that carbon dioxide could arise by the aerobic oxidation of carbon monoxide (CO). Genes for aerobic CO dehydrogenases (from Actinobacteria, Binatia, and CSP-1 genomes) were among the most expressed in these sediments. Analogous to findings from soil systems, it is possible that atmospheric carbon monoxide is a major energy source supporting persistent aerobic heterotrophic bacteria in deprived, or dynamic organic carbon environments [54]. Based on the genomic inventory of these HUM-V genomes, 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 [54].

Since heterotrophic respiration and carbon monoxide oxidation would generate carbon dioxide in these sediments, we next tracked microorganisms that could use this carbon source autotrophically (Fig. 3, **Additional File 2, Additional File 5**). The ability to fix carbon was prevalent, encoded by 75% of HUM-V microbial genomes. In fact, this metabolism was represented by multiple fixation pathways from 18 different lineages, demonstrating both functional and taxonomic redundancy. Specifically, this includes 4 different pathways (e.g., Calvin-Benson-Bassham cycle, reductive TCA cycle, 3-HydroxyPropionate /4-HydroxyButyrate cycle, 3-Hydroxypropionate bi-cycle) from members of nitrifying lineages (Thaumarchaeota and Nitrospirota) (discussed below), as well as from organisms with heterotrophic capabilities like Binatia, CSP1_3, Proteobacteria, Woeseiaceae, and Acidobacteria (**Additional File 3:**

Figure S1, **Additional File 2**). Collectively our multi-omics data suggest that sediment microbial respiration is likely decoupled from river respiration, since some microbially produced carbon dioxide would be lost to supporting autotrophy. Our research further resolves the carbon economy in HZ sediments, implying that the net effect of carbon dioxide emissions from rivers could depend on the balance between carbon dioxide production from heterotrophy and carbon monoxide, as well as consumption by autotrophs.

Microbial metaproteomics supports theoretical inferences derived from geochemistry

The ratio of total element carbon (C) and total nitrogen (N) (e.g., C/N) is a geochemical indicator often used to assess the possible microbial metabolisms that can be supported in an ecosystem [55, 56]. Here the C/N ratios of these sediments were relatively low to other sediments at 6.4 ± 1.1 across the samples (**Additional File 7**). Biogeochemical theory posits that C/N values less than 15 would indicate rapid microbial mineralization of organic nitrogen to release inorganic nitrogen [57]. This theory also states that C/N ratios less than 10 may indicate ammonium is released to the surrounding environment, allowing sufficient concentrations to simultaneously support the assimilatory needs of heterotrophs and energy needs of nitrifiers, allowing for their co-occurrence [56]. Our multi-omics data offered a new opportunity to substantiate these geochemical inferences by profiling the possible substrates and microbial activity of nitrogen mineralizers and nitrifiers in river sediments.

Given the prevalence of ammonium in all 33 sediment samples (0.28–11.22 μg $gram^{-1}$) (**Additional File 3: Figure S8, Additional File 7**), we next examined our metaproteomic data for peptidases, genes that could contribute to the mineralization of organic nitrogen into amino acids and free ammonium. Hinting at the relevance of this metabolism, the gene expression of peptidases (n = 31) was 3 times more abundant and prevalent than glycoside hydrolase genes modulating organic C transformations (**Additional File 5**). In support of active microbial N mineralization, hydrophobic, polar, and hydrophilic amino acids were prevalent (more so than sugars) in the H^1-NMR characterized metabolites (**Additional File 3: Figure S8**).

We focused our analyses on the putative extracellular peptidases, as these were most likely to shape organic nitrogen pools in the sediment. We categorized expressed peptidase families as either amino acid releasing (end terminus cleaving, e.g., M28) or peptide releasing (endocleaving, e.g., S08A, M43B, M36, MO4) (Fig. 4, **Additional File 5**). Linking these expressed peptidases to our genomes, members of the Actinobacteria, Thermoproteota, and Methylomirabilota, and Binatia are likely candidates for driving the mineralization of organic N. We then profiled amino acid transporters that were expressed, revealing uptake of branched chain amino acids, glutamate, osmoprotectants, spermidine/putrescine, and peptides (Fig. 4). This profiling indicated synergy and competitions for this organic N resource in these sediments.

We propose that in HZ sediments extracellular peptidases are a shared public good whose cost of production is assumed by certain individuals with benefits to the entire community [58]. In some cases, taxa that mineralized organic N were consumers of the resulting products, as genomes in the

Actinobacteria and Binatia expressed external peptidases genes and the genes for transporting the organic N products (Fig. 4, linkages shown). In other instances, members of the Proteobacteria, Thermoplasmatota, and CSP1-3 could be functioning as cheater cells that expressed only genes for intracellular transport and benefitting from peptidases produced by others. Our findings reinforce that cooperative interactions based on cross-feeding and public goods are likely at the core of many processes relevant to organic carbon (Fig. 3) and nitrogen (Fig. 4) cycling in these sediments.

Consistent with established conceptual geochemical theory, we showed the lower C:N ratios (< 10) of these sediments not only supported mineralization which could be a source of free ammonium in these sediments, but also nitrification. Supporting this, ammonium was detected in all sediments (average concentration 2.6 µg/gram of sediment) (**Additional File 3: Figure S8, Additional File 7**). Proteomics confirmed ammonium (NH_4^+) oxidation to nitrite was performed by Archaeal Nitrososphaeraceae (formerly Thaumarchaeota), with ammonia monooxygenase proteins being one of the most prevalent and highly expressed functional proteins (top 5%) across this dataset (**Additional File 5**). The next step in nitrification, nitrite oxidation to nitrate was inferred from nitrite oxidoreductase peptides assigned to 5 genomes belonging to 2 new species (Nitro_40CM-3_1, Nitro_NS7_3, Nitro_NS7_4, Nitro_NS7_5, and Nitro_NS7_14) (**Additional File 3: Figure S9**, **Additional File 5**, see sheet metabolism info). Both nitrifying lineages had the capacity for carbon dioxide fixation with the reductive tricarboxylic acid (TCA) cycle (e.g., ATP-citrate lyase) in Nitrospiraceae genomes, and 3-HydroxyPropionate/4-HydroxyButyrate (3HP/4HB) encoded by the Nitrososphaeraceae. We did not detect genomic evidence for comammox or anammox and thus aerobic, chemolithoautotrophic nitrification supported by a metabolic partnership between bacteria and archaea occurred in the presence of heterotrophs as predicted by C/N ratios.

Similarly, others have reported the prominence of nitrifying lineages from the archaeal thaumarcheotal Thermoproteota and bacterial Nitrospirota both by 16S rRNA [59] and genome-resolved metagenomics [12, 13] in HZ sediments. Here we nearly doubled the genomic sampling of these river nitrifiers, assigning unique gene expression patterns to 3 and 17 genomes from Nitrososphaeraceae and Nitrospiraceae respectively, including the first genomic sampling of new genera and species (Fig. 2). Our co-expression data indicate that metabolic handoffs between archaeal ammonia oxidizers and bacterial nitrite oxidizers may be an unaccounted-for biogenic source of nitrate in these sediments (**Additional File 5**). This suggests the activity of nitrifiers could be an underappreciated modulator of nitrous oxide fluxes from oligotrophic HZ sediments, both through their indirect stimulation of denitrifiers and their own contributions to this greenhouse flux [60]. Taken together, the archaeal-bacterial nitrifying mutualism outlined here appears well adapted to the low nutrient conditions present in many HZ sediments, warranting future research on the variables that constrain nitrification rates (i.e., ammonium availability, dissolved oxygen, pH) and their role as driver of nitrogen fluxes from these systems [61].

Denitrification is encoded by novel and taxonomically diverse lineages in HZ sediments

Beyond the possible biogenic sources of nitrate, we identified from nitrification, these HZ sediments receive significant allochthonous nitrate from groundwater. When river stage decreases, groundwater discharges through the HZ sediments, bringing nitrate concentrations to over 20 mg/L [2, 62]. In support of an important influence of nitrate from either source, HUM-V genomes with the capacity for nitrate reduction spanned diverse taxonomies, with NarG or NapX encoded in 11 genomes from the Actinobacteriota, Binatia, Gammaproteobacteria, and Myxococcota (**Additional File 3: Figure S1**). However, our proteomic evidence for nitrate reduction was detected in less than 10% of the 33 sediment samples, with unique peptides assigned to Binatia NarG from a single sample.

