Han Li-Li (Orcid ID: 0000-0002-8105-1672)

Title: Diversity and potential biogeochemical impacts of viruses in bulk and rhizosphere soils

Running title: Virome analysis of agricultural soils

Li Bi[a,b], Dan-Ting Yu[c], Shuai Du[a,b], Li-Mei Zhang[a,b], Li-Yu Zhang[d], Chuan-Fa Wu[e], Chao Xiong[a,b], Li-Li Han[a,b*], Ji-Zheng He[c,f]

[a]State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

[b]University of Chinese Academy of Sciences, Beijing 100049, China

[c]Institute of Geography, Fujian Normal University, Fuzhou 350007, China

[d]School of Environment and Energy, Shenzhen Graduate School, Peking University, Shenzhen 518055, China

[e]Key Laboratory of Agro-ecological Processes in Subtropical Region, Institute of Subtropical Agriculture, Chinese Academy of Sciences, Hunan 410125, China

[f]Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia

***Corresponding author**

Address: Research Centre for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, 100085, China

Telephone: +86-10-6284 9500; Fax: +86-10-62953251

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.15010

E-mail address: llhan@rcees.ac.cn

Originality-Significance Statement

Viruses, as a major food web component, play an important role in the marine environment. However there is little information about viruses in agricultural soil, and in particular, the diversity and potential roles of viruses in the rhizosphere remain poorly explored. In this study, we studied viral abundance, diversity and potential biogeochemical impacts in both bulk and rhizosphere soils. Our results highlight that viruses had high diversity in agricultural soil and viral communities were affected by plant roots. Viruses can contribute to biogeochemical cycling by reprogramming bacterial metabolism via putative auxiliary metabolic genes (AMGs) or virus-encoded metabolic genes, and this would help us understand biogeochemical element cycling in an agro-ecosystem.

Summary

Viruses can affect microbial dynamics, metabolism and biogeochemical cycles in aquatic ecosystems. However, viral diversity and functions in agricultural soils are poorly known, especially in the rhizosphere. We used virome analysis of eight rhizosphere and bulk soils to study viral diversity and potential biogeochemical impacts in an agro-ecosystem. The order *Caudovirales* was the predominant viral type in agricultural soils, with *Siphoviridae* being the most abundant family. Phylogenetic analysis of the terminase large subunit of *Caudovirales* identified high viral diversity and three novel groups. Viral community composition differed significantly between bulk and rhizosphere soils. Soil pH was the main environmental driver of viral community structure. Remarkably, abundant auxiliary carbohydrate-active enzyme (CAZyme) genes were detected in viromes, including glycoside hydrolases, carbohydrate esterases, and carbohydrate-binding modules. These results demonstrate that virus-encoded putative auxiliary metabolic genes (AMGs) or metabolic genes that may change bacterial metabolism and indirectly contribute to biogeochemical cycling, especially carbon cycling, in agricultural soil.

Introduction

Viruses are the most abundant and genetically diverse biological entities on earth (Edwards and Rohwer, 2005), and they probably infect all cellular life forms. Environmental virome studies have mainly focused on aquatic environments, especially marine waters (192 viromes from marine while 9 viromes from soils until 2016) (Cobian-Guemes et al., 2016). These studies demonstrated that viruses play a pivotal role in ocean ecosystems (Rohwer and Thurber, 2009). By lysing approximately one-third of microorganisms per day in the ocean, viruses contribute to nutrient release in the form of dissolved organic carbon (DOC) and particulate organic carbon (POC). This process affects geochemical cycles and modulates host populations and diversities (Suttle, 2007; Brum and Sullivan, 2015). Viruses also drive host evolution via virus-mediated horizontal gene transfer (Mann et al., 2003; Suttle, 2007). Metagenomics-based data indicate that viruses can impact biogeochemical cycling by reprogramming the metabolism of their hosts via the expression of virus-encoded auxiliary metabolic genes (AMGs) (Roux et al., 2016c). Those AMGs have a variety of metabolic functions, including carbon, nitrogen and sulfur metabolism (Roux et al., 2016c; Middelboe and Brussaard, 2017), photosynthesis (Sullivan et al., 2006) and phosphate scavenging (Hsieh and Wanner, 2010). Recently, viruses in some terrestrial ecosystems have been found to encode

AMGs, such as *phoH* and carbon metabolism related genes, suggesting a role in biogeochemical cycles (Adriaenssens et al., 2015; Jin et al., 2019). However, there is less information about the ecological roles of viruses in agricultural soils.

Previous viromic studies have focused on limited soil types, including deserts, agricultural soils, forest soils, and wetlands (Williamson et al., 2017). Metagenomic analysis revealed that dsDNA tailed bacteriophages were the most abundant in some American agricultural soils (Liang et al., 2019). Viral community structure has been reported to be controlled by several soil factors. In Antarctic soils, where the family *Siphoviridae* predominates, viral community structure was driven by pH, calcium content, and altitude (Adriaenssens et al., 2017). Viral communities shifted along a thawing permafrost peatland soil, and were correlated with host community composition, pH, soil moisture content, and soil depth (Emerson et al., 2018).

Some information exists on the abundance and taxonomic diversity of viruses in soils (Williamson et al., 2005; Han et al., 2017). Although Rhizosphere, the narrow zone surrounding and affected by roots, is a hotspot of soil microbial activity and nutrient element cycling (Bastian et al., 2009; Emerson, 2019; Kuzyakov and Razavi, 2019), viruses in the rhizosphere are poorly studied. Rhizosphere and rhizosphere microorganisms are important for plant health and sustainable agriculture. The rhizosphere differs from the surrounding bulk soil in a range of chemical and physical

properties, including nutrient concentrations, pH, redox potential, and partial pressures of O_2 and CO_2 (Hinsinger et al., 2005). Microbial communities in the rhizosphere differ from those in bulk soils (Mendes et al., 2013; Bakker et al., 2015). Additionally, different fertilization regimes can also affect soil microbial communities (Zhao et al., 2019), but their impact on viral communities remains unexplored.

