Depth-related variability in viral communities in highly stratified sulphidic mine tailings

Shao-Ming Gao
Sun Yat-Sen University
Axel Schippers
Federal Institute for Geosciences and Natural Resources (BGR)
Nan Chen
Sun Yat-Sen University
Yang Yuan
Sun Yat-Sen University
Miao-Miao Zhang
Sun Yat-Sen University
Qi Li
Sun Yat-Sen University
Bin Liao
Sun Yat-Sen University
Wen-Sheng Shu
South China Normal University
Li-Nan Huang (eseshln@mail.sysu.edu.cn)
Sun Yat-Sen University https://orcid.org/0000-0002-4881-7920

Research

Keywords: Stratified mine tailings, Viruses, Diversity, Functions, Auxiliary metabolic genes

Posted Date: December 6th, 2019

DOI: https://doi.org/10.21203/rs.2.18336/v1

Abstract

Background: Recent studies have significantly expanded our knowledge of viral diversity and functions in the environment. Exploring the ecological relationships between viruses, hosts and the environment is a crucial first step towards a deeper understanding of the complex and dynamic interplays among them.

Results: Here, we obtained extensive 16S rRNA gene amplicon, metagenomics sequencing and geochemical datasets from different depths of two highly stratified sulphidic mine tailings cores with steep geochemical gradients especially pH, and explored how variations in viral community composition and functions were coupled to the co-existing prokaryotic assemblages and the varying environmental conditions. Our data showed that many viruses in the mine tailings represented novel genera, based on gene-sharing networks. *Siphoviridae* and *Myoviridae* dominated the classified viruses in the surface tailings and deeper layers, respectively. Both viral richness and normalized coverage increased with depth in the tailings cores and were significantly correlated with geochemical properties, for example, pH. Viral richness was also coupled to prokaryotic richness (Pearson's $r = 0.75$, $P < 0.01$). The enrichment of prophages in the surface mine tailings revealed a preference of lysogenic viral lifestyle in more acidic conditions. Community-wide comparative analyses clearly showed that viruses in the surface tailings acquired genes related to low pH adaptation from archaea while viruses in the deeper layers contained genes mainly annotated as conventional viral functions. Notably, abundant auxiliary sulfate reduction genes were identified from the deeper tailings layers and they were widespread in viruses predicted to infect diverse bacterial phyla.

Conclusions: Overall, our results revealed a depth-related distribution of viral populations in the extreme and heterogeneous tailings system. The viruses may interact with diverse hosts and dynamic environmental conditions and likely play a role in the functioning of microbial community and modulate sulfur cycles *in situ*.

Background

Viruses are abundant and critical components of microbial communities in the environment [1]. Historically, studies of viral diversity have largely relied on culture-dependent techniques with well recognized limitations, including especially the inconsistency between morphological and genetic taxon identification [2]. While maker gene surveys have revolutionized our understanding of cellular systematics and diversity, such approaches cannot be adopted in viral ecology studies due to absence of a phylogenetically informative universal marker owing to the mosaic nature of viral genome organization [3]. To tackle these problems, recent works have employed metagenomic sequencing to discover viral sequences from a wide variety of habitats including marine and freshwater environments [3, 4, 5], soils [6, 7], and extreme environments [8, 9, 10]. These studies often reveal the existence of diverse viral assemblages in nature, whose members remain largely uncharacterized ('unknown virosphere'), and significantly improve our understanding of the ecological roles of viruses in Earth's major ecosystems [4, 11]. A current challenge is to move beyond the two basic questions, i.e., what is there and what is it doing,

to a more in-depth analysis of the dynamic interplay between viruses, microbes and environmental conditions [12].

Viruses can substantially affect the ecology, evolution and physiology of their hosts in natural settings by causing host mortality, facilitating horizontal gene transfer, and influencing biogeochemical cycles via production of dissolved organic matter through cell lysis or participate in host metabolisms with auxiliary metabolic genes (AMGs) [13, 14]. In the meantime, viruses are intracellular obligatory parasites that repurpose the host cell machinery to replicate; thus, prokaryotic hosts play a key role in regulating viral populations [9]. Population oscillations of viruses and their hosts have been documented [5] and reviewed [15] in natural and cultivated environments. Furthermore, geochemical conditions may also have a significant influence on viral populations via direct or indirect mechanisms. Analyses of viruses in the pelagic upper-ocean revealed that viral communities are locally structured by environmental conditions that affect host community structure [16]. Additionally, the AMGs in viral genomes are obtained by horizontal gene transfer from their hosts, and exhibit parallel depth-stratified host adaptations [14]. All these aspects imply a complicated interaction between viruses, hosts and the environments.

Acid mine drainage (AMD) is a worldwide environmental problem that arises largely from microbially-mediated oxidative dissolution of sulphidic ores exposed to oxygen and water during mining activities [17]. These environments are characterized by low pH and high concentrations of metals and sulfate, representing an extreme environment for life. AMD environments are well recognized as model systems for the study of microbial community structure, functions and evolution due to their reduced complexity and have been studied extensively by cultivation-independent molecular approaches [18, 19, 20]. Meanwhile, several investigations with a specific focus on viruses in AMD systems have been reported. These early works documented a major influence of minerals (via attachment) on viral abundance [21, 22], unveiled the coevolution relationships between viruses and their specific hosts [23], and uncovered viruses infecting cells of the archaeal lineages of ARMAN and Thermoplasmatales [24]. In contrast, while waste tailings dumps are an important source of AMD around the globe [20], relatively little is known about the microbial diversity and ecology in these harsh, highly heterogeneous environments [25, 26], and the indigenous viral communities have never been investigated. Mine tailings dumps are typically stratified into distinct geochemical zones, reflecting progressive oxidation of sulfide minerals in the tailings and indicating that each of these zones is shaped by organisms with specific metabolic traits [26]. Thus, mine tailings offer unique possibilities to resolve complex biological interactions and to explore the relationship between these dynamic interactions and multivariate geochemistry.

Here we report the analysis of two highly stratified tailings cores sampled from a sulphidic tailings impoundment of a Pb/Zn mine where extremely low pH and metal-rich drainage is a persistent feature. The composition of both the prokaryotic and of viral populations in different sections of the cores was resolved by 16S rRNA gene high-throughput sequencing and recovering viral sequences from metagenomic datasets, respectively. We assessed how prokaryotic and viral communities varied along

the tailings depth profiles and examined how the down-core stratification of viral diversity and functions were related to the co-existing prokaryotic assemblages and tailings geochemistry.

Materials And Methods

Study site, sampling and physicochemical analyses

The Fankou Pb/Zn sulphidic mine tailings site (25˚2′56.5″N, 113˚39′48.5″E) is located in Shaoguan, Guangdong province, China. Extremely acidic, heavy metals rich drainage is a persistent feature due to microbially mediated dissolution of sulfide minerals in the tailings at this site. Previous 16S rRNA surveys have documented vertical stratification of geochemistry and prokaryotic populations, with acidophilic archaea, mostly *Ferroplasma spp.* in the *Thermoplasmatales* predominant in the upper layers of tailings (oxidized zones and the oxidation front) [26]. Two tailings cores (inner diameter, 8 cm; length, 60 cm) were sampled from an area covered with AMD using a sampling collector in October 2017. After retrieval, the cores were immediately sectioned into distinct layers based on their physical feature and appearance (e.g., colors), yielding six layers for core A and five layers for core B (Additional file 1: Figure S1). Each of the 11 tailings layers was collected in 50 ml sterile tubes, kept in an icebox and transported to the laboratory, where the samples were stored at 4 °C prior to subsequent analyses.

Air-dried subsamples were analyzed with standard methods for the determination of total organic carbon (TOC) (TOC-VCPH; Shimadzu, Columbia, MD), total nitrogen (TN) and phosphorus (TP) (SmartChem; Westco Scientific Instruments Inc., Brookfield, CT). The pH and electrical conductivity (EC) were measured in a 1:2.5 (w/v) aqueous solution using a pH meter and an EC meter. HCl-extractable ferrous iron was determined by the 1, 10-phenanthroline method at 530 nm [27], and sulfate (SO_4^{2-}) was measured by a $BaSO_4$-based turbidimetric method [28]. Total concentrations of heavy metals (including Pb, Zn, Cu, Cr, Mn, and As) and sulfur (TS) were determined by inductively coupled plasma optical emission spectrometry (ICP-OES; Optima 2100DV, PerkinElmer, Wellesley, MA) and an elemental analyzer (Vario EL, Elementar, Germany), respectively.