Based on gene expression data, we inventoried other steps in the denitrification pathway. Nitrite was reduced via nitrifier and denitrifier reduction to nitric oxide from archaeal ammonia oxidizers of the Nitrososphaeraceae active in 79% of metaproteome samples, and from Gammaproteobacterial Burkholderia in a single sample, respectively. The role of nitrite reduction by Nitrososphaeraceae is still under investigation but could be used for detoxification [63]. Genes for converting nitric oxide to nitrous oxide were not detected in proteomics, but we did find evidence that the Desulfobacterota genome (Desulf_UBA2774_1) expressed the *nos* gene for reducing nitrous oxide to nitrogen gas. Phylogenetic analysis suggest this organism used a "Clade II" *nos* sequence type adapted for low atmospheric concentrations of nitrous oxide (**Additional File 3: Figure S2**), and consistent with our genome metabolic summary did so without encoding other steps of the denitrification pathway [64]. Notably, the capacity for denitrification exists beyond those detected in proteomics, as Binatia encoded dissimilatory nitrite reduction to ammonium (DNRA) and the potential for nitrous oxide production via *nor* was encoded by two Gammaproteobacteria (Steroid-FEN-1191_1, Steroid_1) and a member of the Myxococcota (Anaerom_1).

In summary, our data adds to the growing realization that complete denitrification by single microorganism is likely the exception rather than the rule in natural systems [65], including the HZ [66]. In support of this, none of the genomes reconstructed here encoded a complete denitrification pathway for reducing nitrate to nitrous oxide or dinitrogen gas (**Additional File 3: Figure S1**). Similarly, our proteomics data hinted that separate microbial members likely catalyzed each step of the denitrification pathway (**Additional File 3: Figure S4**). This suggests cross-organism inorganic nitrogen exchange would be necessary for nitrogen gas flux, such that physical processes (e.g., advection, diffusion) or the spatial colocalization of microorganisms, as well as organic carbon availability, may have disproportionate impacts on flux of nitrous oxide and dinitrogen from these sediments.

HUM-V identifies new microbial and viral players in hyporheic zone carbon and nitrogen cycling

The creation of a genome database expanded upon prior amplicon-based surveys, allowing us to assign new metabolic functions to microbes and even viruses in hyporheic sediments. While HUM-V contains genomes from phyla (CSP1-3, Eisenbacteria) and classes (Binatia, MOR-1 in Acidobacteria) composed entirely of uncultivated members (Fig. 1d), here we focus our analysis on the Binatia, as we recovered 7

genomes (one which included a complete 16S rRNA gene), they recruited peptides, and they also played key roles in carbon and nitrogen cycling. Using the 16S rRNA gene (from Binatia_7), we inventoried the distribution of closely related species to our HUM-V genomes (> 97% similarity) in the Sequence Read Archive (SRA) samples, to uncover the ecological distribution of these organisms from soils, as well as a wide variety of terrestrial, terrestrial-aquatic, marine samples (Fig. 5), indicating the processes uncovered by proteomics here are likely applicable to a wide range of ecosystems.

A recent comparative genomics analysis on Binatota MAGs provided a first assessment of their metabolic potential, indicating genes for methylotrophy, alkane degradation, and pigment production were distributed across the phylum [67]. These HUM-V genomes belong to a class and family denoted UBA9968. Contrary to their prior metabolic inventory, HUM-V UBA9968 MAGs do not encode the potential for methanol oxidation, and we identified a new role in the decomposition of aromatic compounds from plant biomass (phenylpropionic acid, phenylacetic acid, salicylic acid), and xenobiotics (phthalic acid) (Fig. 5). We provide the first proteomic evidence for any members of the Binatia, supporting their roles in aerobically oxidizing carbon monoxide, producing extracellular peptidases, and in denitrification. Together these findings illustrate the power of HUM-V paired proteomes to illuminate new roles for members of uncultivated, previously enigmatic lineages in HZ carbon and nitrogen cycling.

The relatively high proteomic recruitment of viruses sampled in HUM-V (Fig. 2d) suggested important viral contributions in these sediments. *In silico* analysis assigned a putative host to 29% of the 111 viral genomes linking 18 microbial genomes that belong to bacterial members in Acidobacteriota, Actinobacteriota, CSP1-3, Eisenbacteria Methylomirabilota, Myxococcota, Nitrospirota, and Proteobacteria (**Additional File 3: Figure S10**, **Additional file 2**, **Additional file 6**). Analysis of the metaproteomes for these phage-impacted microorganisms revealed these hosts expressed genes for nitrification (Nitrospiraceae) as well as carbon monoxide oxidation and nitrogen mineralization (Actinobacteria) (Fig. 6). Additionally, HUM-V phage genomes encode auxiliary metabolic genes with the potential to enhance microbial metabolism of carbon (CAZymes), sulfur (sulfate adenyl transferase), and nitrogen (amidase to cleave ammonium) (**Additional File 3: Figure S11**, see Additional File 3 supplemental text). We also show viral abundances were better predictors of total carbon and nitrogen percentages relative to microbial genome abundances (**Additional File 3: Figure S12, Additional File S9**, see Additional File 3 supplemental text). Together, these HUM-V enabled results indicate viral infections may contribute to river sediment functioning and raise the question to whether enhanced viral interrogation might provide a means to improved ecosystem or biogeochemical models in these systems.

Conclusions

To our knowledge this study represents one of the first genome-resolved microbial and viral enabled proteomic studies in river sediments. Using genome-resolved proteomics with complementary metabolites (detected by NMR, FTICR-MS), and geochemistry we begin to illuminate the microbial contributions to processes well known to occur but previously poorly defined mechanistically in river sediments (e.g., nitrogen mineralization). We also show how multi-omic tools can uncover previously

enigmatic processes which may directly impact river respiration (e.g., carbon monoxide oxidation). While river carbon and nitrogen budgets are often quantified by direct measurements of inputs and the concentration of inorganic and organic compounds exported from rivers, what is missing today is an appreciation for the microbially and virally mediated sources and sinks for key intermediates (e.g., carbon dioxide, ammonium, nitrate), the degree to which these compounds are recycled and exchanged, and the underlying microbial metabolic lifestyles that catalyze this interconnected carbon and nitrogen biogeochemistry. Here, we have created a conceptual framework that elaborates on these missing ideas.

Empowered by our individual process-based metaproteomic analyses (Figs. 3–6), we created a conceptual model outlining the microbial conversions of carbon and nitrogen in these hyporheic sediments (Fig. 7). Heterotrophic oxidation of organic carbon derived from plant (and likely microbial) biomass supported by oxygen and nitrogen respiring populations produce carbon dioxide. In addition, metaproteomics divulged that aerobic carbon monoxide oxidation may also be a source of carbon dioxide. Like organic carbon, the organic nitrogen in microbial and plant biomass could be mineralized to release ammonium in these sediments. Together inorganic pools of nitrogen (ammonium) and carbon (carbon dioxide) sustain the coordinated activity of nitrifying populations. Together our findings put forth an integrated framework that advances microbial roles in hyporheic carbon and nitrogen transformations, yielding insights that could inform research strategies to reduce existing predictive uncertainties in river corridor models.

Abbreviations

HUM-V: Hyporheic Uncultured Microbial and Viral (HUM-V) database

MAG: Metagenome assembled genome

vMAG: Viral metagenome assembled genome

HZ: Hyporheic zone

FTICR-MS: Fourier-transform ion cyclotron resonance mass spectrometry

NMR: Nuclear magnetic resonance

DNA: Deoxyribonucleic acid

rRNA: Ribosomal ribonucleic acid

tRNA: Transfer ribonucleic acids

JGI: Joint Genome Institute

EMSL: Environmental Molecular Sciences Laboratory

OSU: The Ohio State University

NCBI: National Center for Biotechnology Information

SRA: Sequence Read Archive

Gbp: Giga base pair

IDBA-UD: Iterative De Bruijn Graph De Novo Assembler – Uneven Depth

metaSPAdes: metagenomic St. Petersburg genome Assembler

GTDB: Genome Taxonomy Database

ANI: Average Nucleotide Identity

DRAM: Distilled and Refined Annotation of MAGs

IMG/MER: Integrated Microbial Genomes / Microbiomes Expert Review

PFAM database: Protein Family Database

MUSCLE: Multiple Sequence Comparison by Log-Expectation

MAFFT: Multiple Alignment using Fast Fourier Transform

nar: respiratory nitrate reductase

nxr: nitrite oxidoreductase

nos: nitric oxide synthase

DNRA: Dissimilatory Nitrate Reduction to Ammonium

MIUViG: Minimum information about an Uncultivated Virus Genome

ICTV: International Committee on Taxonomy of Viruses

PHYRE2: Protein Homology / AnalogY Recognition Engine 2.0

sPLS: sparse Partial Least Squares Regression

MS: Mass Spectrometry

MS / MS: Tandem mass spectrometry

LC-MS/MS: Liquid Chromatography Tandem Mass Spectrometry

FDR: False Detection Rate

PSM: Peptide Spectrum Matches

NSAF: Normalized spectral abundance frequency

MHz: Megahertz

DSS-d6: 3-(Trimethylsilyl)-1-propanesulfonic acid-d6 sodium salt

ppm: parts per million

RANCM: Rank and AssigN Confidence to Metabolites

TOCSY: Total Correlation Spectroscopy

HSQC: Heteronuclear single-quantum correlation spectroscopy

gHSQCAD: Gradient-enhanced Heteronuclear Single Quantum Coherence with Adiabatic Pulses