In this study, we analyzed eight viromes from an agricultural ecosystem in southwestern China. We studied (i) taxonomic and functional diversity of viruses in agricultural soils; (ii) effects of soil habitats (rhizosphere and bulk soil) and agricultural management practices on viral communities; (iii) potential biogeochemical role of viruses in agricultural soil.

Results

Overview of soil viromes and virus abundance

Viral DNA from eight soil samples was extracted and sequenced, and yielded a range of 9.5 to 25 million clean reads (Table 1). Only 0.26% to 1.54% reads were classified as viruses and most reads (85%–94.5%) were unknown. The clean reads were assembled into 237 contigs at least 10 kb in length and VirSorter predicted 94 "confident" viral contigs (>10 kb in size and Category1, 2, or 5) (Table S1). The abundance of virus-like particles (VLPs) in bulk soils ranged from 5.04×10^8 to 1.45×10^9 g^{-1} dry weight, and 5.53×10^8 to 1.3×10^9 g^{-1} dry soil in rhizosphere soils (Table

S2). In bulk soils, virus-to-microbe ratios (VMRs) were higher in soil samples with fertilization treatments (QJB3, 4, and 7) than in control soil (QJB1). In rhizosphere soils, the abundances of viruses and microbes in inorganic nitrogen fertilizer treatment soil (QJR3) were nearly twice those of other samples (QJR1 and 4), but there was no significant change in VMRs. The abundance of microbes in QJR7 (with inorganic nitrogen fertilizer and *Klebsiella variicola* W12) was higher, which led to a decrease of VMRs in QJR7 (Table S2). However, the differences between bulk and rhizosphere soils and different fertilizer treatments in viral abundances, microbial abundances, and VMR were not significant (ANOVA $P > 0.05$).

Viral community composition and environmental driving factors

A total of 57 viral families, 352 genera and 4,373 species were identified from the eight viromes. Among these, dsDNA viruses were predominant, with primary assignment to the order *Caudovirales*. At the family level, *Siphoviridae* (34.8%–88.3%) was dominant in most samples (Fig. 1). The ssDNA *Circoviridae* family and dsDNA *Myoviridae* family also accounted for a large fraction of viromes. Followed by those three families, other viral families including *Podoviridae*, *Microviridae*, *Mimiviridae*, *Phycodnaviridae*, and *Nanoviridae* were also detected in QJ soils. The proportions of predominant viral families in bulk soils (QJB1 and QJB3, and QJB4 and QJB7) were similar, while those in rhizosphere soils were distinct from each

other.

We compared the effects of different nutrient addition treatments and soil habitats on viral community using three statistical approaches (PERMANOVA, ANOSIM and MRPP). The viral communities were significantly different between bulk soils and rhizosphere soils, but similar among different treatments (Table 2). However, these findings are based on 8 viromes which is a small dataset for statistical analyses, and need to be further tested with more soil samples. Non-metric multidimensional scaling (NMDS) analysis was used to compare the viral community between bulk and rhizosphere soils and showed that viromes from bulk and rhizosphere soil formed two separate clusters (ANOSIM R = 0.521, P = 0.032) (Fig. 2A). Linear discriminant analysis effect size (LEfSe) analysis identified four families, five subfamilies and seventeen genera that showed significantly different abundances between bulk and rhizosphere soils (Fig. 2B). A total of 61.5% belonged to *Caudovirales*, and most of these belonged to *Siphoviridae*. Four families (*Myoviridae*, *Microviridae*, *Lavidaviridae* and *Papillomaviridae*), all sub-families and fourteen genera were enriched in bulk soils, whereas only three genera were significantly enriched in rhizosphere soils.

Redundancy analysis (RDA) was used to identify the major factors contributing to the distribution of viral communities. The first and second RDA axis interpreted 40.61%

and 23.06% of the total variability of viral species, respectively (Fig. S1). Among those physicochemical properties, only soil pH had a significant impact on viral communities ($P < 0.05$).

Phylogenetic tree of *terL* gene

As *Caudovirales* were the most abundant viruses in QJ agricultural soil samples and the terminase large subunit (*terL*) is conserved in different families of *Caudovirales*, phylogenetic analysis of *terL* gene was used to assess the diversity and genetic distance of *Caudovirales* in bulk and rhizosphere soils. A total of 74 *terL* amino acid sequences were retrieved from contigs and used in the phylogenetic analysis (Fig. 3). Many sequences were affiliated with *Siphoviridae*, followed by *Podoviridae* and *Myoviridae*. Only sequences from bulk soil were divided into *Myoviridae*, which was consistent with the LefSe analysis showing that *Myoviridae* was enriched in bulk soil. These sequences formed 5 major groups within the *Siphoviridae* and *Podoviridae* families. Three groups (QJ group1, 2 and 3) phylogenetically distant from the known reference sequences illustrated previously uncharacterized diversity for *Caudovirales* in agricultural soil, and could be viewed as novel major branches, depending on the reference protein used.