DNA extraction and 16S rRNA amplicon and metagenomic sequencing

Total community genomic DNA was extracted using the FastDNA Spin kit (MP Biomedicals, Irvine, CA) according to the manufacturer's instructions. The V4 region of bacterial and archaeal 16S rRNA genes was amplified with prokaryotic universal primers F515 (5'-GTGCCAGCMGCCGCGGTAA-3') and R806 (5'-GGACTACVSGGGTATCTAAT-3') [29]. A sample-specific 8-bp error-correcting barcode was added to the reverse primer. PCR amplification was conducted in triplicate in 50-μl reaction mixtures following the thermal cycling procedure described previously [30, 31]. Replicate PCR reactions from each sample were pooled and concentrated and purified using a QIAquick Gel Extraction Kit (Qiagen, Chatsworth, CA). A single composite sample was prepared by combining an approximately equimolar amount of PCR product from each tailings sample and then sequenced on an Illumina MiSeq platform (Illumina, San Diego, CA) (250bp, paired end reads). To obtain metagenomic data, extracted DNA was purified using a

QIAquick Gel Extraction Kit (Qiagen, Chatsworth, CA), quantified with Qubit (Thermo Fisher Scientific, Australia), and then randomly amplified using Illustra^{TM} GenomiPhi^{TM} V3 DNA Amplification Kit (GE Health Care, United Kingdom). The amplified products were used for library preparation with NEBNext Ultra II DNA Prep Kit (New England Biolabs, Ipswich, MA) and sequenced with MiSeq Reagent Kit v3 on an Illumina MiSeq platform (150bp, paired end reads). Finally, 50-Gigabyte sequence data was obtained for each of the samples.

Processing of 16S rRNA and metagenomic sequence data

16S rRNA raw data were processed and analyzed with the Mothur software package (version 1.38.1) and QIIME (1.9.0) [32, 33]. Briefly, obtained short reads were noise reduced to minimize sequencing error by using the commands of 'shhh.flows' and 'pre.cluster' in Mothur [32]. Then, putative chimeric sequences were identified and removed by using Chimeric Uchime [34]. Pair-end reads were assembled via the 'make.contigs' command, and the primers and barcodes in assembled sequences were removed using the 'trim.seqs' commond [32]. Operational taxonomic units (OTUs) were identified by clustering assembled sequences at the 97% similarity level using UCLUST algorithm [34]. Taxonomic classification of the phylotypes was determined based on the Ribosomal Database Project at a default threshold of 80% [35]. Finally, the non-rarified OTU table (table of counts of OTUs on a per-sample basis with singleton OTUs excluded) and OTU taxonomy were converted to a 'biom' format to obtain prokaryotic community composition at different taxonomic levels by using the script of 'summarize_taxa_through_plots.py' in QIIME [33, 36, 37].

Metagenomic reads were quality filtered and trimmed using in-house Perl scripts [38]. A trim quality threshold of 20 was used and reads containing more than 5 'N' were discarded. All quality-controlled reads from a tailings core were cross-assembled using SPAdes 3.9.0 and kmers of 21, 33, 55, 77, 99, 127 under the '–meta' mode [39]. Genes were predicted by Prodigal 2.6.3 (with the parameters set as "-p meta -g 11 -f gff -q -m -c") [40], and functional annotation was performed through assignment of predicted proteins to the Pfam 32.0 [41], Kyoto Encyclopedia of Genes and Genomes (KEGG) database [42], and Non-supervised Orthologous Groups (EggNOG v5.0.0) [43]. Briefly, predicted proteins were compared to Pfam database by using the InterProscan 5.0 software with settings of "-appl Pfam -irplookup" and the lowest E-value as the best hits. Additionally, blastp was used to assign viral proteins to KEGG and EggNOG database to get KO and COG terms (E-value: 10^{-5}).

To access the dynamics of individual scaffolds and genes, sequencing reads from each library were mapped onto sequences using Bowtie2 with default parameters [44]. The normalized coverage for a given scaffold or gene was computed as the average scaffold or gene coverage (that is, the number of nucleotides mapped to the scaffold or gene divided by the scaffold or gene length) divided by the number of reads in a given library and multiplied by the mean value of the number of reads in the 11 libraries [5].

Identification and clustering of viral scaffolds

Three methods were applied to identify viral scaffolds in the metagenomic assemblies: (1) viral protein families generated with isolate reference viruses and viral scaffolds identified from a collection of geographically and ecologically diverse samples according to metadata from the Integrated Microbial Genomes with Microbiome Samples (IMG/M) system [45], (2) VirSorter software based on the identification of viral hallmark genes, enrichment in hypothetical proteins and other viral signatures [46], and (3) VirFinder software applying a k-mer frequency based machine learning method [47]. First, viral protein family models were used as a bait to screen metagenomic scaffolds longer than 5 kb and then filtered by inspecting the number of genes covered with viral protein families, Pfams and KO terms, as previously described [45]. Next, metagenomic scaffolds longer than 3kb were processed with VirSorter using the Viromes database [46]. Predicted viral scaffolds in the categories 1 and 2 were then manually curated as described previously [48]. For scaffolds in the categories 4 and 5, only predicted prophage regions were retained. Then, VirFinder was applied to search all scaffolds longer than 1kb, q-values were computed for the identified viral scaffolds, and the scaffolds having q-values < 0.05 were predicted as viruses. Finally, if the viral scaffolds predicted by viral protein families and VirFinder contain a prophage prediction, these viral scaffolds were removed from the viral sequence pools identified by these two methods before all identified viral scaffolds were merged.

All viral scaffolds were clustered into viral OTUs (vOTUs) using the parameters 95% average nucleotide identity and 85% alignment fraction of the smallest scaffolds [49]. To place the viral scaffolds in the context of known viruses, a gene-content based network analysis was used to cluster viral scaffolds into viral clusters (VCs). Briefly, predicted proteins from viral scaffolds were clustered with predicted proteins from isolate reference viruses in the NCBI database (dsDNA viruses, ssDNA viruses and retroviruses combined) [50] based on all versus-all blastp search with an E-value of 10^{-3}, and protein clusters were defined with the Markov clustering algorithm and processed using vConTACT v.2.0 [51, 52].

Reconstruction of prokaryotic genomes and host prediction of viral scaffolds

All cross-assembled scaffolds longer than 2.5kb were binned using MetaBAT v2.12.1 [53], MaxBin v2.2.2 [54], Abawaca v1.00 (https://github.com/CK7/abawaca), and Concoct v0.4.0 [55] with default parameters, considering tetranucleotide frequencies, scaffolds coverage and GC content, and then the results were combined using DASTool [56]. Bins were further manually curated to obtain high-quality genomes using RefineM v0.0.24 [57]. In detail, the automatic binning methods may separate a “true” genome bin into two or more smaller, separate bins. Bins that shared a similar coverage range, GC content and identical taxonomic classifications as determined by CheckM v1.0.7 [58] were grouped into a single bin. Additionally, scaffolds with incongruent taxonomic classification and incongruent 16S rRNA genes were removed as implemented in RefineM v0.0.24 [57]. The completeness and contamination of genome bins were assessed using CheckM v1.0.7 [58], and genomes estimated to be more than 50% complete and less than 10% contaminated were classified using the genome taxonomy database (GTDB-Tk v0.3.0) [59].

Viral scaffolds were putatively linked to their hosts *in silico* [60]. Briefly, these linkages were based on (1) shared genomic content between viral scaffolds and host genomes, (2) prophages identified in host genomes, and (3) sequence similarity between spacers in microbial CRISPR regions and in the viral scaffolds. All viral scaffolds were compared to the recovered host genomes (E-value $\leq 10^{-3}$, bit score $\geq$ 50, alignment length $\geq$ 2.5 kb and identity $\geq$ 70%) using blastn [4]. Viral sequences identified as prophage were matched to their corresponding host genomes. CRISPR spacers were recovered from metagenomic scaffolds using metaCRT with default parameters [61]. Extracted spacers were compared to viral scaffolds using blastn with thresholds of no mismatches over the whole spacer length and an E-value $\leq 10^{-10}$ [1, 4].

Analysis of AMGs

Viral genes predicted by Prodigal [40] were assigned to EggNOG v5.0.0 database [43] using blastp (threshold of 50 for bit score and 10^{-5} for E-value). Viral AMGs assigned as COG0175 (sulfate reduction) were identified in the viral genomes [62], and then compared to the protein sequences in EggNOG v5.0.0 database [43] (blastp, threshold of 50 for bit score and 10^{-3} for E-value) to recruit relevant reference sequences (up to 20 for each viral AMG sequence) [4]. These sets of viral AMGs and related protein sequences were then aligned with Muscle v3.8.31 [63] and filtered by TrimAL 1.2rev59 [64] to remove columns comprised of more than 95% gaps. Phylogenetic trees were reconstructed using RAxML (version 8.2.8 with the parameters set as "-f a -m GTRGAMMA -n boot -c 25 -p 12345 -x 12345") [65]. The resulting newick file with the best tree topology determined as with the best likelihood score was uploaded to iTOL v4 [66] for visualization and formatting.

Statistical analyses

All statistical analyses were implemented with various packages within the statistical program R. Pearson correlations were performed using 'rcorr' function (Hmisc package) to assess the relationships between the diversity of viruses, prokaryotes and environmental variables in all samples. Bray–Curtis distances were used to construct the dissimilarity matrices for prokaryotic and viral community structure and function profiles, whereas Euclidean distances were calculated using standardized environmental variables (vegan 2.5-4). Permutational multivariate analysis of variance ('Adonis' function; 999 permutations) was used to test for significant differences between classified groups of samples (vegan 2.5-4). Mantel tests were performed to reveal the correlations between the dissimilarity matrices (vegan 2.5-4). Statistical significance of differences in normalized coverage of a given gene or COG between two datasets was determined using non-parametric *Wilcoxon* t-test (unpaired), with confidence intervals at 99% significance and Benjamini–Hochberg correction ($P < 0.05$).