CAZymes: Carbohydrate-Active Enzymes

GH: Glycoside Hydrolase

CO: Carbon Monoxide

TCA: Tricarboxylic acid

C: Carbon

N: Nitrogen

C:N: Carbon / Nitrogen Ratio

$NH_4^{+:}$ Ammonium

3HP/4HB: 3-HydroxyPropionate/4-HydroxyButyrate

ATP: Adenosine triphosphate

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and material: The datasets supporting the conclusions of this article are publicly available. Sequencing data are available in NCBI under bioproject PRJNA576070, with MAGs deposited under biosamples SAMN18867633-SAMN18867734 and 16S rRNA amplicon sequences under accession numbers SRX9312157-SRX9312180, and vMAGs have been temporarily deposited in Zenodo doi 10.5281/zenodo.5124937. Viral genomes fasta file is publicly available within the following repository link: https://doi.org/10.5281/zenodo.5124937. Metaproteomics data are deposited in the MassIVE database under accession MSV000087330. Metabolomics data are publicly available and deposited in Zenodo doi https://doi.org/10.5281/zenodo.5076253. Additional datasets supporting the conclusions of this article are included within the article (and its additional files).

Competing interests: The authors declare they have no competing interests.

Funding: This work was supported by the Subsurface Biogeochemical Research (SBR) program (DE-SC0018170); the National Sciences Foundation Division of Biological Infrastructure [#1759874]; and the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Environmental System Science (ESS) program through subcontract from the River Corridor Scientific Focus Area project at Pacific Northwest National Laboratory. J.R.R. is funded by the National Science Foundation (NRT-DESE) [1450032], support for A Trans-Disciplinary Graduate Training Program in Biosensing and Computational Biology at Colorado State University. The NMR data, FTICR-MS data and MS-proteomics data in this work was collected using instrumentation in the Environmental Molecular Science Laboratory (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory. Pacific Northwest National Lab is operated by Battelle for the DOE under Contract DE-AC05-76RL01830. Metagenomic sequencing for this research was performed by the Joint Genome Institute via a large-scale sequencing award (Award 1781) and at the Genomics Shared Resource Core at The Ohio State University Comprehensive Cancer Center supported by P30 CA016058.

Authors' contributions: Using the CRediT contributor roles,author contributions can be defined as follows:Conceptualization,JRR, MAB, JCS, and KCW; Data curation, JRR, MAB, and BBM; Formal analysis, JRR, MAB, BBM, GJS, LMS, RAD; Funding acquisition, JCS, KCW; Project administration, RAD, JCS, KCW; Investigation, JRR, MAB, BBM, GJS, SOP, CDN, DWH; Supervision, MSL, EBG, DWH, JCS, KCW; Writing- original draft, JRR, MAB, BBM, GJS, KCW; Writing- review and editing, JRR, MAB, BBM, GJS, JCS, KCW.

Acknowledgements: The authors would like to thank Tyson Claffey and Richard Wolfe for Colorado State University server management; Sandy Shew for management of computing resources retained from The Ohio State University Unity cluster; Dr. Pearlly Yan at the Genomics Shared Resource Core at The Ohio State University Comprehensive Cancer Center for management of metagenomic sequencing; and Dr. J John for the continuous support.

Author's information: J.R.R. and M.A.B. contributed equally to this work.

References

1. Boulton AJ, Findlay S, Marmonier P, Stanley EH, Valett HM. The functional significance of the hyporheic zone in streams and rivers. Annu Rev Ecol Syst. Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA; 1998;29:59–81.
2. Stegen JC, Johnson T, Fredrickson JK, Wilkins MJ, Konopka AE, Nelson WC, et al. Influences of organic carbon speciation on hyporheic corridor biogeochemistry and microbial ecology. Nat Commun. Nature Publishing Group; 2018;9:1–11.
3. Newcomer ME, Hubbard SS, Fleckenstein JH, Maier U, Schmidt C, Thullner M, et al. Influence of hydrological perturbations and riverbed sediment characteristics on hyporheic zone respiration of CO2 and N2. J Geophys Res Biogeosciences. Wiley Online Library; 2018;123:902–22.
4. Trimmer M, Grey J, Heppell CM, Hildrew AG, Lansdown K, Stahl H, et al. River bed carbon and nitrogen cycling: state of play and some new directions. Sci Total Environ. Elsevier; 2012;434:143–58.
5. 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. Elsevier; 2020;715:136920.
6. Hu M, Chen D, Dahlgren RA. Modeling nitrous oxide emission from rivers: a global assessment. Glob Chang Biol. Wiley Online Library; 2016;22:3566–82.
7. Beaulieu JJ, Tank JL, Hamilton SK, Wollheim WM, Hall RO, Mulholland PJ, et al. Nitrous oxide emission from denitrification in stream and river networks. Proc Natl Acad Sci. National Acad Sciences; 2011;108:214–9.
8. Caruso A, Boano F, Ridolfi L, Chopp DL, Packman A. Biofilm-induced bioclogging produces sharp interfaces in hyporheic flow, redox conditions, and microbial community structure. Geophys Res Lett. Wiley Online Library; 2017;44:4917–25.
9. Naegeli MW, Uehlinger U. Contribution of the hyporheic zone to ecosystem metabolism in a prealpine gravel-bed-river. J North Am Benthol Soc. North American Benthological Society; 1997;16:794–804.
10. Battin TJ, Kaplan LA, Newbold JD, Hendricks SP. A mixing model analysis of stream solute dynamics and the contribution of a hyporheic zone to ecosystem function. Freshw Biol. Wiley Online Library; 2003;48:995–1014.
11. Zhang M, Daraz U, Sun Q, Chen P, Wei X. Denitrifier abundance and community composition linked to denitrification potential in river sediments. Environ Sci Pollut Res. Springer; 2021;1–12.
12. Liu S, Wang H, Chen L, Wang J, Zheng M, Liu S, et al. Comammox Nitrospira within the Yangtze River continuum: community, biogeography, and ecological drivers. ISME J. Nature Publishing Group; 2020;14:2488–504.
13. Pinto OHB, Silva TF, Vizzotto CS, Santana RH, Lopes FAC, Silva BS, et al. Genome-resolved metagenomics analysis provides insights into the ecological role of Thaumarchaeota in the Amazon River and its plume. BMC Microbiol. BioMed Central; 2020;20:1–11.

14. 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. Elsevier; 2018;642:742–53.
15. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. Narnia; 2012;28:1420–8.
16. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. Cold Spring Harbor Lab; 2017;27:824–34.
17. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. PeerJ Inc.; 2019;2019:e7359.
18. Wu M, Scott AJ. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics. Oxford University Press; 2012;28:1033–4.
19. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat. Biotechnol. Nature Publishing Group; 2017. p. 725–31.
20. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. Nature Publishing Group; 2017;11:2864–8.
21. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res. :gkaa621.
22. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2018;36:1925–1927.
23. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. Oxford University Press; 2004;32:1792–7.
24. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. Oxford University Press; 2002;30:3059–66.
25. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. Oxford University Press; 2009;26:1641–50.
26. Castelle CJ, Hug LA, Wrighton KC, Thomas BC, Williams KH, Wu D, et al. Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. Nat Commun. Nature Publishing Group; 2013;4:2120.
27. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010;26:1608–15.

28. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. Nature Publishing Group; 2019;37:420–3.
29. Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. Non-classical protein secretion in bacteria. BMC Microbiol. BioMed Central; 2005;5:1–13.
30. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. PeerJ. PeerJ Inc.; 2015;3:e985.
31. Roux S, Adriaenssens EM, Dutilh BE, Koonin E V., Kropinski AM, Krupovic M, et al. Minimum information about an uncultivated virus genome (MIUVIG). Nat Biotechnol. Nature Publishing Group; 2019;37:29–37.
32. Merchant N, Lyons E, Goff S, Vaughn M, Ware D, Micklos D, et al. The iPlant collaborative: cyberinfrastructure for enabling data to discovery for the life sciences. PLoS Biol. Public Library of Science San Francisco, CA USA; 2016;14:e1002342.
33. Jang H Bin, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol. Nature Publishing Group; 2019;37:632–9.
34. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. Nature Publishing Group; 2015;10:845–58.
35. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. Alignment-free oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. Nucleic Acids Res. Oxford University Press; 2017;45:39–53.
36. Borton MA, Hoyt DW, Roux S, Daly RA, Welch SA, Nicora CD, et al. Coupled laboratory and field investigations resolve microbial interactions that underpin persistence in hydraulically fractured shales. Proc Natl Acad Sci. National Academy of Sciences; 2018;115:E6585–E6594.
37. Daly RA, Borton MA, Wilkins MJ, Hoyt DW, Kountz DJ, Wolfe RA, et al. Microbial metabolisms in a 2.5-km-deep ecosystem created by hydraulic fracturing in shales. Nat Microbiol. Nature Publishing Group; 2016;1:16146.
38. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. Nature Publishing Group; 2012;9:357.
39. Lê Cao K-A, Rossouw D, Robert-Granié C, Besse P. A sparse PLS for variable selection when integrating omics data. Stat Appl Genet Mol Biol. De Gruyter; 2008;7.
40. Nicora CD, Burnum-Johnson KE, Nakayasu ES, Casey CP, White III RA, Chowdhury TR, et al. The MPLEx protocol for multi-omic analyses of soil samples. J Vis Exp JoVE. MyJoVE Corporation; 2018;
41. Elias JE, Gygi SP. Target-decoy search strategy for mass spectrometry-based proteomics. Proteome Bioinforma. Springer; 2010. p. 55–71.
42. Kim S, Pevzner PA. MS-GF+ makes progress towards a universal database search tool for proteomics. Nat Commun. Nature Publishing Group; 2014;5:5277.

43. McGivern BB, Tfaily MM, Borton MA, Kosina SM, Daly RA, Nicora CD, et al. Decrypting bacterial polyphenol metabolism in an anoxic wetland soil. Nat Commun. Nature Publishing Group; 2021;12:1–16.
44. Weljie AM, Newton J, Mercier P, Carlson E, Slupsky CM. Targeted profiling: quantitative analysis of 1H NMR metabolomics data. Anal Chem. ACS Publications; 2006;78:4430–42.
45. Joesten WC, Kennedy MA. RANCM: a new ranking scheme for assigning confidence levels to metabolite assignments in NMR-based metabolomics studies. Metabolomics. Springer; 2019;15:5.
46. Moon K, Jeon JH, Kang I, Park KS, Lee K, Cha C-J, et al. Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. Microbiome. Springer; 2020;8:1–15.
47. Wolf YI, Silas S, Wang Y, Wu S, Bocek M, Kazlauskas D, et al. Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. Nat Microbiol. Nature Publishing Group; 2020;5:1262–70.
48. Peduzzi P. Virus ecology of fluvial systems: a blank spot on the map? Biol Rev. Wiley Online Library; 2016;91:937–49.
49. Püttker S, Kohrs F, Benndorf D, Heyer R, Rapp E, Reichl U. Metaproteomics of activated sludge from a wastewater treatment plant–A pilot study. Proteomics. Wiley Online Library; 2015;15:3596–601.
50. Rudney JD, Xie H, Rhodus NL, Ondrey FG, Griffin TJ. A metaproteomic analysis of the human salivary microbiota by three-dimensional peptide fractionation and tandem mass spectrometry. Mol Oral Microbiol. Wiley Online Library; 2010;25:38–49.
51. Solden LM, Naas AE, Roux S, Daly RA, Collins WB, Nicora CD, et al. Interspecies cross-feeding orchestrates carbon degradation in the rumen ecosystem. Nat Microbiol. Nature Publishing Group; 2018;3:1274.
52. Graham EB, Crump AR, Resch CT, Fansler S, Arntzen E, Kennedy DW, et al. Deterministic influences exceed dispersal effects on hydrologically-connected microbiomes. Environ Microbiol. Wiley Online Library; 2017;19:1552–67.
53. Yang F, Bogdanov B, Strittmatter EF, Vilkov AN, Gritsenko M, Shi L, et al. Characterization of Purified c-Type Heme-Containing Peptides and Identification of c-Type Heme-Attachment Sites in Shewanella o neidenis Cytochromes Using Mass Spectrometry. J Proteome Res. ACS Publications; 2005;4:846–54.
54. Cordero PRF, Bayly K, Leung PM, Huang C, Islam ZF, Schittenhelm RB, et al. Atmospheric carbon monoxide oxidation is a widespread mechanism supporting microbial survival. ISME J. Nature Publishing Group; 2019;13:2868–81.
55. Xia X, Zhang S, Li S, Zhang L, Wang G, Zhang L, et al. The cycle of nitrogen in river systems: sources, transformation, and flux. Environ Sci Process \& Impacts. Royal Society of Chemistry; 2018;20:863–91.
56. Strauss EA, Lamberti GA. Regulation of nitrification in aquatic sediments by organic carbon. Limnol Oceanogr. Wiley Online Library; 2000;45:1854–9.

57. Brust GE. Management strategies for organic vegetable fertility. Saf Pract Org food. Elsevier; 2019. p. 193–212.
58. Cavaliere M, Feng S, Soyer OS, Jiménez JI. Cooperation in microbial communities and their biotechnological applications. Environ Microbiol. Wiley Online Library; 2017;19:2949–63.
59. Gibbons SM, Jones E, Bearquiver A, Blackwolf F, Roundstone W, Scott N, et al. Human and environmental impacts on river sediment microbial communities. PLoS One. Public Library of Science; 2014;9:e97435.
60. Wrage N, Velthof GL, Van Beusichem ML, Oenema O. Role of nitrifier denitrification in the production of nitrous oxide. Soil Biol Biochem. Elsevier; 2001;33:1723–32.
61. Strauss EA, Richardson WB, Bartsch LA, Cavanaugh JC, Bruesewitz DA, Imker H, et al. Nitrification in the Upper Mississippi River: patterns, controls, and contribution to the NO3- budget. J North Am Benthol Soc. 2004;23:1–14.
62. Stegen JC, Fredrickson JK, Wilkins MJ, Konopka AE, Nelson WC, Arntzen E V, et al. Groundwater–surface water mixing shifts ecological assembly processes and stimulates organic carbon turnover. Nat Commun. Nature Publishing Group; 2016;7:1–12.
63. Bartossek R, Nicol GW, Lanzen A, Klenk H-P, Schleper C. Homologues of nitrite reductases in ammonia-oxidizing archaea: diversity and genomic context. Environ Microbiol. Wiley Online Library; 2010;12:1075–88.
64. Jones CM, Spor A, Brennan FP, Breuil M-C, Bru D, Lemanceau P, et al. Recently identified microbial guild mediates soil N 2 O sink capacity. Nat Clim Chang. Nature Publishing Group; 2014;4:801–5.
65. Kuypers MMM, Marchant HK, Kartal B. The microbial nitrogen-cycling network. Nat Rev Microbiol. Nature Publishing Group; 2018;16:263–76.
66. Nelson WC, Graham EB, Crump AR, Fansler SJ, Arntzen E V, Kennedy DW, et al. Distinct temporal diversity profiles for nitrogen cycling genes in a hyporheic microbiome. PLoS One. Public Library of Science San Francisco, CA USA; 2020;15:e0228165.
67. Murphy CL, Sheremet A, Dunfield PF, Spear JR, Stepanauskas R, Woyke T, et al. Genomic Analysis of the Yet-Uncultured Binatota Reveals Broad Methylotrophic, Alkane-Degradation, and Pigment Production Capacities. MBio. Am Soc Microbiol; 2021;12.