Functional composition of viromes

The putative functions of the annotated open reading frames (ORFs) from viral reads

were visualized using MEGAN, and 17,259 ORFs were functionally classified by comparison with the SEED database. Twenty-two functional categories were assigned to the eight viromes (Fig. 4A). The dominant annotation of SEED (19.6%–36.6%) was the subsystem "Phages, prophages, transposable elements, plasmids" (Fig. 4A). "Phage head and packaging" presented the largest proportion of this group, followed by "phage packaging machinery" and "Phage Family Inoviridae", while "Listeria phi-A118-like prophages", "Phage photosynthesis", "T4-like cyanophage core proteins" and seven other functions were also identified, but were less abundant (Table S3). Other SEED functional categories, such as "Cofactors, Vitamins, Prosthetic Groups, Pigments" (6.35%–15.84%) and "Amino Acids and Derivatives" (5.94%–17.83%) were the most abundant following the phage ones (Fig. 4A). Other functions relating to phosphorus, protein, RNA and DNA metabolisms and photosynthesis were also identified in QJ agricultural soils. Although there was no significant difference of subsystem category (level 1) in relative abundances between bulk and rhizosphere soil, "Phage entry and exit", "Bacterial Cytoskeleton", "DNA structural proteins, bacterial" and "Polyamine Metabolism" at SEED subsystem level 2 annotations, were enriched in bulk soils whereas no gene was significantly enriched in rhizosphere soils (Fig. 4B). "Carbohydrates" was abundant in the eight soil samples, suggesting that viruses may contribute to modulation of carbon cycling.

Auxiliary carbohydrate metabolism genes in soil viromes

To clarify the viral role in carbon cycling in agricultural soils, viral contigs (> 10 kb) were identified by Virsorter and only VirSorter categories 1, 2 and 5 were considered for annotating potential carbohydrate-active enZYmes (CAZymes) on dbCAN web server (Table S4). A total of 48 viral ORFs identified as CAZymes belonged to three CAZyme functional classes (CBM_families (Carbohydrate-binding modules), CE_families (Carbohydrate Esterases), and GH_families (Glycoside hydrolases)), with primary assignment to glycoside hydrolases (Fig. 5). Forty genes were identified as lysozyme or chitinase which were used to degrade the host cell wall. The remaining genes were identified as acetyl xylan esterase, cutinase, α-amylase, etc. (Table S4). Many of CAZymes were identified from QJB7. GH23 and GH24 were the most abundant CAZymes in QJ agricultural soil, while CE5, GH0, and GH13 were present at low abundance. CE3 and GH0 were only identified in rhizosphere soils, while CE5, GH13, GH19 and GH73 only existed in bulk soils (Fig. 5).

Map analysis of CAZymes**_containing viral genomes**

CAZymes_containing viral genomes were identified by linkage information or genomic context (Fig. S2). One-third of CAZymes genes were identified from complete viral genomes annotated with the PHAge Search Tool. Two CAZymes_ containing complete genomes separately from bulk soil (QJB7) and rhizosphere soil

(QJR7) were shown in Fig. S2. QJR7_NODE7 showed 100% genetic similarity to other two genomes from QJR3 and QJR4 (QJR3_NODE5, QJR4_NODE12), and had 70.21% nucleotide identity with *Xanthomonas* virus Xp10. JQB7_NODE22 had 85.19% identity with *Salmonella* phage IME207. These genomes have 'high-confidence' viral genomic contexts, for common viral genes, such as viral structural genes and terminases were found in genomic regions both downstream and upstream of CAZymes (Fig. S2).

Discussion

Predominant viral families and novel viral groups in QJ agricultural soil

Taxonomic diversity analysis revealed that *Caudovirales* were the major viral group. *Siphoviridae* was particularly abundant in QJ agricultural soils, especially in QJR4. This result is consistent with results from southeastern USA agricultural soil and Antarctic soil (Adriaenssens et al., 2017; Liang et al., 2019). The ssDNA viruses like *Circoviridae* and *Microviridae* were also abundant (Fig. 1). They are the dominant taxa in some soil habitats, like agricultural soils and mangrove soils (Han et al., 2017; Jin et al., 2019). However, the high occurrence of ssDNA viruses might be caused by the bias of multiple displacement amplification (MDA) using the phi29 polymerase (Kim and Bae, 2011; Marine et al., 2014). We recovered ssDNA and dsDNA viruses using the Accel-NGS 1S Plus DNA Library Kit, which was efficient for quantitative

amplification when targeting both dsDNA and ssDNA viruses (Roux et al., 2016a). Methods for study of the soil virome are still being optimized (Trubl et al., 2019). Compared with other soil virome studies, we found a highly diverse array of viral families (57 viral families) in this study. Nucleocytoplasmic large DNA viruses (NCLDVs) (e.g. *Mimiviridae*, *Phycodnaviridae*, *Poxviridae*, *Iridoviridae*, *Ascoviridae*, and *Marseilleviridae*) were identified in these agricultural soils. Most of them accounted for a small proportion due to the size-fractionated method. NCLDVs were highly represented in some surface soils (Liang et al., 2019), and the phylogenetic diversity of giant viruses were recently identified in forest soil using cultivation-independent metagenomics and mini-metagenomics (Schulz et al., 2018). Many novel viruses have been discovered in different soil habitats. Several novel clades of *Caudovirales* and ssDNA eukaryotic viruses were found in mangrove soils using phylogenetic analyses (Jin et al., 2019). In this study, phylogenetic tree of *terL* gene identified three novel groups of *Siphoviridae*. Although *Caudovirales* were the best-studied viral group to date, these novel groups suggest that *Caudovirales* in agricultural soils are relatively uncharacterized and understudied. More than 85% of the reads showed no-hits against the databases, and previous studies have shown that most viral environmental sequences are underrepresented in current databases (Hurwitz and Sullivan, 2013), and the large component of uncharacterized viromes is

typically labeled as "viral dark matter" (Reyes et al., 2012).