Results

Physicochemical stratification of mine tailings

Both tailings cores showed steep gradients of physicochemical properties (Fig. 1). pH values shifted from extremely acidic at the surface layers to near neutral at the deeper layers, while electronic conductivity (EC) declined with depth along the vertical profiles. Both total organic carbon (TOC) and total phosphorus (TP) exhibited an increase with depth. The ratio of Fe^{2+} to total Fe increased dramatically with depth, contrasting to the decrease in the ratio of SO_4^{2-} to total sulfur (TS). This indicated a shift from an oxidative environment at the surface tailings to a reductive condition at the deeper layers. For detailed physicochemical parameters of the tailings samples, see Additional file 2: Table S1 in the Supplemental material.

Diversity and distribution of viral and prokaryotic communities

Application of viral protein families-based pipeline [45], VirSorter [46] and VirFinder softwares [47] to predict viral sequences in the two cross-assembled metagenomic assemblies led to the identification of 2690 putative DNA metagenomic viral scaffolds. Reticulate classification of the viral scaffolds with classified isolate viruses was conducted by assessing shared gene contents; this allowed grouping viral genomes at approximately the genus level into VCs [49]. 211 VCs were identified across the 11 tailings samples, of which 155 did not contain any isolate viruses and 56 clusters were taxonomically affiliated with and corresponded to double-stranded DNA (dsDNA) and single-stranded DNA (ssDNA) viruses (Additional file 2: Table S2). The number of VCs in each tailings layer ranged from 64 to 205 and generally increased with depth (Additional file 1: Table S2). Examination of relative abundance of VCs in each tailings layer (calculated as the cumulative normalized coverage of its members divided by the total normalized coverage of viruses in that community) showed that the classified viruses accounted for 5.2%~49.7% of all viral communities, most of which were assigned as one of the three families (*Myoviridae*, *Siphoviridea*, and *Podoviridae*) in the *Caudovirales* order (Fig. 2a and Additional file 2: Table S2). Distribution of different viral groups in the 11 tailings samples showed a contrasting pattern: relative abundance of *Siphoviridae* and *Rudiviridae* decreased gradually from surface tailings to deeper layers, whereas relative abundance of *Myoviridae* and *Podoviridae* increased gradually from surface to deeper layers except in A6 where *Siphoviridae* increased significantly again and *Myoviridae* decreased (Fig. 2a).

The barcoded 16S rRNA gene sequencing generated 1,742,197 quality sequences from the 11 tailings samples, with a range of 43,424 to 134,565 sequences per community (Additional file 2: Table S3). A total of 3371 phylotypes were defined at a 97% sequence similarity cutoff; most (99%) of which could be assigned to a taxonomic group (phylum) by the RDP classifier (80% threshold). The prokaryotic phylotype richness generally increased with depth (ranging from 398 to 2321 in each sample), coincident with vertical distribution of viral diversity. Examination of relative abundance of the dominant lineages also showed contrasting patterns: while archaeal phylotypes were most abundant in the surface tailings layers, those of bacteria were most frequently detected in the deeper layers. Specifically, *Euryarchaeota* represented 67% and 80% of the total sequences of the surface tailings (A1 and B1, respectively), whereas *Proteobacteria*, *Nitrospirae* and *Firmicutes* collectively accounted for 77% and 90% of the total communities in the deeper layers (A6 and B5, respectively) (Fig. 2b).

Correlations between viral communities, prokaryotic communities, and geochemical data

Strong correlations were observed between viral communities, prokaryotic communities and geochemical data. Specifically, the two tailings cores exhibited similar increases in the number of vOTUs with increasing prokaryotic richness along the depth profiles as expected (Fig. 1 and Fig. 3a). Meanwhile, the number of vOTUs and the overall normalized coverage of viruses were also significantly correlated with measured geochemical parameters, for example, pH (Fig. 3b, c).

Euclidean distance based principal components analysis (PCA) and Bray-Curtis distance-based principle coordinate analysis (PCoA) were applied to further reveal the clustering patterns of physicochemical properties, and prokaryotic communities, and viral communities of the tailings, respectively (Fig. 3d-f). Results showed that physicochemical properties and prokaryotic and viral communities (OTU level) of samples from the vertical profiles of the tailings cores were apparently separated between surface and deeper layers, indicating a significant depth related variability in the biotic and abiotic signals and the potential correlations between them. In support of this, Mantel test analysis revealed that viral community dissimilarity (estimated between all pairwise combinations of samples) increased with an increasing difference in the prokaryotic community (Mantel's $r = 0.52$, $P < 0.001$) and geochemical characteristics (Mantel's $r = 0.31$, $P < 0.001$). Notably, viral communities were also apparently separated between the two tailings cores (Fig. 3f). Mantel test analysis revealed that viral community dissimilarity was most related to TOC (Mantel's $r = 0.47$, $P < 0.001$), while prokaryotic community dissimilarity was most significantly related to EC (Mantel's $r = 0.65$, $P = 0.002$) (Additional file 1: Figure S2). Thus, different variation patterns of TOC (increased with depth gradually in Core A but dramatically in Core B) may lead to distinct distribution patterns of viruses between the two cores.

Next, we performed extensive genome reconstruction for the bacteria and archaea present in the tailings cores to resolve putative hosts of the identified viruses. This resulted in a total of 435 draft prokaryotic genomes. These genomes were then screened for genomic features linking viruses to potential hosts. Protospacers were identified in 4 viral scaffolds, and 32 prophages were matched to their hosts (Additional file 2: Table S4). Together, putative hosts from 13 bacterial and archaeal phyla were predicted for 36 viral scaffolds, and a total of 36 virus-host pairs were identified (Additional file 2: Table S4). Notably, the relative abundance of prophages exhibited a depth related profiles in the two cores and correlated with pH significantly (Pearson's $r = -0.68$, $P = 0.022$) (Additional file 1: Figure S3).

Community-wide comparative gene profiles

To explore the metabolic capabilities and function diversity of viral communities associated with different depths, cluster of orthologous group (COG) annotation of viral genomes was performed by comparing the predicted viral proteins against the EggNOG database (5.0.0) [43], and the normalized coverage of each COG was calculated. Bray-Curtis distance-based PCoA again revealed strong primary clustering of viral COGs by depth (Fig. 4a). Further analysis indicated that 230 out of 2975 COGs displayed significantly ($P < 0.05$) different normalized coverage between the surface tailings and deeper layers (Additional file 2: Table S5). We defined a COG with a significantly higher or lower normalized

coverage in the surface tailings than that in the deeper layers viral communities as an indicator COG. Accordingly, 20 and 210 indicator COGs were identified for the surface communities and deeper layers communities, respectively (Additional file 2: Table S5). Interestingly, most of the indicator COGs in the surface tailings were assigned as archaeal clusters of orthologous genes (arCOGs) that are mainly involved in energy production and conversion (COG C), cell wall/membrane/envelope biogenesis (COG M), amino acid transport and metabolism (COG E), carbohydrate transport and metabolism (COG G) and inorganic ion transport and metabolism (COG P). This result was consistent with the predominance of archaeal phylotypes in the corresponding surface prokaryotic communities. In contrast, the deeper layers viral communities harbored a large proportion of higher indicator COGs related to coenzyme transport and metabolism (COG H), signal transduction mechanisms (COG T), replication, recombination and repair (COG L) and transcription (COG K) (Fig. 4b and Additional file 2: Table S5). To further illustrate potential links between viral functions and compositions, we analyzed the relative abundance and composition of viral genomes that encoded the indicator COGs. These viruses accounted for a significant proportion of the total viral communities in the surface and deeper layers (23% and 17% in A1 and A4, respectively) (Fig. 5). Taxonomic classification of these viral genomes further revealed that viruses encoding the indicator COGs in the surface layers were mainly classified as *Rudiviridae*, while families of *Caudovirales* order and *Ascoviridae* primarily encoded the indicator COGs in the deeper layers (Fig. 5).

Case study of AMGs

Having illustrated the community-wide functional profiles, we next sought to identify the virus-encoded AMGs that could modify host metabolism during infection. Given the observed lower ratio of SO_4^{2-}/TS (Fig. 1) and higher abundance of sulfate-reducing bacteria such as *Firmicutes* and *Proteobacteria* (Fig. 2b) in the deeper tailings layers, genes related to sulfate reduction were selected for subsequent analysis. Intriguingly, we found 9 viral scaffolds harbored genes participating in sulfate reduction (COG0175) (Fig. 6a and Additional file 2: Table S6), which are important for the conversion of sulfate to sulfide [62]. To further confirm the origin of these viral genes, 136 homologs from 11 prokaryotic phyla were recruited and combined to build a phylogenentic tree (Fig. 6b), and their putative hosts were predicted as nearest-neighbors. The phylogenetic analysis showed that the sulfate reduction genes in the viral genomes ‘CoreA_NODE_22178’ and ‘CoreA_NODE_4680’ clustered with their counterparts from *Firmicutes*, indicating that these AMGs might be acquired from this widely distributed bacterial lineage. This result was in agreement with our prediction of *Firmicutes* as the putative host of viral genome ‘CoreA_NODE_22178’ (Additional file 2: Table S4). However, the hosts of other sulfate reduction genes were uncertain, as they clustered with sulfate reduction genes from different phyla (Fig. 5b). Nonetheless, reads mapping to the 9 viral genomes and the 9 sulfate reduction genes showed that either the genomes or the genes were enriched in the deeper layers of the tailings cores (Fig. 6c), implying the potential impact of viral sulfate reduction on the sulfur cycles *in situ*.