Figures

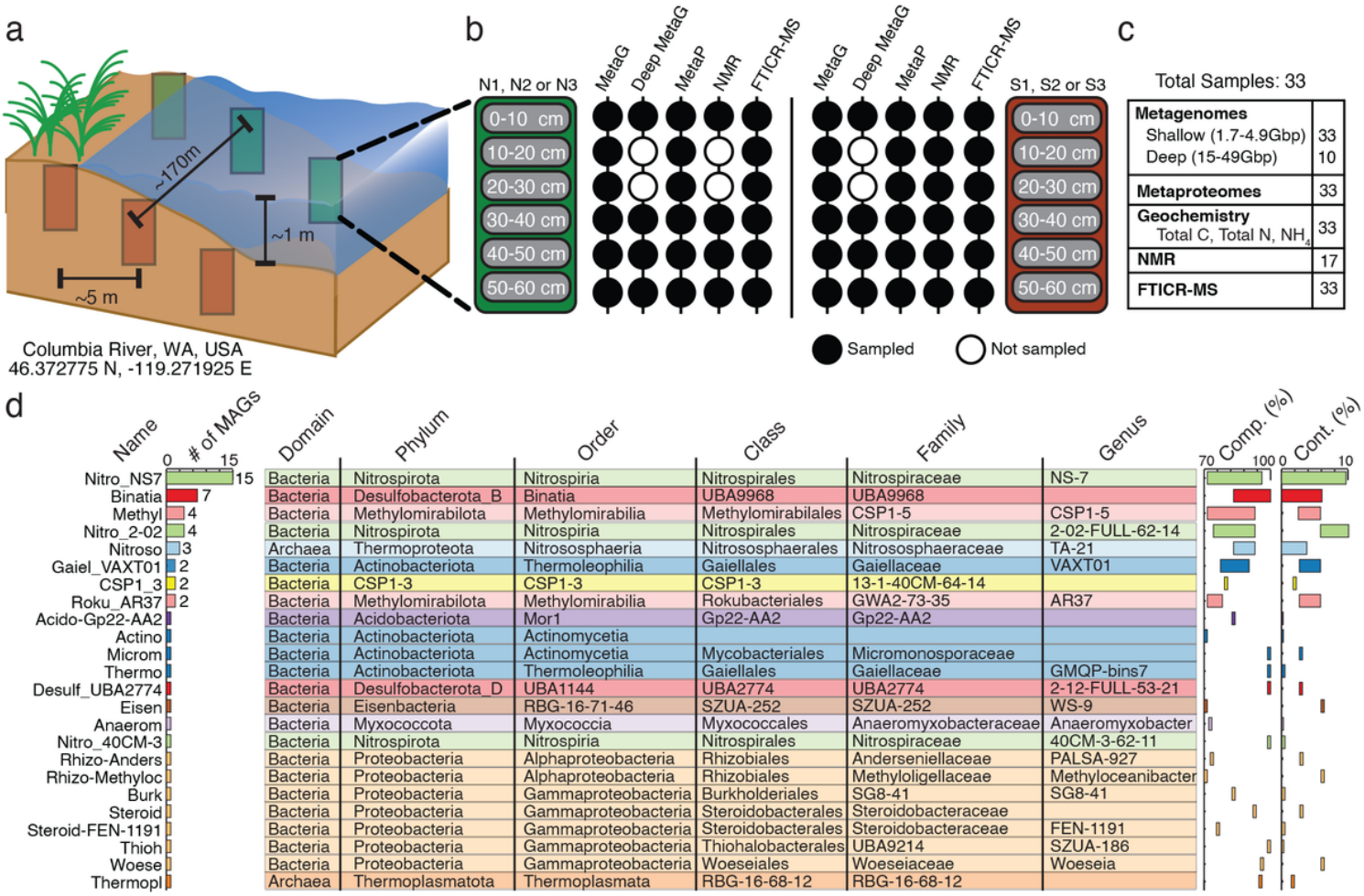

Figure 1

Overview of hyporheic zone sampling and the microbial genomes included in the HUM-V database a) Samples were collected from two transects, each with 3 sediment cores, with each core sectioned into 10-centimeter segments from 0-60 centimeters in depth and paired with metaproteomics and geochemistry. b) Schematic of the data types available for each of the depth samples within a core. Black-filled circles indicate depth samples for a particular data type, while the open circles denote missing analyses (due to limited sample availability). c) Summarized catalog of the total samples for each analysis. d) The total recovered genomes (# of MAGs), taxonomic string, inferred genome completion (Comp., %), and contamination (Cont., %) for the dereplicated microbial genomes retained in HUM-V.