Comparisons of viral taxonomic diversity between bulk and rhizosphere soils

Four families, *Microviridae*, *Myoviridae*, *Lavidaviridae* and *Papillomaviridae* were more abundant in bulk soil than in rhizosphere soil (Fig. 2B). *Microviridae* is a typical ssDNA bacteriophage and mainly infects enterobacteria, intracellular parasitic bacteria, and spiroplasma (Brentlinger et al., 2002). *Gokushovirinae*, a subfamily of *Microviridae*, was also enriched in bulk soils, and ubiquitous in Chinese agricultural soils (Han et al., 2017). Remarkably, in our study, *Pradovirus* was significantly enriched in rhizosphere soil (Fig 2B), and may infect *Ralstonia solanacearum* an important soil-borne plant pathogen (Addy et al., 2018). This suggests that roots may accommodate special viruses such as *Pradovirus* which could function as plant disease biocontrol agents. Recently, Shen and his team successfully used phage treatment to decrease the incidence of disease of tomato (Wang et al., 2019), but the use of phage biocontrol is still poorly studied. Many microbes have been reported as biocontrol agents, for example, *Pseudomonas* isolated from maize rhizospheres and *Rhodopseudomonas palustris* colonizing tobacco phyllosphere (Costa et al., 2006; Su et al., 2019).

In our study, the viral community in maize rhizosphere soil formed a separate cluster from the community in bulk soil (Fig. 2A). Previous study showed that land use could

change the diversity of viruses (Munoz-Arenas et al., 2020). Viromes from different biomes, like seawater, freshwater, sediment, desert, soil and air, tend to be clustered according to the sample medium types (Roux et al., 2016b; Han et al., 2017). Additionally, the heterogeneity of geochemical parameters, even over short spatial scales in wetlands control viral community composition and structure (Martins et al., 2018). We found that pH had a significant relationship with viral communities (Fig. S1). This was consistent with the results from Antarctic soils and thawing permafrost peatland soils (Adriaenssens et al., 2017; Emerson et al., 2018). Soil pH has been reported to be the most important abiotic driver of microbial community structure of bulk and rhizosphere soils (Fan et al., 2017). It may indirectly influence the viral diversity through impacts on host community composition. Additionally, plants, such as *Arabidopsis thaliana*, produce a range of specialized triterpenes that could influence and shape the microbial community within and around its roots (Huang et al., 2019). It is also argued that root exudates can also impact the host community composition, thus affecting the viral community. A new study by Starr and his colleagues uncovered a high diversity of RNA viruses in bulk and rhizosphere soil, and viral and host communities were impacted by the presence of root litter instead of growing roots (Starr et al., 2019). Overall, more studies of the soil virome are needed to decipher the major factors driving viral community composition in soil, the

underlying mechanisms as well. It is very interesting that a dialogue between microbiology and archaeology showed that the oldest sequenced RNA virus genome is a 1000-year-old maize virus (Brussow, 2020), therefore combined analysis of RNA viruses was important for expanding our understanding of the entire viral characteristics in soil which would be helpful for the use of viruses in agricultural productivity.

Potential biogeochemical impacts of viruses in agricultural soil

Viruses impact biogeochemical cycling via top-down (lysing dominant microbial hosts) and bottom-up (carrying auxiliary metabolic genes) controls in different environments (Brum and Sullivan, 2015; Trubl et al., 2018). For example, viruses infecting C-cycling-related microorganisms like methanogens and methanotrophs have been identified in a thawing permafrost peatland soil (Emerson et al., 2018; Trubl et al., 2018). In this study, predicted ORFs related to carbohydrate metabolism accounted for 4.31%–11.81% of viral functions classified by the SEED database (Fig. 4). Furthermore, a total of 48 CAZymes were identified by the dbcan server. These genes included glycoside hydrolases, carbohydrate esterases and carbohydrate-binding modules. Glycoside hydrolases that break down complex organic matter were the most abundant in our soils, which supported previous findings (Emerson et al., 2018). However, most CAZyme genes in QJ agricultural

soils were different from those in mangrove soil (Jin et al., 2019). We suppose that viral-encoded CAZymes could be environment-specific because of different carbohydrate compositions or associated host communities. As the degradation of polysaccharide is a complex process, abundant carbon processing related AMGs or CAZymes in numerous environments may boost host metabolism and promote viral propagation (Breitbart et al., 2007; Anderson et al., 2017).

This study represents an initial survey of the viral abundance, compositional and functional diversity, and potential biogeochemical impacts in bulk and rhizosphere soils. Viral community composition in rhizosphere soil was significantly different from that in bulk soil. Soil pH was the main driver of viral community composition in our datasets. Future studies based on large-scale sampling and metagenomics analysis are needed to determine whether other physicochemical properties and root exudates affect viral community composition and explain the underlying mechanisms. Abundant CAZymes were identified in our study, suggesting that viruses can contribute to biogeochemical cycling. Overall, this study expands our understanding of interactions among plant-soil-virus systems.

Experimental Procedures

Sample collection and physiochemical properties

The field experiment was established in spring of 2016 with a maize-barley rotation

system in Qujing (QJ, 25°09'40.8"N, 104°01'51.5"E, 1.925 km) in the Yunnan province, China. The soil type in QJ is classified as red soil (Paleudults in the USDA Soil Taxonomy). Soil samples were collected during the maize heading stage in August 2018. Three replicates of 4 treatments were arranged in a randomized block design in 30-m^2 plots, and the plot was 1 m away from each other. Treatments consisted of no additional fertilizer, nitrogen addition (160 kg N (urea) ha^{-1}), nitrogen addition (160 kg N (urea) ha^{-1}) plus straw (3000 kg ha^{-1}), and nitrogen addition (160 kg N(urea) ha^{-1}) plus the diazotroph *Klebsiella variicola* W12 (500 L ha^{-1} (1×10^{12} CFU/mL)) (Table S5). The fertilizer application rates of phosphorus (P) and potassium (K) were identical for all treatments at a rate of 90 kg ha^{-1} P_2O_5 and 90 kg ha^{-1} K_2O for each crop season. We sampled rhizosphere and bulk samples from each plot with five replicate locations. For bulk soil, at each location, the topsoil (0-10 cm) 15 cm away from plants was collected. For rhizosphere soil, 2 to 3 maize plants linked to a nearby bulk soil were excavated. After gently shaking away the loosely attached soil, the tightly attached soil on the roots (0-15 cm) was collected, and it served as the rhizosphere soil sample. Soil samples collected from 3 replicates were homogenized respectively, and in total we had 8 samples, namely 4 bulk soil (QJB1, QJB3, QJB4, QJB7), 4 rhizosphere soil (QJR1, QJR3, QJR4, QJR7). A total of 4 kg of each bulk and rhizosphere soil sample was collected and transported at 4℃ back to the

laboratory.