Discussion

The depth-stratified physicochemical and biological profile in the Fankou Pb/Zn sulphidic mine tailings site has enabled an in-depth exploration of the variation of viral communities in the context of geochemical changes. While many viral ecology studies have employed size-based enrichment of viral particles to generate the metagenomes (viral metagenomes or viromes) [3, 4, 5], we performed metagenomic sequencing on total genomic DNA extracted directly from the mine tailings. This would allow recovery of sequences from not only temperate viruses that are either integrated into host genomes or present as episomal elements in the host cell, but also free virus particles present in the original samples. While metagenomics has brought new opportunities to the rapidly progressing field of viral ecology, identification of putative viral sequences in the sequence datasets remains a major challenge. Previous studies have employed both viral protein families and VirSorter software [1, 48]. However, benchmarking of the two computational approaches demonstrated that the viral protein family-based pipeline had a better precision (although the recall rate was higher with VirSorter) in a synthetic metagenome and they rarely behaved in a similar way to metagenomes from natural communities [1]. An additional software VirFinder had the advantage of identifying novel and relatively shorter viral sequences [47]. Recent studies have successfully applied both VirSorter and VirFinder to identify new viruses in metagenomes [67, 68], indicating combined benefits of using multiple tools. Thus, we employed separately these methods in our study and merged the identified viral scaffolds data, uncovering a large proportion of unclassified viral genomes in the Fankou mine tailings (Fig. 2a). Unknown virospheres have recently been discovered in many other habitats such as marine environments, acidic hot springs and permafrost soils [4, 8, 9]. Given that the reticulate classification method of viral sequences uses shared gene-content information [2, 49], and that currently the isolated archaeal viruses are largely outnumbered by bacteriophages [69, 70], it is likely that archaeal viruses may account for a substantial fraction of the unclassified viral scaffolds in our study, especially in the archaea-predominating surface tailings. It is noteworthy that the family *Ascoviridae*, which mainly infects lepidopteran larvae was identified in our datasets (Fig. 2a) [71]. This is likely because the mine tailings impoundment is an open natural environment surrounded by mountains and thus it is possible that the analyzed tailings samples may contain some insect remains. Biases associated with the bioinformatic procedures can not be ruled out, however. Specifically, small fragments of larger genomes often have low statistical power in the automated viral taxonomy tool (vConTACT v.2.0) [52], and this may lead to misclassification of novel viral sequences.

Samples from both cores share a common depth-stratified pattern in the overall composition of geochemistry, prokaryotic communities, and viral communities (Fig. 3d-f). While it is unclear whether variations in viral communities were directly driven by their hosts or by geochemical changes along the tailings profiles, our results provided quantitative evidence that viral diversity increases with depth in the highly stratified mine tailings at this site (Fig. 1). Noteworthy, previous 16S rRNA gene surveys have identified pH as major driver of prokaryotic community composition at local- or large-scales in the extreme AMD and associated environments [72, 73]. Our current metagenomics analysis demonstrated that pH is also one of the major factors shaping the relatively under-studied viral world (Fig. 3b and c). That viral richness and normalized coverage increased with increasing pH along the depth profiles is

somewhat expected because both viruses and their prokaryotic hosts tend to be sensitive to acidic pH [21, 22]. This would also explain the observed less variability of both prokaryotic and viral populations at lower pH values (Fig. 3a and b).

It is widely accepted that viruses depend on their prokaryotic hosts to successfully replicate. We hypothesised that viruses tend to be more temperate and symbiotic with hosts in extreme conditions, and this was supported by the significant negative correlations between the relative abundance of prophages and pH (Additional file 1: Figure S3). Our results are consistent with previous studies which suggested that the lysogenic state should be favored under extreme conditions (for example, low nutrients, low productivity or heat) [74]. This is a readily comprehensible pattern as lysogeny can enhance phage and host survival, particularly under adverse conditions [75]. Thus, the enrichment of prophages in the surface layers of the mine tailings might enable not only the detection of virus-host links, but also allow a glimpse of viral lysogeny decisions.

Viral communities with diverse taxa in natural environments may exhibit distinct functional profiles in response to the varying biotic and abiotic factors [14, 76]. Comparative analysis of viral community gene profiles showed that metabolic patterns were significantly different between surface tailings and deeper layers (Fig. 4a) and, although found in all tailings layers, many indicator COGs had distinct, depth-related distribution (Additional file 2: Table S5). The microorganisms populating surface tailings face multiple environmental stresses including extreme acidity, oligotrophy, and high EC. The viral populations associated with these prokaryotic communities may help their hosts cope with these stresses via AMGs. A careful check on the functional profiles of viral communities provided insights into these potential interactions. For example, microbial homeostatic mechanisms for low pH include a highly impermeable cell membranes and active secondary membrane transporter proteins [77]. Striking functional parallels were observed in the surface tailings viral communities. First, biosynthesis of archaeal membrane lipids may be augmented by arCOG00570 and KOG3638. Second, the influx of protons may be augmented by the major facilitator superfamily (MFS) arCOG00130 which represents the largest group of secondary active membrane transporters [78]. Additionally, given the limited TOC and high EC levels in the surface tailings, the MFS-directed augmentation of carbohydrate and inorganic ion transport and metabolism may be also critical for the survival of prokaryotic hosts in this stressful environments. Intriguingly, viruses encoding the indicator COGs in the surface layers were mainly assigned as *Rudiviridae* (Fig. 5), which are known to infect hyperhermophilic members of archaea [79]. As *Rudiviridae* were found increasing in the surface tailings (Fig. 2a), it is reasonable to speculate that the indicator COGs encoded by this viral group may contribute to the predominance of archaea in these samples. Compared with the more readily identifiable microbial adaptations in the surface extreme environments, the category of functions in the deeper layers showed strong consistency with conventional viral functions, such as phage integrase (COG4974), phage tail tap measure protein (COG5412), histidine kinase (COG4191) and transposition (COG2801) (Fig. 4b), suggesting that viruses in the less extreme deeper tailings layers are more similar to currently known viruses, and thus more easily annotated. This finding is in agreement with the fact that known viruses are largely isolated from non-extreme environments.

The role of viruses in regulating the sulfur cycle was previously described in deep ocean viral communities [4, 80]. Interestingly, our analyses showed that viral genes (COG0175) participating in sulfate reduction process were abundant in the deeper tailings layers (Fig. 6a), which were characterized by lower ratio of SO_4^{2-}/TS. While AMD typically contains elevated levels of sulfate due to oxidative dissolution of sulfide minerals, sulfate can be reduced by acidophilic sulfate reducers in anoxic microenvironments (e.g., the underlying sediments) [81, 82], leading to the formation of sulfide and sulfide-mediated metal precipitation, an important mechanism for the bioremediation of AMD environments. Meanwhile, sulfur plays a central role in many essential biomolecules like iron-sulfur (Fe-S) clusters, sulfur-containing amino acids and cofactors and thus is a essential element for all life [83]. However, reduction of sulfate to sulfide is needed prior to microbial incorporation. Thus, our results indicate a potential contribution of viruses to this important step of the sulfur cycle in the deeper part of the tailings environment.

Phylogenetic analysis showed that two of the viral sulfate reduction genes originated from *Firmicutes*, corroborating with the result that one of the viral genomes carriying the sulfate reduction genes was predicted to infect *Firmicutes* (Fig. 6b and Additional file 2: Table S4). Sulfate reduction was widespread in *Firmicutes* and *Proteobacteria*, as exemplified by phylogenetic analysis of homologous genes within COG0175 (Fig. 6b). Sulfate reduction bacteria (SRB) are a phylogenetically diverse group of anaerobes that use sulfate as a terminal electron acceptor during degradation of simple organic matters, and thus play important roles in both the sulfur and carbon cycles [84, 85]. Our geochemical data showed that TOC increased while the ratio of SO_4^{2-}/TS decreased in the deeper tailings layers (Fig. 1), indicating a possible enrichment SRB, which is further evidenced by the coincident higher relative abundance of *Firmicutes* and *Proteobacteria* in these samples (Fig. 2b). The abundant viral sulfate reduction AMGs in the deeper layers possibly may facilitate these SRB to utilize sulfate in energy conservation and thus proliferate in the oxygen-depleting environment, which in turn benefit the replication and reproduction of associated viruses (Fig. 6c).

Conclusions

Although the field of viral ecology is rapidly evolving owing to recent developments of sequencing and bioinformatics methods, the viral communities populating various extreme environments remain relatively underexplored. Our comprehensive analysis of the mine tailings cores has revealed a largely novel, depth-stratified viral community that shows strong correlations with co-occurring prokaryotic assemblages and geochemical gradients. The environmental conditions associated with different oxidations stages of mine tailings (deep layers of the cores represent unaltered, pH-neutral tailings material whereas top layers represent highly oxidized and acidified tailings) apparently have a profound impact on the viral populations and their functions. Future simulated experiments of oxidative dissolution of sulphidic mine tailings or sulfide minerals, coupled with extensive time-series sampling and analysis, will provide more detailed insights into viral dynamics and their interplay with prokaryotic populations and geochemical conditions during the process of acid generation.