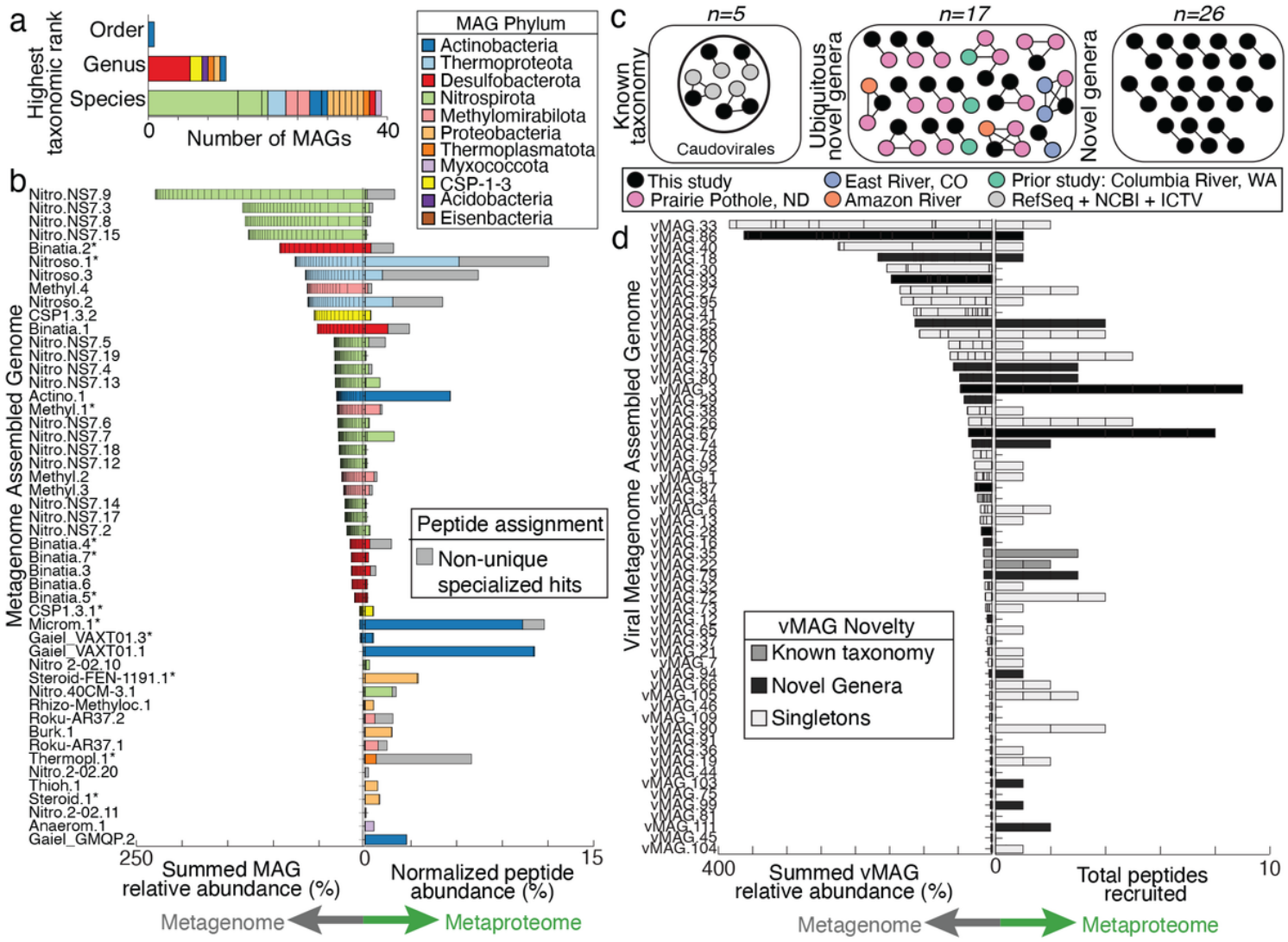

Figure 2

HUM-V database coupled to metaproteomics reveals active members in the hyporheic zone microbiome. a) Stacked bar graph indicates the taxonomic novelty of the de-replicated MAGs colored by phylum and stacked according to the first empty position within the taxonomic string provided by the Genome Taxonomy Database GTDB-Tk. Each color represents a MAG phylum according to the MAG Phylum legend, a coloring maintained across this manuscript. b) Butterfly plot reports the summed genomic relative abundance across all samples (left side) and the normalized mean proteomic relative abundance (right side) for dereplicated MAGs (55 total, 49 shown), with bars colored by phylum. A butterfly plot with all MAGs is shown in Additional File 3: Figure S3. MAGs that contain a partial or complete 16S rRNA sequence are denoted with and asterisk (*). Non-unique peptide assignment is defined in the methods and is shown with grey bars. c) Similarity network of the few vMAGs from our study (black) that clustered to viruses belonging to the default RefSeq, ICTV and NCBI Taxonomy databases (gray), as well as clustering of our vMAGs to other freshwater, publicly available dataset we mined (Pink, Purple, Orange, and Turquoise). The remaining clusters of viruses that were novel (e.g., did not cluster with prior viral genomes) are shown, with the full network file including singletons shown in Additional File 6. d) Butterfly

plot showing summed genomic relative abundance (left side) and total peptides recruited for each vMAG population (total 111, 58 shown). Bars are colored by clustering of vMAGs from this study with (i) viruses of known taxonomy in RefSeq, ICTV and NCBI Taxonomy (dark grey), (ii) novel genera, both only from this study and ubiquitous (black), and no clustering from any database (light grey, singletons).

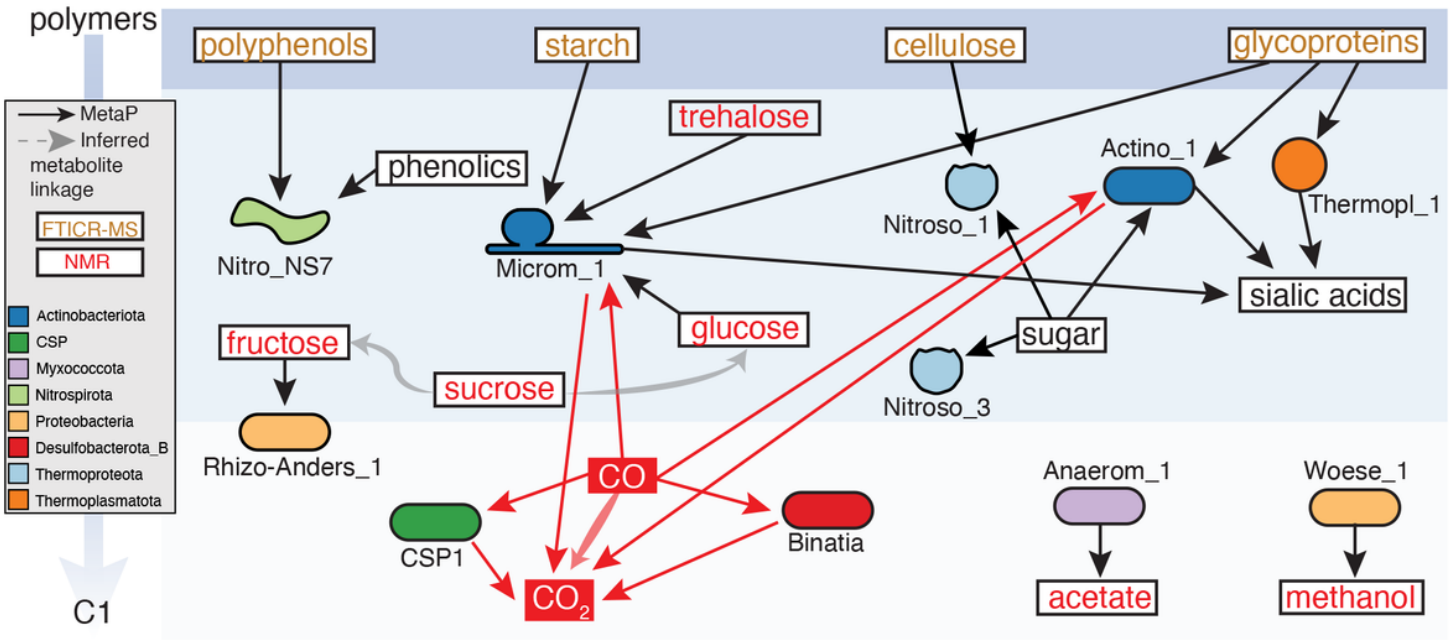

Figure 3

Metaproteomics and metabolomics reveal microbial metabolic handoffs that support carbon cycling in river sediments. Detected metabolites are given in boxes, with NMR-detected compounds listed in red, polymers from FTICR-MS in orange, undetected metabolites in black. These polymers were inferred from FTICR-MS assigned biochemical classes and the specificity of CAZymes detected in metaproteome, where starch and cellulose were within the “polysaccharide-like” class and glycoproteins were in the “amino sugar-like” class. MAG-resolved metaproteome information is indicated by solid arrows, with MAG shape colored by phylum. Red arrows indicate processes leading to CO2 production, while black arrows indicate other microbial carbon transforming genes expressed in the proteome. Shaded bold arrows indicate chemical connections, where (1) grey indicates a metabolite was detected along with putative downstream products (e.g., sucrose conversion to glucose) but metaproteomic lacked evidence for the transformation or (2) red indicates a metabolite not measured but metaproteomic evidence supported transformation (e.g., CO conversion to CO2).