Sieved (2 mm) soil samples were stored at 4℃ for virus extraction and physicochemical analysis. Soil samples were dried for 12 h at 105℃and reweighted to measure soil water content (SWC). Soil pH was determined in a 1:2.5 soil/water suspension using a pH-meter (Professional Meter PP-20, Sartorius, Germany). Dissolved organic carbon (DOC) was extracted by 0.5 M K_2SO_4 and determined using a TOC analyzer (Multi N/C 3100, Analytikjena, Germany). NH_4^+-N and NO_3^--N were extracted with 1 M KCl and measured by a Continuous Flow Analyzer (SAN++, Skalar, Holland). Available phosphorus (AP) was measured using the Olsen method (Olsen, 1954).

Virus and microbe counting by epifluorescence microscopy

Viral and microbial abundances in each soil sample were estimated using epifluorescence microscopy (EFM) (Danovaro and Serresi, 2000; Han et al., 2017). Approximately 3 g of each sample was suspended in 30 mL of glycine buffer (250 nM, pH=8.5), vortexed for 15 min, and then centrifuged at 3,000 × g for 4 min at 4℃. The supernatant was filtered through a 0.45 μm Millex filter to remove larger particles. Finally, 200 μL of subsample was suspended in 800 μL of sterile deionized water and vacuum-filtered through a 0.02 μm Anodisc filter membrane. The dried filter membrane was stained for 15 min using SYBR Green I (Invitrogen, Eugene, Oregon,

USA) working solution (1:400) in darkness at room temperature. Subsequently, the stained Anodisc membrane was observed under EFM (Nikon, Melville, NY, USA). All samples were carried out in triplicate. For each sample, the abundances of VLPs and microbial cells were calculated from at least 15 fields of view. VMRs were calculated.

Viral DNA extraction, metagenomic library preparation and sequencing

Methods for extracting soil viromes were as previously described (Han et al., 2017; Yu et al., 2018), and viruses were extracted from all samples within ten days after transporting to the laboratory. Briefly, for each sample, 600 g of soil was suspended in 3 L of glycine buffer (250 nM, pH=8.5) and shaken for 15 min. After centrifugation, the supernatant was filtered sequentially through 1 mm, 0.45 μm and 0.20 μm tangential flow filters (QuixStand, GE Healthcare Life Sciences, Pittsburgh, PA, USA). Then the viral particles were concentrated using 30 kDa tangential flow filter system, followed by concentration in a 30 kDa centrifugal ultrafiltration tube (Merck Millipore Ltd., Tullagreen, Ireland) at 4000×g until the final volume was less than 1 mL. The viral concentrates were treated with DNase I (Thermo Fisher Scientific, Lithuania, EU) (10 units DNaseI/100 μL) and incubated at 37℃ for 1 h to remove free nucleic acids. The presence of free and contaminating bacterial DNA was checked by PCR amplification of the 16S rRNA gene with universal primers 27F/1492R (Suzuki

et al., 2000).

Viral DNA was extracted from the concentration of virus particles using the Qiagen AllPrep PowerViral DNA/RNA extraction kit (Qiagen, Hilden, Germany), following the manufacturer's protocol. Then viral DNA was fragmented to 350 bp by Covaris (Covaris, Woburn, MA, USA) and purified by Zymo Research DNA Clean & Concentrator kit (Zymo Research, Orange, CA, USA). Eight libraries were performed with the Accel-NGS 1S Plus DNA Library Kit (Swift Biosciences, Ann Arbor, MI, USA) according to manufacturer's protocol. The libraries were sequenced using an Illumina X10 at Nanjing Puwikon, Co., Ltd (Nanjing, China) to generate 150 bp pair-end reads.

Bioinformatic analyses of the virome

We used the fastp tool to remove adapters and filter low-quality reads (Chen et al., 2018). Reads mapped to the NCBI UniVec database (ftp://ftp.ncbi.nlm.nih.gov/pub/UniVec, March 2017) using bbmap with default parameters to remove mobile genetic elements. SortMeRNA v2.1 was used to identify and remove rRNA reads (Kopylova et al., 2012). Clean reads were classified using Kraken against the NCBI reference sequences (RefSeq, accessed March 2019) to identify bacterial or archaeal reads (Wood and Salzberg, 2014). For viral taxonomic identification, clean reads were compared against the NCBI non-redundant (nr)

database, Refseq virus database and all phage database from the PHAST website (until August, 2018) (Pruitt et al., 2007; Zhou et al., 2011; Agarwala et al., 2016) using diamond (BLASTx, thresholds of $1e^{-5}$ on E-value and 50 on bit score) (Buchfink et al., 2015). For functional assignments, ORFs of viral reads were predicted by Prodigal (Hyatt et al., 2010) and each predicted gene was compared to NR and RefSeq viral protein database using the diamond (BLASTp , E-Value < $1e^{-5}$ and identity > 60%) (Buchfink et al., 2015). Viral functions were annotated with the SEED database (May 2015) of MEGAN 6 with default parameters (Huson et al., 2007).