Abbreviations

Electronic conductivity (EC) Total phosphorus (TP)

Total sulfur (TS) **TOC:** Total organic carbon

VC: Viral cluster **vOTU:** Viral operational taxonomic unit

AMG: Auxiliary metabolic gene **COG**: Clusters of orthologous group

NCBI: National Center for Biotechnology Information

EggNOG: Evolutionary genealogy of genes: Non-supervised Orthologous Groups

PCA: Principal components analysis **PCoA:** Principle coordinate analysis

PCR: Polymerase chain reaction **SRB:** Sulfate reduction bacteria

Declarations

Acknowledgements

Not applicable

Funding

This work was supported by the National Natural Science Foundation of China (No. 31570500, 31870111 and 40930212).

Availability of data and materials

Raw reads of prokaryotic 16S rRNA gene amplicons and metagenomics are available for download from the Short Reads Archive with NCBI BioProject accession no. PRJNA515819.

Authors' contributions

SMG, NC and MMZ conducted the experiments and collected the data. SMG, QL and YY analysed the data. LNH, SMG, and YY wrote the initial draft of the manuscript while AS, BL and WSS provided substantial feedback.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare no conflict of interest.

References

1. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC. Uncovering Earth's virome. Nature. 2016;536(7617):425.
2. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. Reticulate representation of evolutionary and functional relationships between phage genomes. Mol Biol Evol. 2008;25(4):762-77.
3. Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, Brussaard CPD, Dutilh BE, Thompson FL. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. Nat Commun. 2017;8:15955.
4. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature. 2016;537(7622):689.
5. Arkhipova K, Skvortsov T, Quinn JP, McGrath JW, Allen CC, Dutilh BE, McElarney Y, Kulakov LA. Temporal dynamics of uncultured viruses: a new dimension in viral diversity. ISME J. 2018;12(1):199-211.
6. Adriaenssens EM, Kramer R, Van Goethem MW, Makhalanyane TP, Hogg I, Cowan DA. Environmental drivers of viral community composition in Antarctic soils identified by viromics. Microbiome. 2017;5(1):83.
7. Yu DT, Han LL, Zhang LM, He JZ. Diversity and distribution characteristics of viruses in soils of a marine-terrestrial ecotone in east China. Microb Ecol. 2018;75(2):375-386.
8. Bolduc B, Wirth JF, Mazurie A, Young MJ. Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis. ISME J. 2015;9(10):2162-77.
9. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, Singleton CM, Solden LM, Naas AE, Boyd JA, Hodgkins SB, Wilson RM, Trubl G, Li C, Frolking S, Pope PB, Wrighton KC, Crill PM, Chanton JP, Saleska SR, Tyson GW, Rich VI, Sullivan MB. Host-linked soil viral ecology along a permafrost thaw gradient. Nat Microbiol. 2018;3(8):870-880.
10. Daly RA, Roux S, Borton MA, Morgan DM, Johnston MD, Booker AE, Hoyt DW, Meulia T, Wolfe RA, Hanson AJ, Mouser PJ, Moore JD, Wunch K, Sullivan MB, Wrighton KC, Wilkins MJ. Viruses control dominant bacteria colonizing the terrestrial deep biosphere after hydraulic fracturing. Nat Microbiol. 2019;4(2):352-361.
11. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, Darzi Y, Audic S, Berline L, Brum J, Coelho LP, Espinoza JCI, Malviya S, Sunagawa S, Dimier C, Kandels-Lewis S, Picheral M, Poulain J, Searson S; Tara Oceans coordinators, Stemmann L, Not F, Hingamp P, Speich S, Follows M, Karp-

Boss L, Boss E, Ogata H, Pesant S, Weissenbach J, Wincker P, Acinas SG, Bork P, de Vargas C, Iudicone D, Sullivan MB, Raes J, Karsenti E, Bowler C, Gorsky G. Plankton networks driving carbon export in the oligotrophic ocean. Nature. 2016;532(7600):465-470.

12. Forterre P. The virocell concept and environmental microbiology. ISME J. 2013;7(2):233-6.
13. Breitbart M. Marine viruses: truth or dare. Annu Rev Mar Sci, 2012;4:425-448.
14. Hurwitz BL, Brum JR, Sullivan MB. Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean virome. ISME J. 2015;9(2):472.
15. Brockhurst MA, Koskella B. Experimental coevolution of species interactions. Trends Ecol Evol. 2013;28(6):367-375.
16. Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, De VC, Gasol JM. Ocean plankton. Patterns and ecological drivers of ocean viral communities. Science. 2015;348(6237):1261498.
17. Singer PC, Stumm W. Acidic mine drainage: the rate-determining step. Science. 1970;167(3921):1121-3.
18. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature. 2004;428(6978):37-43.
19. Denef VJ, Mueller RS, Banfield JF. AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. ISME J. 2010;4(5):599-610.
20. Huang LN, Kuang JL, Shu WS. Microbial ecology and evolution in the acid mine drainage model system. Trends Microbiol. 2016;24(7):581-593.
21. Kyle JE, Pedersen K, Ferris FG. Virus mineralization at low pH in the Rio Tinto, Spain. Geomicrobiology Journal. 2008;25(7-8):338-345.
22. Kyle JE, Ferris FG. Geochemistry of virus–prokaryote interactions in freshwater and acid mine drainage environments, Ontario, Canada. Geomicrobiology Journal. 2013;30(9):769-778.
23. Andersson AF, Banfield JF. Virus population dynamics and acquired virus resistance in natural microbial communities. Science. 2008;320(5879):1047-50.
24. Comolli LR, Banfield JF. Inter-species interconnections in acid mine drainage microbial communities. Front Microbiol. 2014;5:367.
25. Chen LX, Li JT, Chen YT, Huang LN, Hua ZS, Hu M, Shu WS. Shifts in microbial community composition and function in the acidification of a lead/zinc mine tailings. Environ Microbiol. 2013;15(9):2431-44.
26. Huang LN, Zhou WH, Hallberg KB, Wan CY, Li J, Shu WS. Spatial and temporal analysis of the microbial community in the tailings of a Pb-Zn mine generating acidic drainage. Appl Environ Microbiol. 2011;77(15):5540-4.
27. Hill AG, Bishop E, Coles LE, McLaughlan EJ, Meddle DW, Pater MJ, Watson CA, Whalley C. Standardized general method for the determination of iron with 1,10-phenanthroline. Analyst.

1978;103:391-396.
28. Chesmin L, Yien CH. Turbidimetric determination of available sulphate. Soil Sci Soc Am Proc. 1951;15:149-151.
29. Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R, Fierer N. Examining the global distribution of dominant archaeal populations in soil. ISME J. 2011;5(5):908-17.
30. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods. 2008;5(3):235-7.
31. Fierer N, Hamady M, Lauber CL, Knight R. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. Proc Natl Acad Sci U S A. 2008;105(46):17994-9
32. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537-41.
33. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-6.
34. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27(16):2194-200.
35. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73(16):5261-7.
36. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol. 2014;10(4):e1003531.
37. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE. Human genetics shape the gut microbiome. Cell. 2014 Nov 6;159(4):789-99.
38. https://github.com/eco-gaoshaom/in-house-scripts.
39. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77.
40. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
41. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. The Pfam protein families database in 2019. Nucleic Acids Res. 2019;47(D1):D427-D432.

42. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44(D1):D457-62.
43. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309-D314.
44. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9.
45. Paez-Espino D, Pavlopoulos GA, Ivanova NN, Kyrpides NC. Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data. Nat Protoc. 2017;12(8):1673-1682.
46. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. PeerJ. 2015;3:e985.
47. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome. 2017;5(1):69.
48. Roux S, Hallam SJ, Woyke T, Sullivan MB. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. Elife. 2015;4.
49. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR, Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonté JM, Lee KB, Malmstrom RR, Martinez-Garcia M, Mizrachi IK, Ogata H, Páez-Espino D, Petit MA, Putonti C, Rattei T, Reyes A, Rodriguez-Valera F, Rosario K, Schriml L, Schulz F, Steward GF, Sullivan MB, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Thurber RV, Webster NS, Whiteson KL, Wilhelm SW, Wommack KE, Woyke T, Wrighton KC, Yilmaz P, Yoshida T, Young MJ, Yutin N, Allen LZ, Kyrpides NC, Eloe-Fadrosh EA. Minimum information about an uncultivated virus genome (MIUViG). Nat Biotechnol. 2019;37(1):29-37.
50. https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239, Accessed 27 Dec 2017.
51. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002;30(7):1575-84.
52. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol. 2019;37(6):632-639.
53. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;3:e1165.
54. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016; 32(4):605-7.
55. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11(11):1144-6.

56. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol. 2018;3(7):836-843.
57. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017;2(11):1533-1542.
58. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043-55.
59. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018 Nov;36(10):996-1004.
60. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. Computational approaches to predict bacteriophage-host relationships. FEMS Microbiol Rev. 2016;40(2):258-72.
61. Rho M, Wu YW, Tang H, Doak TG, Ye Y. Diverse CRISPRs evolving in human microbiomes. PLoS Genet. 2012;8(6):e1002441.
62. Valdés J, Veloso F, Jedlicki E, Holmes D. Metabolic reconstruction of sulfur assimilation in the extremophile Acidithiobacillus ferrooxidans based on genome analysis. BMC Genomics. 2003;4(1):51.
63. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113.
64. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972-3.
65. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3.
66. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47(W1):W256-W259.
67. Ahlgren NA, Fuchsman CA, Rocap G, Fuhrman JA. Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode amoC nitrification genes. ISME J. 2019;13(3):618-631.
68. Gregory AC, Zablocki O, Howell A, Bolduc B, Sullivan MB. The human gut virome database. BioRxiv. 2019:655910.
69. Pietilä MK, Demina TA, Atanasova NS, Oksanen HM, Bamford DH. Archaeal viruses and bacteriophages: comparisons and contrasts. Trends Microbiol. 2014;22(6):334-44.
70. Snyder JC, Bolduc B, Young MJ. 40 Years of archaeal virology: Expanding viral diversity. Virology. 2015;479-480:369-78.

71. Asgari S, Bideshi DK, Bigot Y, Federici BA, Cheng XW, ICTV report consortium. ICTV virus taxonomy profile: Ascoviridae. J Gen Virol. 2017;98(1):4-5.
72. Kuang JL, Huang LN, Chen LX, Hua ZS, Li SJ, Hu M, Li JT, Shu WS. Contemporary environmental variation determines microbial diversity patterns in acid minedrainage. ISME J. 2013;7(5):1038-50.
73. Liu J, Hua ZS, Chen LX, Kuang JL, Li SJ, Shu WS, Huang LN. Correlating microbial diversity patterns with geochemistry in an extreme and heterogeneous environment of mine tailings. Appl Environ Microbiol. 2014;80(12):3677-86.
74. Stewart FM, Levin BR. The population biology of bacterial viruses: why be temperate. Theor Popul Biol. 1984;26(1):93-117.
75. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. ISME J. 2017;11(7):1511-1520.
76. Hurwitz BL, Westveld AH, Brum JR, Sullivan MB. Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. Proc Natl Acad Sci U S A. 2014;111(29):10714-9.
77. Baker-Austin C, Dopson M. Life in acid: pH homeostasis in acidophiles. Trends Microbiol. 2007;15(4):165-71.
78. Yan N. Structural biology of the major macilitator superfamily transporters. Annu Rev Biophys. 2015;44:257-83.
79. Prangishvili D, Krupovic M. A new proposed taxon for double-stranded DNA viruses, the order "Ligamenvirales". Arch Virol. 2012;157(4):791-5.
80. Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ. Sulfur oxidation genes in diverse deep-sea viruses. Science. 2014;344(6185):757-60.
81. Sánchez-Andrea I, Rodríguez N, Amils R, Sanz JL. Microbial diversity in anaerobic sediments at Rio Tinto, a naturally acidic environment with a high heavy metal content. Appl Environ Microbiol. 2011;77(17):6085-93.
82. Sánchez-Andrea I, Knittel K, Amann R, Amils R, Sanz JL. Quantification of Tinto River sediment microbial communities: importance of sulfate-reducing bacteria and their role in attenuating acid mine drainage. Appl Environ Microbiol. 2012 Jul;78(13):4638-45.
83. Rückert C. Sulfate reduction in microorganisms-recent advances and biotechnological applications. Curr Opin Microbiol. 2016;33:140-146.
84. Muyzer G, Stams AJ. The ecology and biotechnology of sulphate-reducing bacteria. Nat Rev Microbiol. 2008;6(6):441-54.
85. Zhou J, He Q, Hemme CL, Mukhopadhyay A, Hillesland K, Zhou A, He Z, Van Nostrand JD, Hazen TC, Stahl DA, Wall JD, Arkin AP. How sulphate-reducing microorganisms cope with stress: lessons from systems biology. Nat Rev Microbiol. 2011;9(6):452-66.

Figures

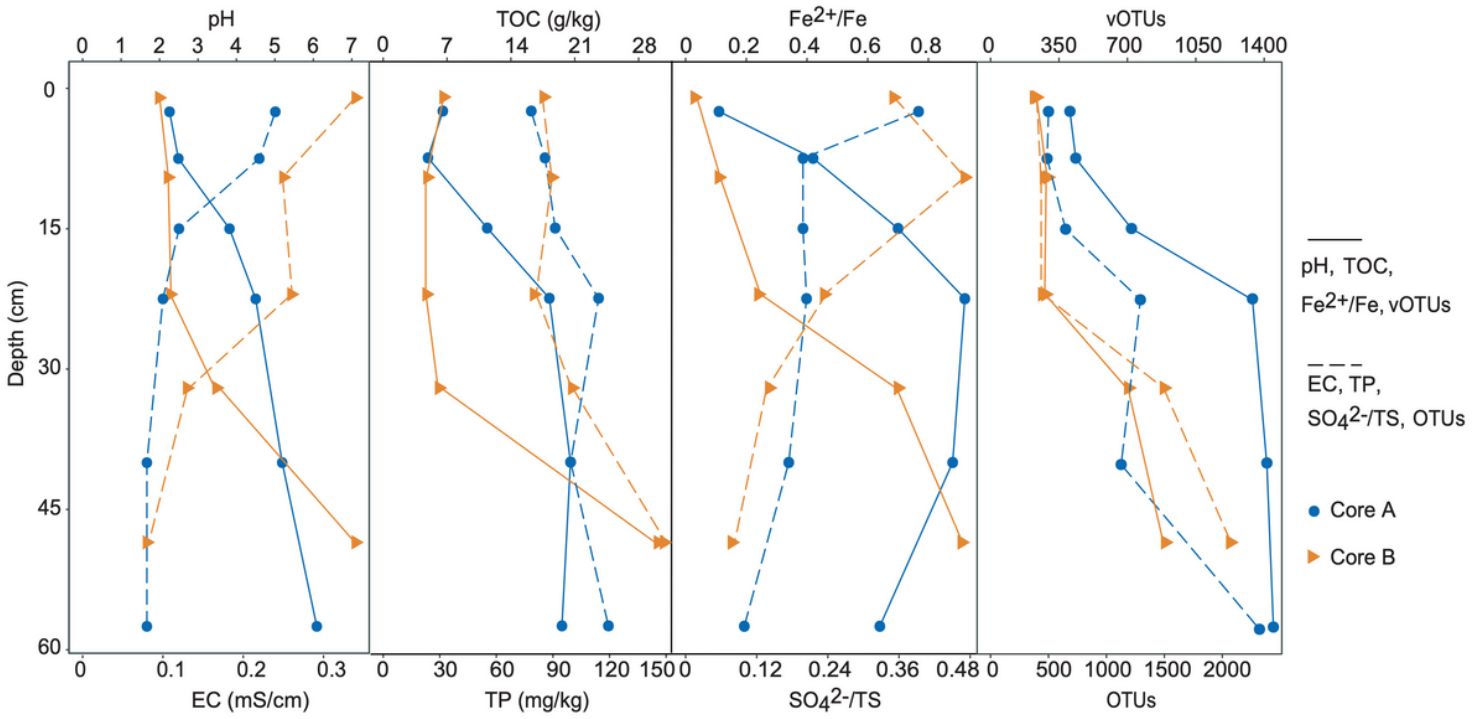

Figure 1

Vertical profiles of physicochemical and biodiversity data for the two tailing cores from the Fankou Pb/Zn Mine located in Guangdong Province, China. The intermediate depth of each layer is taken as the depth of each sample. EC, electronic conductivity; TOC, total organic carbon; TP, total phosphorus; TS, total sulfur; vOTUs, the number of viral OTUs; OTUs, the number of prokaryotic OTUs.

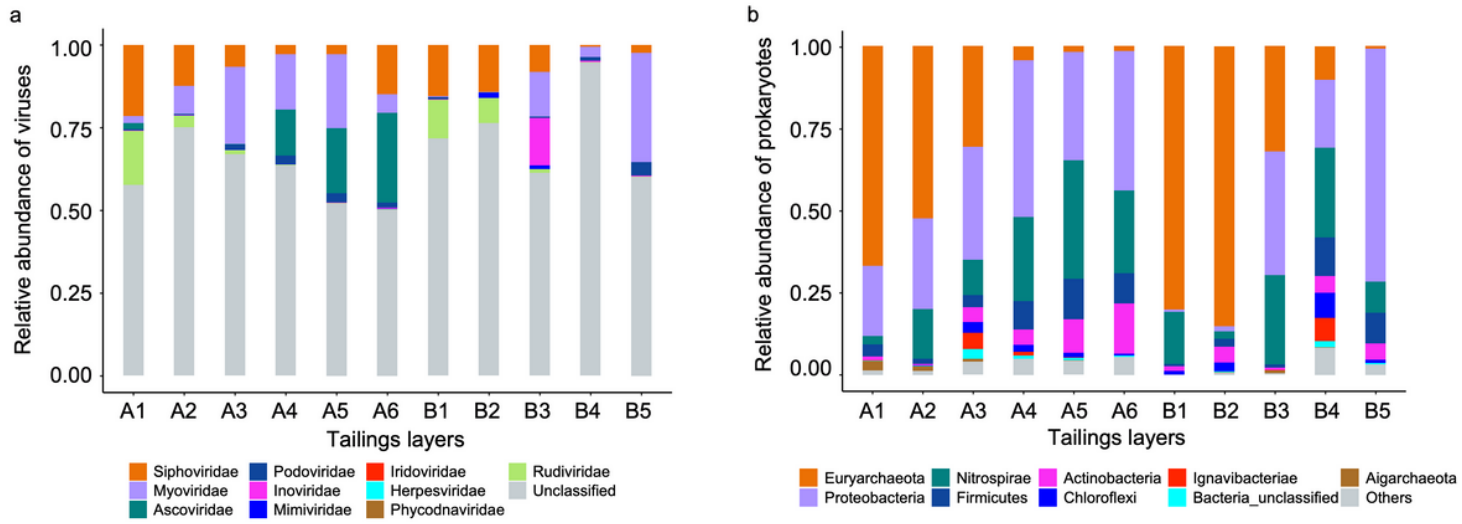

Figure 2

Relative abundance of (a) viruses (family level) and (b) prokaryotes (phylum level) in the 11 depth-stratified mine tailings layers revealed by metagenomics and 16S rRNA gene amplicon sequencing, respectively.