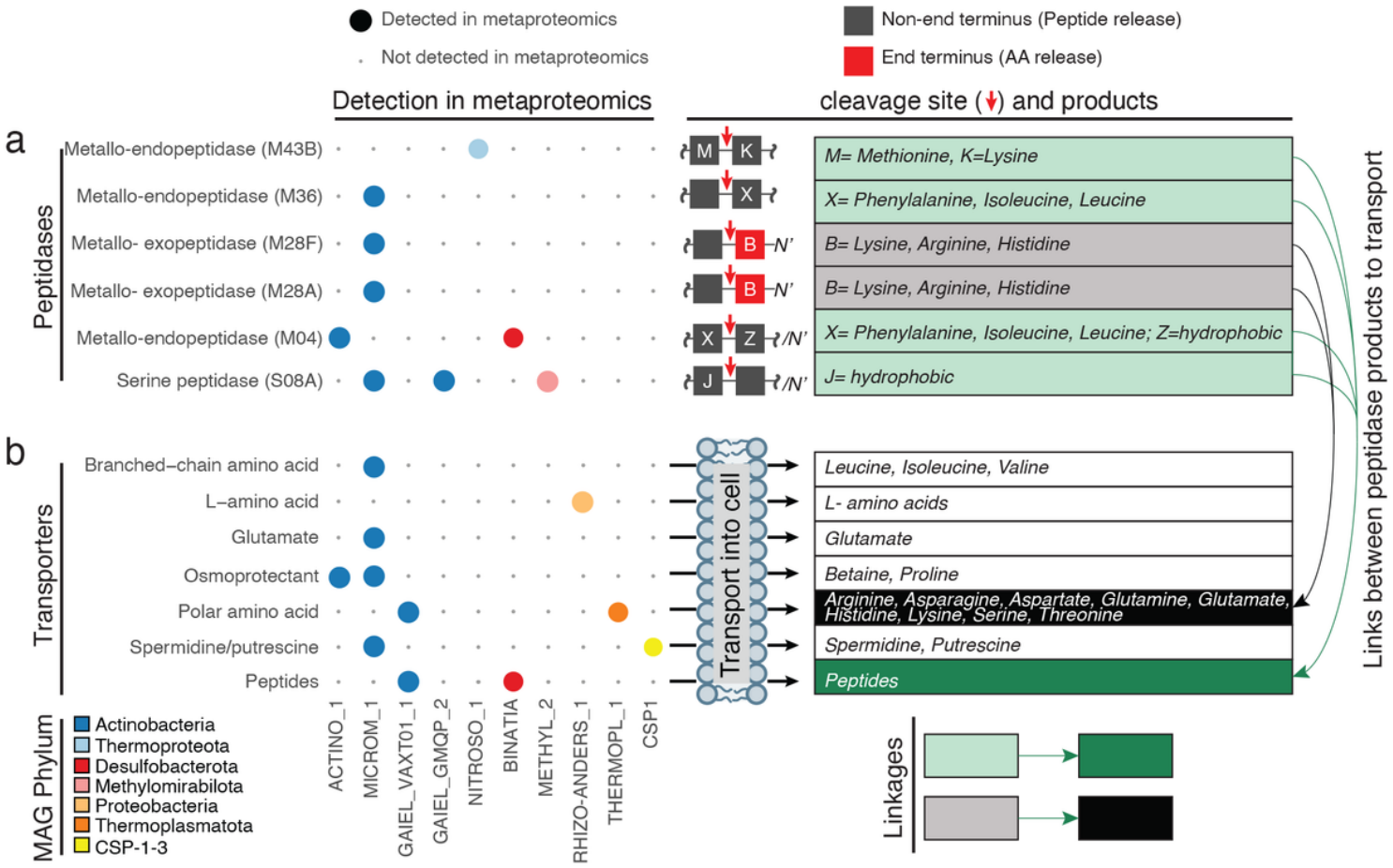

Figure 4

Organic nitrogen mineralization and cellular transport are active microbial processes in river sediments. Bubble plots indicate the expressed genes that were uniquely assigned to specific genome including (a) extracellular peptidases and (b) cellular transporters for organic nitrogen. Unique peptides detected in at least 3 samples are reported as bubbles and colored by phylum. Table on the right shows putative amino acids cleaved or transported by respective peptidases or transporters, shades of color (green or grey) denote peptides that are cleaved into amino acids that could be transported, providing linkages between extracellular organic nitrogen transformation and transport of nitrogen into the cell. White boxes indicate an organic nitrogen transporter that recruited peptides but could not be linked to outputs of specific peptidases.

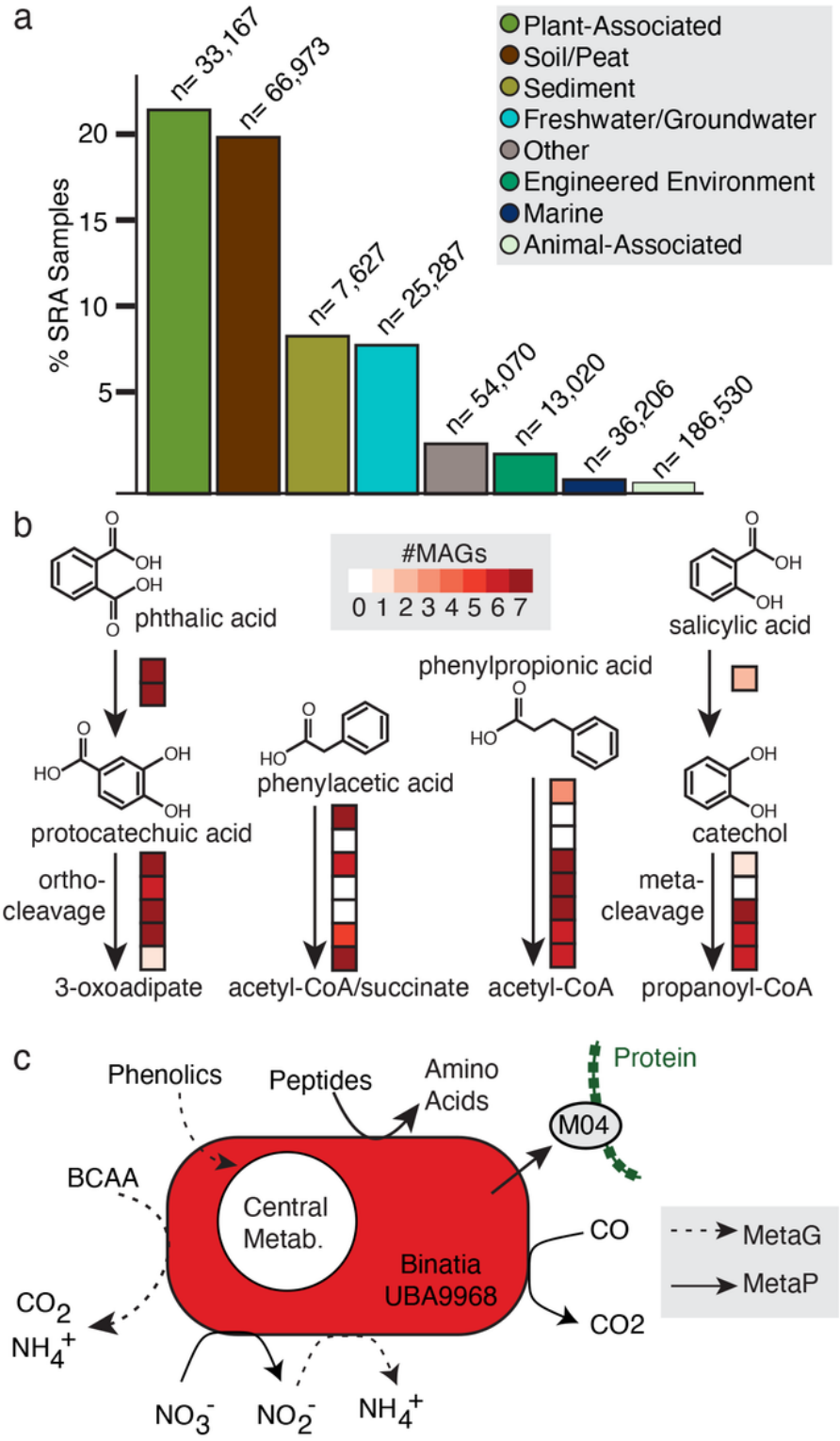

Figure 5

Uncultured Binatia are widely dispersed across ecosystems and express nitrogen and carbon genes in situ. (a) 16S rRNA comparisons between the HUM-V Binatia genomes and other microbiome datasets reveal that these uncultivated HUM-V species are prevalent in terrestrial and aquatic microbiomes. Graph shows the % of samples in the SRA database with reads assigned to Binatia. (b) The seven HUM-V genomes reconstructed here encode diverse pathways for transforming phenolic compounds. Pathways

are shown with boxes corresponding to each enzyme in the pathway, colored by the number of Binatia MAGs that encode each step. Gene information shown here is reported in detail in Additional File 5. (c) Metabolic genome cartoon for the major functions encoded by the Binatia MAGs. Dotted arrows are functions encoded in the metagenome, with black arrows corresponding to metaproteome-detected enzymes. (BCAA, Branched chain amino acids).