Clean reads were assembled into contigs by metaSPAdes with default parameters (Nurk et al., 2017). Viral contigs (at least 10 kb in length) were identified using the VirSorter tool by virome database with the "virome decontamination" method (Roux et al., 2015; Roux et al., 2019), and contigs of "category 1", "category 2" and "category 5" (higher confidence predictions) were selected for further analysis, which aims to minimize potential cellular genome contamination. ORFs of viral contigs were predicted by Prodigal (Hyatt et al., 2010). CAZymes from these viral ORFs were automated on the dbCAN2 meta server based on CAZyme family-specific HMMER (E-Value < $1e^{-15}$, coverage > 0.35), DIAMOND (E-Value < $1e^{-102}$) and Hotpep (Frequency > 2.6, Hits > 6) together (Zhang et al., 2018). Complete viral genomes

were identified using PHAST (Zhou et al., 2011). Genome structures were generated by Easyfig (Sullivan et al., 2011). The similarity of two complete genomes was calculated by the BLAST method (Zhang et al., 2000). Metadata for our virome analyses according to the MIUViG standards was shown in Table S6 (Roux et al., 2019).

A phylogenetic tree of *terL* gene amino acid sequences (encoding the large subunit of the terminase) of *Caudovirales* was generated using MEGA6 software (Tamura et al., 2013). Contigs homologous to the marker gene were selected using HMMER against reference sequences, as previously described by Liu et al. (Liu et al., 2018). All selected amino acid sequences were aligned using MUSCLE with default alignment parameters. The phylogeny tree with 500 bootstraps was constructed using the Jones-Taylor-Thornton (JTT) model with the maximum likelihood method in MEGA6. The output was visualized by Evolview v2 (He et al., 2016).

Statistical analyses of the virome

The effects of treatments and soil habitats (rhizosphere and bulk soils) on viral community compositions (at species level) were tested by three statistical methods (analysis of similarities (ANOSIM), multi-response permutation procedure (MRPP), and non-parametric multivariate analysis of variance (PERMANOVA)), using the "vegan" package with 999 permutations in R (Dixon, 2003; Anderson and Walsh,

2013). The viral community compositions (species-level) were analyzed by NMDS ordinations and visualized by the "ggplot2" package in R (Ginestet, 2011). The NMDS analysis was based on Bray-Curtis dissimilarity distances and calculated on the relative abundance of species-level taxonomic assignment, and statistically evaluated by ANOSIM. LEfSe was used to identify differentially abundant viruses (from order level to genus level) between bulk and rhizosphere soils, based on $P < 0.05$ and an LDA score > 2.0 (Segata et al., 2011). RDA was used to examine the relationship between viral community compositions and environmental factors (i.e. SWC, pH, DOC, NH_4^+-N, NO_3^--N and AP). Before conducting RDA analysis, the environmental factors were selected based on the variation inflation factors (VIFs) (<10) (O'Brien, 2007). The significant relationships between viral community compositions and environmental factors were calculated by the "envfit" (permutation = 999) on the R platform ($P < 0.05$).

Data availability

The raw sequence files obtained from this research were submitted to the NCBI Sequence Read Archive (SRA) with accession number PRJNA588894.

Acknowledgments

We thank Drs. Forest Rohwer, Cynthia Silveira and Ana Georgina Cobián Güemes (San Diego State University) for data analysis and manuscript review, Pengpeng Xu

(Nanjing Puwikon, Co., Ltd) for technical support in bioinformatics analysis, Zhongwang Jing and Anhui Ge (University of Chinese Academy of Sciences) for statistics analyses. This work was supported by National Key R&D Program (grant number: 2017YFD0200600), the National Science Foundation of China (grant number: 41771289 and 41571248), the Strategic Priority Research Pro-gram of the Chinese Academy of Sciences (grant number: XDB15020200) and the China Scholarship Council.

Conflict of Interest

The authors declare no conflicts of interest.

References

Addy, H.S., Farid, M.M., Ahmad, A.A., and Huang, Q. (2018) Host range and molecular characterization of a lytic Pradovirus-like Ralstonia phage RsoP1IDN isolated from Indonesia. *Arch Virol* **163**: 3409-3414.

Adriaenssens, E.M., Kramer, R., Van Goethem, M.W., Makhalanyane, T.P., Hogg, I., and Cowan, D.A. (2017) Environmental drivers of viral community composition in Antarctic soils identified by viromics. *Microbiome* **5**: 83.

Adriaenssens, E.M., Van Zyl, L., De Maayer, P., Rubagotti, E., Rybicki, E., Tuffin, M., and Cowan, D.A. (2015) Metagenomic analysis of the viral community in Namib Desert hypoliths. *Environ Microbiol* **17**: 480-495.

Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E. et al. (2016) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **44**: D7-D19.

Anderson, C.L., Sullivan, M.B., and Fernando, S.C. (2017) Dietary energy drives the dynamic response of bovine rumen viral communities. *Microbiome* **5**: 155.

Anderson, M.J., and Walsh, D.C. (2013) PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? *Ecol Monogr* **83**: 557-574.

Bakker, M.G., Chaparro, J.M., Manter, D.K., and Vivanco, J.M. (2015) Impacts of bulk soil microbial community structure on rhizosphere microbiomes of Zea mays. *Plant Soil* **392**: 115-126.

Bastian, F., Bouziri, L., Nicolardot, B., and Ranjard, L. (2009) Impact of wheat straw decomposition on successional patterns of soil microbial community structure. *Soil Biol Biochem* **41**: 262-275.

Breitbart, M., Thompson, L.R., Suttle, C.A., and Sullivan, M.B. (2007) Exploring the Vast Diversity of Marine Viruses. *Oceanography* **20**: 135-139.