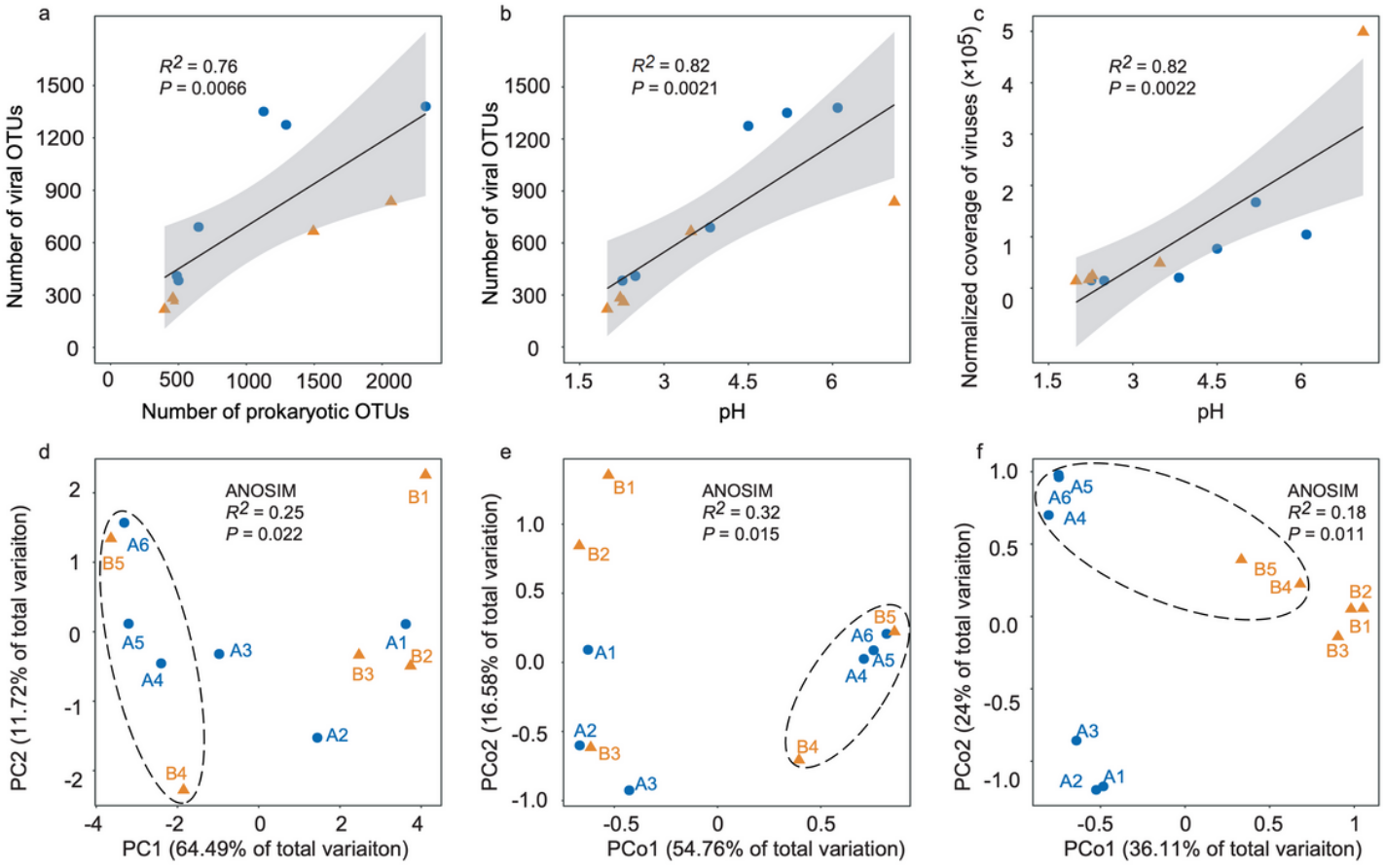

Figure 3

(a-c) Significant Pearson correlations between prokaryotic community, viral community, and pH. (d-f) Principal components analysis (PCA) of geochemical data as derived from Euclidean dissimilarities, and Principle coordinate analysis (PCoA) of prokaryotic community and viral community as derived from Bray-Curties dissimilarities. The analysis of similarity (ANOSIM) statistics consider samples grouped by depth (inside and outside the dashed ellipses).

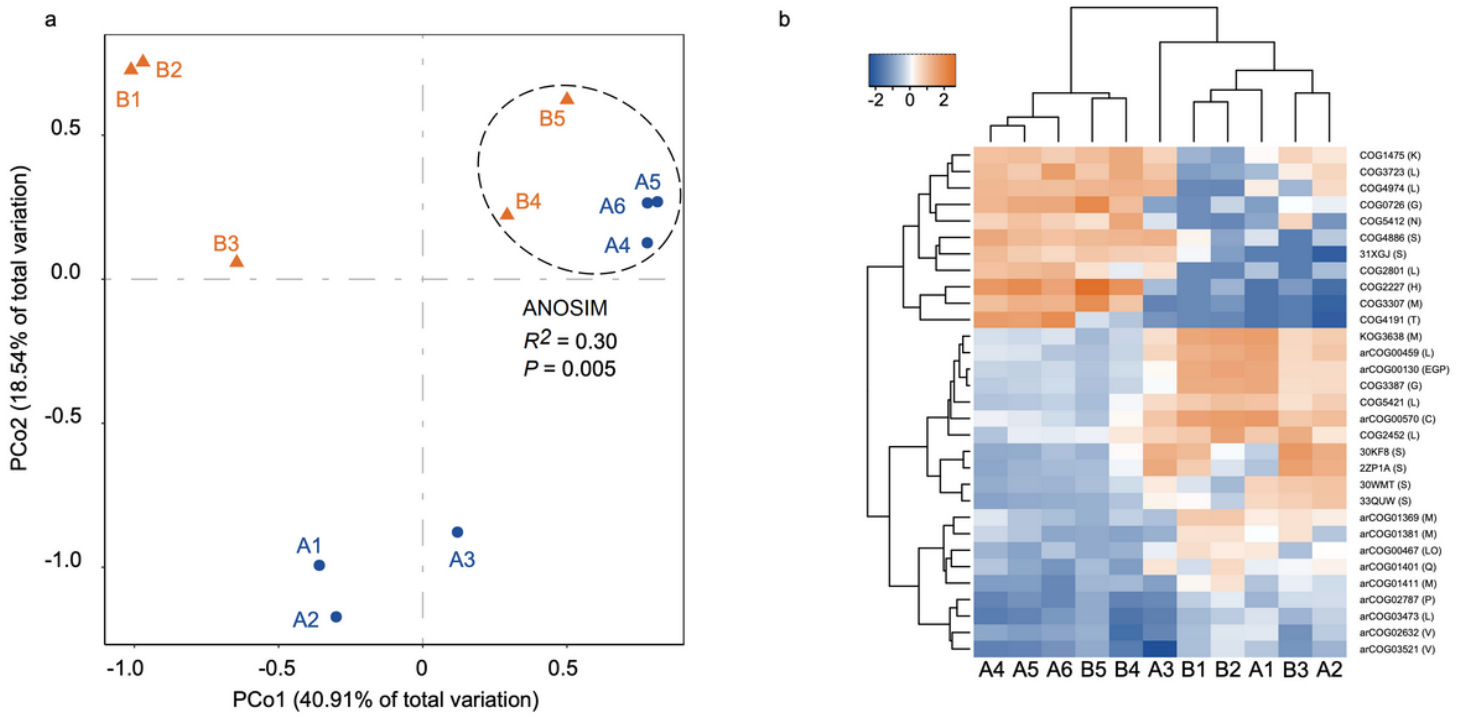

Figure 4

Overview of functional profiles of viral community. a Principle coordinate analysis of viral COGs based on their relative abundance in each community. The analysis of similarity (ANOSIM) statistics consider samples grouped by depth (inside and outside the dashed ellipse). b Hierarchical clustering of the indicator COGs (all indicator COGs in the surface layers and indicator COGs with a relative abundance > 1% in the deeper layers). The relative abundance were log transferred and normalized with a z-score method. COG categories: S: Function unknown; E: amino-acid transport and metabolism; G: carbohydrate transport and metabolism; P: inorganic ion transport and metabolism; L: replication, recombination and repair; O: posttranslational modification, protein turnover, chaperones; C: energy production and conversion; M: cell wall/membrane/envelope biogenesis; Q: secondary metabolites biosynthesis, transport and catabolism; V: defense mechanisms; P: inorganic ion transport and metabolism; K: transcription; G: carbohydrate transport and metabolism; F: nucleotide transport and metabolism; H: coenzyme transport and metabolism; signal transduction mechanisms; N: cell motility.