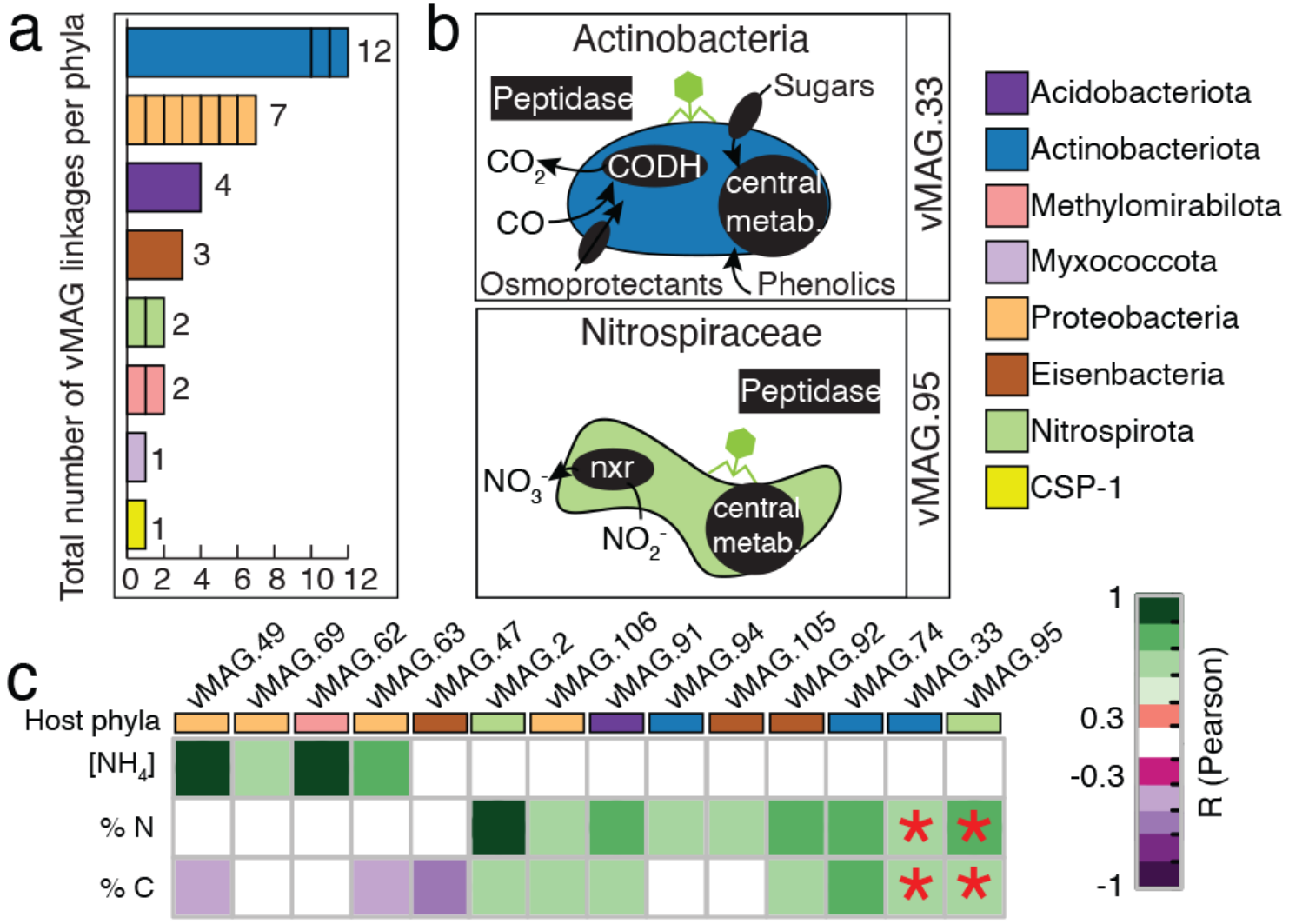

Figure 6

Evidence that viruses could impact microbial host metabolism and river geochemistry. a) Stacked bar chart of the total number of vMAGs (n=32) that have putative host linkages. Each bar represents a phylum and lines within bars indicate the linkages for specific genomes within each phylum. For example, there are three genomes within the Actinobacteriota phylum that collectively have 12 viral linkages and of the three genomes that have linkages, one host has 10 viruses linked, while the other two hosts have 1 virus linked. b) Genome cartoons of microbial metabolisms for two representative genomes that could be predated by vMAGs, with the genes shown in black text boxes denoting processes detected in proteomics. These two microorganisms were selected as examples because they were active members in shaping carbon and nitrogen metabolism in these river sediments but could be impacted by viral

predation; other virus-host relationships are reported in Additional File 3: Figure S10. c) Heatmap reports correlations between a subset of vMAGs with rectangle colors denoting the putative phyla for the respective host. Correlations between these vMAGs and ecosystem geochemistry (NH4 µg/gram, %N, %C) are reported with significant correlation coefficients denoted by purple-green shading according to the legend. Red asterisks (*) indicate the vMAG relative abundance predicted a key environmental variable by sparse partial least squares (sPLS) regression. Note a subset of these predicted vMAGs are shown in 6b.

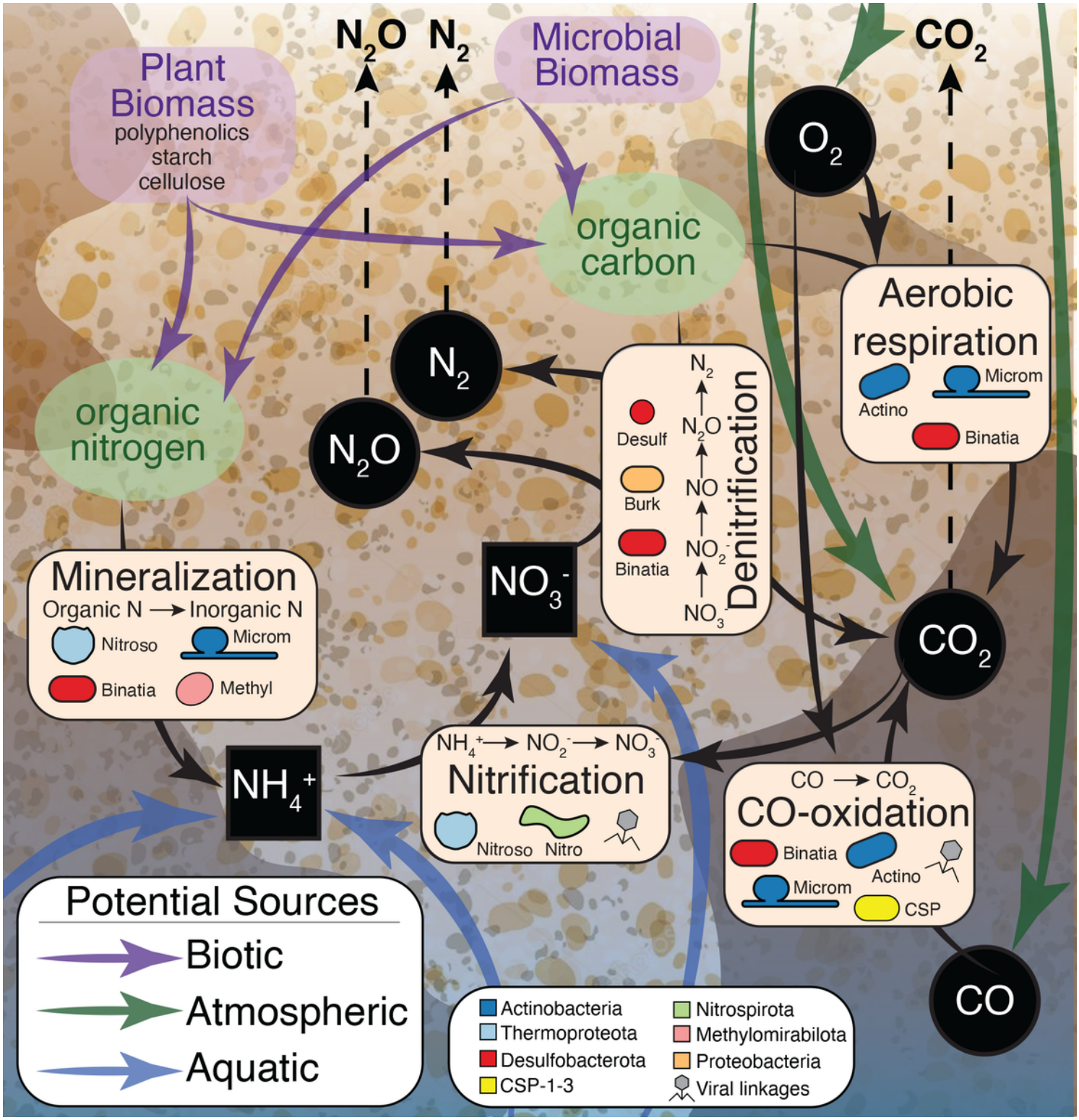

Figure 7

Conceptual model uncovering microbes and processes contributing to carbon and nitrogen cycling in river sediments. Integration of multi-omic data uncovered the microbial and viral effects on carbon and nitrogen cycling in river sediments. Black arrows signify microbial transformations uncovered in our metaproteomic data. Specific processes (e.g., mineralization, nitrification, CO-oxidation, denitrification, and aerobic respiration) are highlighted in beige boxes, with microorganisms inferred to carry out the specific process denoted by overlaid cell shapes colored by phylum. Prior to this research, little was known about the specific enzymes and organisms responsible for river organic nitrogen mineralization and CO-oxidation, thus this research adds new content to microbial roles in carbon and nitrogen transformations in these systems. Possible (biotic, atmospheric, and aquatic) carbon and nitrogen sources are shown by purple, green, and blue arrows respectively. Inorganic carbon and nitrogen sources are shown by black squares (aqueous) and black circles (gaseous) with white text and dashed arrows indicating possible gasses that could be released to the atmosphere. Processes that could be impacted by viruses are marked with grey viral symbols.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1.xlsx
- AdditionalFile2.xlsx
- AdditionalFile3.docx
- AdditionalFile4.zip
- AdditionalFile5.xlsx
- AdditionalFile6.xlsx
- AdditionalFile7.xlsx
- AdditionalFile8.xlsx
- AdditionalFile9.xlsx