Brentlinger, K.L., Hafenstein, S., Novak, C.R., Fane, B.A., Borgon, R., McKenna, R., and Agbandje-McKenna, M. (2002) Microviridae, a family divided: Isolation, characterization, and genome sequence of phi MH2K, a bacteriophage of the obligate intracellular parasitic bacterium Bdellovibrio bacteriovorus. *J Bacteriol* **184**: 1089-1094.

Brum, J.R., and Sullivan, M.B. (2015) Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat Rev Microbiol* **13**: 147-159.

Brussow, H. (2020) Bioarchaeology: a profitable dialogue between microbiology and archaeology. *Microb Biotechnol* **13**: 406-409.

Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**: 59-60.

Chen, S.F., Zhou, Y.Q., Chen, Y.R., and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: 884-890.

Cobian-Guemes, A.G., Youle, M., Cantu, V.A., Felts, B., Nulton, J., and Rohwer, F. (2016) Viruses as Winners in the Game of Life. *Annu Rev Virol* **3**: 197-214.

Costa, R., Gomes, N.C.M., Peixoto, R.S., Rumjanek, N., Berg, G., Mendonca-Hagler, L.C.S., and Smalla, K. (2006) Diversity and antagonistic potential of Pseudomonas spp. associated to the

rhizosphere of maize grown in a subtropical organic farm. *Soil Biol Biochem* **38**: 2434-2447.

Danovaro, R., and Serresi, M. (2000) Viral density and virus-to-bacterium ratio in deep-sea sediments of the Eastern Mediterranean. *Appl Environ Microb* **66**: 1857-1861.

Dixon, P. (2003) VEGAN, a package of R functions for community ecology. *J Veg Sci* **14**: 927-930.

Edwards, R.A., and Rohwer, F. (2005) Viral metagenomics. *Nat Rev Microbiol* **3**: 504-510.

Emerson, J.B. (2019) Soil Viruses: A New Hope. *Msystems* **4**: e00120-00119.

Emerson, J.B., Roux, S., Brum, J.R., Bolduc, B., Woodcroft, B.J., Jang, H.B. et al. (2018) Host-linked soil viral ecology along a permafrost thaw gradient. *Nat Microbiol* **3**: 870-880.

Fan, K.K., Cardona, C., Li, Y.T., Shi, Y., Xiang, X.J., Shen, C.C. et al. (2017) Rhizosphere-associated bacterial network structure and spatial distribution differ significantly from bulk soil in wheat crop fields. *Soil Biol Biochem* **113**: 275-284.

Ginestet, C. (2011) ggplot2: Elegant Graphics for Data Analysis. *J R Stat Soc a Stat* **174**: 245-245.

Han, L.L., Yu, D.T., Zhang, L.M., Shen, J.P., and He, J.Z. (2017) Genetic and functional diversity of ubiquitous DNA viruses in selected Chinese agricultural soils. *Sci Rep-Uk* **7**: 45142

He, Z.L., Zhang, H.K., Gao, S.H., Lercher, M.J., Chen, W.H., and Hu, S.N. (2016) Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res* **44**: W236-W241.

Hinsinger, P., Gobran, G.R., Gregory, P.J., and Wenzel, W.W. (2005) Rhizosphere geometry and heterogeneity arising from root-mediated physical and chemical processes. *New Phytol* **168**: 293-303.

Hsieh, Y.J., and Wanner, B.L. (2010) Global regulation by the seven-component Pi signaling system. *Curr Opin Microbiol* **13**: 198-203.

Huang, A.C.C., Jiang, T., Liu, Y.X., Bai, Y.C., Reed, J., Qu, B.Y. et al. (2019) A specialized metabolic network selectively modulates Arabidopsis root microbiota. *Science* **364**: eaau6389.

Hurwitz, B.L., and Sullivan, M.B. (2013) The Pacific Ocean Virome (POV): A Marine Viral Metagenomic Dataset and Associated Protein Clusters for Quantitative Viral Ecology. *Plos One* **8**: e57355.

Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Res* **17**: 377-386.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform* **11**: 119.

Jin, M., Guo, X., Zhang, R., Qu, W., Gao, B.L., and Zeng, R.Y. (2019) Diversities and potential biogeochemical impacts of mangrove soil viruses. *Microbiome* **7**: 58.

Kim, K.H., and Bae, J.W. (2011) Amplification Methods Bias Metagenomic Libraries of Uncultured Single-Stranded and Double-Stranded DNA Viruses. *Appl Environ Microb* **77**: 7663-7668.

Kopylova, E., Noe, L., and Touzet, H. (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**: 3211-3217.

Kuzyakov, Y., and Razavi, B.S. (2019) Rhizosphere size and shape: Temporal dynamics and spatial stationarity. *Soil Biol Biochem* **135**: 343-360.

Liang, X.L., Wagner, R.E., Zhuang, J., DeBruyn, J.M., Wilhelm, S.W., Liu, F. et al. (2019) Viral

abundance and diversity vary with depth in a southeastern United States agricultural ultisol. *Soil Biol Biochem* **137**: 107546.

Liu, Y.R., Johs, A., Bi, L., Lu, X., Hu, H.W., Sun, D. et al. (2018) Unraveling Microbial Communities Associated with Methylmercury Production in Paddy Soils. *Environ Sci Technol* **52**: 13110-13118.

Mann, N.H., Cook, A., Millard, A., Bailey, S., and Clokie, M. (2003) Marine ecosystems: Bacterial photosynthesis genes in a virus. *Nature* **424**: 741-741.

Marine, R., McCarren, C., Vorrasane, V., Nasko, D., Crowgey, E., Polson, S.W., and Wommack, K.E. (2014) Caught in the middle with multiple displacement amplification: the myth of pooling for avoiding multiple displacement amplification bias in a metagenome. *Microbiome* **2**: 3.