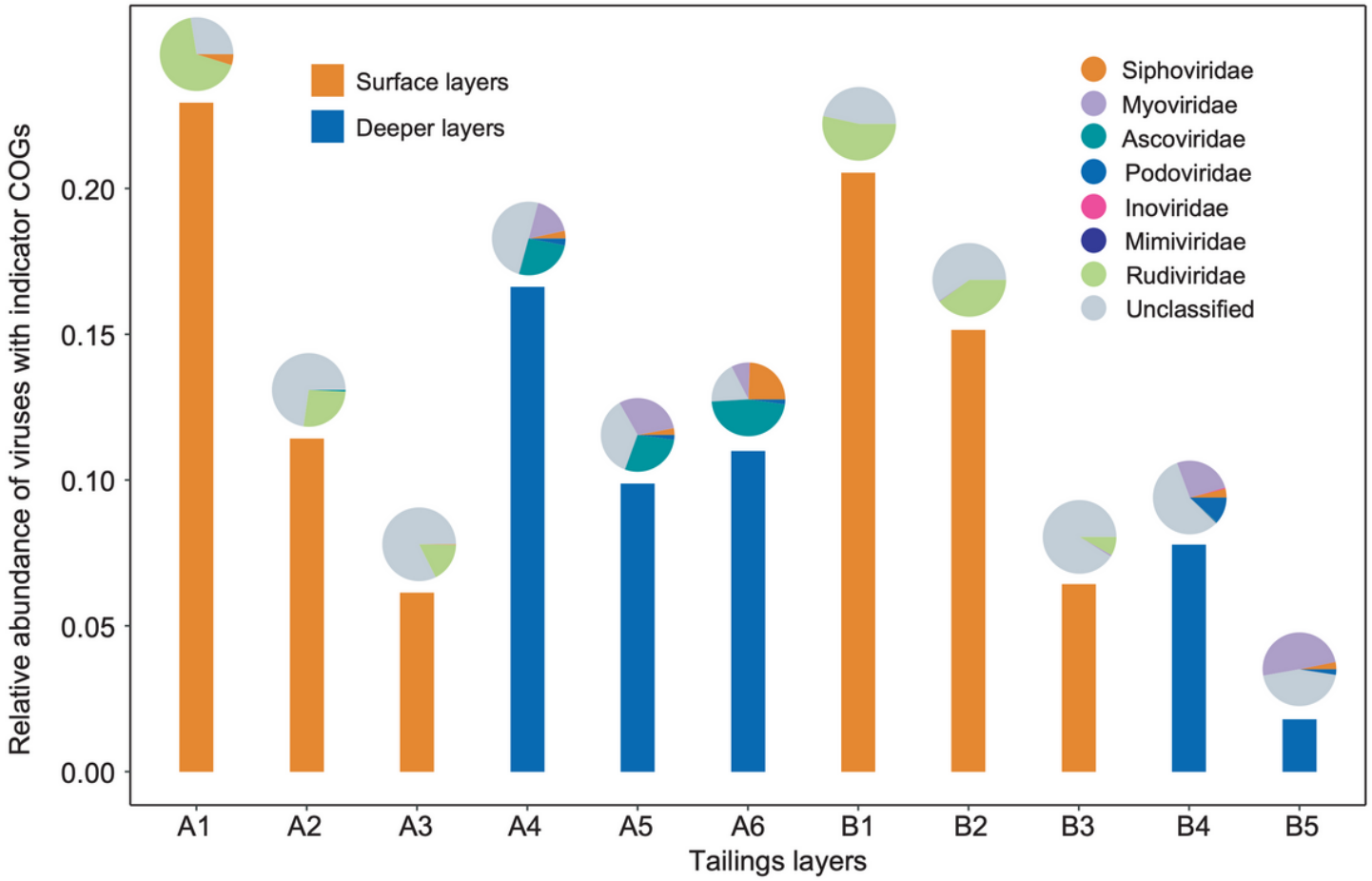

Figure 5

Bar graphs showing the relative abundance of viruses encoding the indicator COGs in surface tailings (orange) and deeper layers (blue) and pie charts showing percent composition of viruses that encode the indicator COGs in each layers.

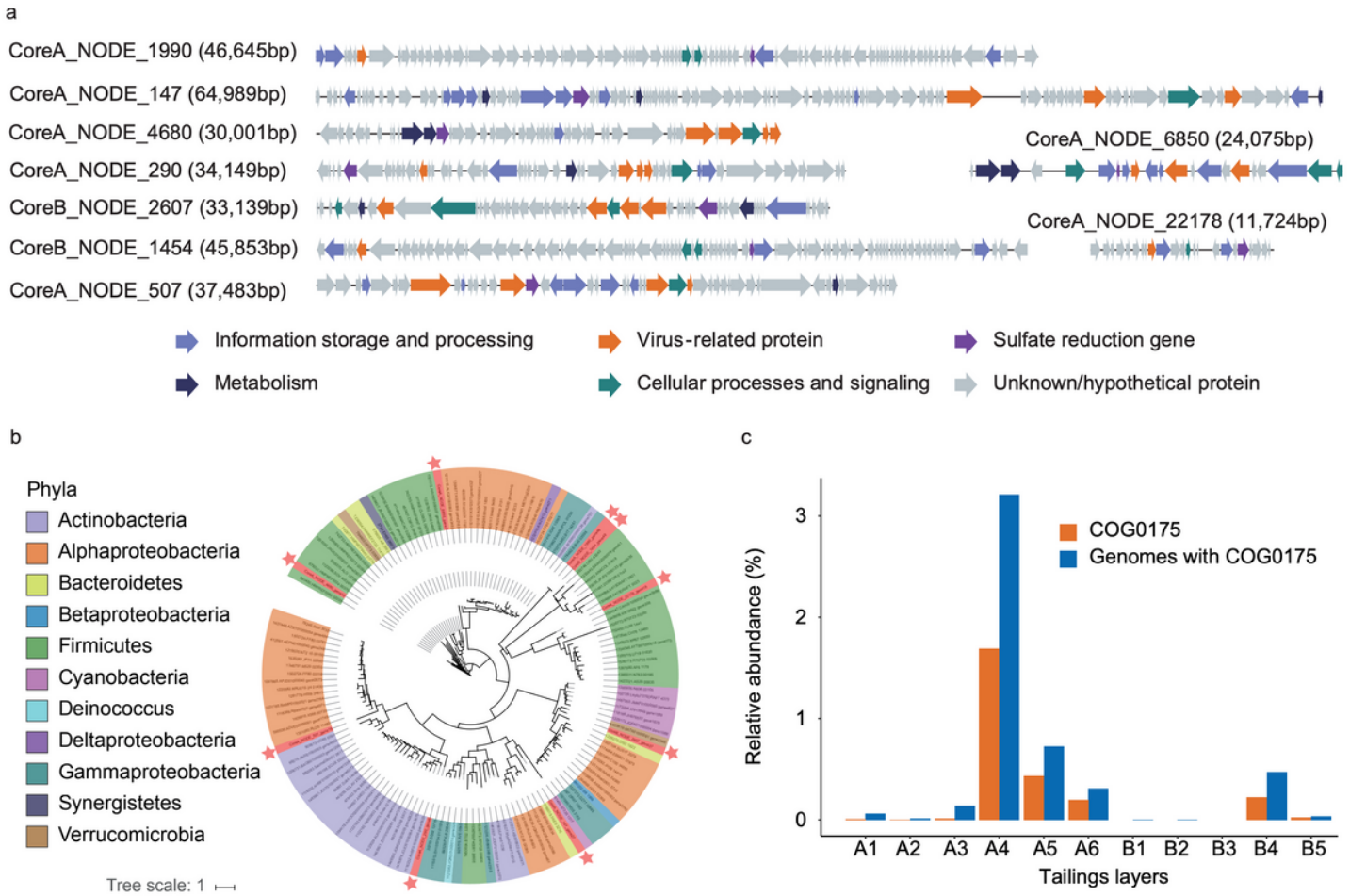

Figure 6

Genomic analysis of viral sulfate reduction genes. a Genome map of 9 viral scaffolds containing sulfate reduction genes. Genes related to information storage and processing are shown in light purple, genes related to metabolism are shown in dark blue, genes related to cellular processes and signaling are shown in green, viral hallmark genes are in orange, sulfur reduction genes (COG0175) are in dark purple and unknown or hypothetical genes are in grey. b Maximum-likelihood phylogenetic tree with sulfate reduction genes from mine tailings viral genomes (indicated by stars) compared to genes found in bacterial reference sequences (Material and methods). The scale bar represents 1 amino acid substitution per site. c Total relative abundance of COG0175 and genomes containing COG0175 in each tailings layer viral genomes.

Supplementary Files

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

- SupplementaryFiles.zip