Martins, P.D., Danczak, R.E., Roux, S., Frank, J., Borton, M.A., Wolfe, R.A. et al. (2018) Viral and metabolic controls on high rates of microbial sulfur and carbon cycling in wetland ecosystems. *Microbiome* **6**: 138.

Mendes, R., Garbeva, P., and Raaijmakers, J.M. (2013) The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *Fems Microbiol Rev* **37**: 634-663.

Middelboe, M., and Brussaard, C.P.D. (2017) Marine Viruses: Key Players in Marine Ecosystems. *Viruses-Basel* **9**: 302.

Munoz-Arenas, L.C., Fusaro, C., Hernandez-Guzman, M., Dendooven, L., Estrada-Torres, A., and Navarro-Noya, Y.E. (2020) Soil microbial diversity drops with land-use change in a high mountain temperate forest: a metagenomics survey. *Env Microbiol Rep* **12**: 185-194.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017) metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27**: 824-834.

O'Brien, R.M. (2007) A caution regarding rules of thumb for variance inflation factors. *Qual Quant* **41**: 673-690.

Olsen, S.R. (1954) Estimation of available phosphorus in soils by extraction with sodium bicarbonate. *US Dep Agric Circ* **939**: 1-19.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**: D61-D65.

Reyes, A., Semenkovich, N.P., Whiteson, K., Rohwer, F., and Gordon, J.I. (2012) Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat Rev Microbiol* **10**: 607-617.

Roesch, L.F.W., Camargo, F.A.O., Bento, F.M., and Triplett, E.W. (2008) Biodiversity of diazotrophic bacteria within the soil, root and stem of field-grown maize. *Plant Soil* **302**: 91-104.

Rohwer, F., and Thurber, R.V. (2009) Viruses manipulate the marine environment. *Nature* **459**: 207-212.

Roux, S., Enault, F., Hurwitz, B.L., and Sullivan, M.B. (2015) VirSorter: mining viral signal from microbial genomic data. *Peerj* **3**: e985

Roux, S., Solonenko, N.E., Dang, V.T., Poulos, B.T., Schwenk, S.M., Goldsmith, D.B. et al. (2016a)

Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *Peerj* **4**: e2777.

Roux, S., Adriaenssens, E.M., Dutilh, B.E., Koonin, E.V., Kropinski, A.M., Krupovic, M. et al. (2019) Minimum information about an uncultivated virus genome (MIUViG). *Nat Biotechnol* **37**: 29-37.

Roux, S., Enault, F., Ravet, V., Colombet, J., Bettarel, Y., Auguet, J.C. et al. (2016b) Analysis of metagenomic data reveals common features of halophilic viral communities across continents. *Environ Microbiol* **18**: 889-903.

Roux, S., Brum, J.R., Dutilh, B.E., Sunagawa, S., Duhaime, M.B., Loy, A. et al. (2016c) Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**: 689.

Schulz, F., Alteio, L., Goudeau, D., Ryan, E.M., Yu, F.Q.B., Malmstrom, R.R. et al. (2018) Hidden diversity of soil giant viruses. *Nat Commun* **9**: 4881.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C. (2011) Metagenomic biomarker discovery and explanation. *Genome Biol* **12**: R60.

Starr, E.P., Nuccio, E.E., Pett-Ridge, J., Banfield, J.F., and Firestone, M.K. (2019) Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. *P Natl Acad Sci USA* **116**: 25900-25908.

Su, P., Zhang, D., Zhang, Z., Chen, A., Hamid, M.R., Li, C. et al. (2019) Characterization of Rhodopseudomonas palustris population dynamics on tobacco phyllosphere and induction of plant resistance to Tobacco mosaic virus. *Microb Biotechnol* **12**: 1453-1463.

Sullivan, M.B., Lindell, D., Lee, J.A., Thompson, L.R., Bielawski, J.P., and Chisholm, S.W. (2006) Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *Plos Biol* **4**: 1344-1357.

Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011) Easyfig: a genome comparison visualizer. *Bioinformatics* **27**: 1009-1010.

Suttle, C.A. (2007) Marine viruses - major players in the global ecosystem. *Nat Rev Microbiol* **5**: 801-812.

Suzuki, M.T., Taylor, L.T., and DeLong, E.F. (2000) Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5 '-nuclease assays. *Appl Environ Microb* **66**: 4605-4614.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* **30**: 2725-2729.

Trubl, G., Roux, S., Solonenko, N., Li, Y.F., Bolduc, B., Rodriguez-Ramos, J. et al. (2019) Towards optimized viral metagenomes for double-stranded and single-stranded DNA viruses from challenging soils. *Peerj* **7**: e7265.

Trubl, G., Jang, H.B., Roux, S., Emerson, J.B., Solonenko, N., Vik, D.R. et al. (2018) Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing. *Msystems* **3**: e00076-00018.

Wang, X., Wei, Z., Yang, K., Wang, J., Jousset, A., Xu, Y. et al. (2019) Phage combination therapies for bacterial wilt disease in tomato. *Nat Biotechnol* **37**: 1513-1520.

Williamson, K.E., Radosevich, M., and Wommack, K.E. (2005) Abundance and diversity of viruses in six Delaware soils. *Appl Environ Microb* **71**: 3119-3125.

Williamson, K.E., Fuhrmann, J.J., Wommack, K.E., and Radosevich, M. (2017) Viruses in Soil Ecosystems: An Unknown Quantity Within an Unexplored Territory. *Annu Rev Virol* **4**: 201-219.

Wood, D.E., and Salzberg, S.L. (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* **15**: R46.

Yu, D.T., Han, L.L., Zhang, L.M., and He, J.Z. (2018) Diversity and Distribution Characteristics of Viruses in Soils of a Marine-Terrestrial Ecotone in East China. *Microb Ecol* **75**: 375-386.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P.Z., Yang, Z.L. et al. (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **46**: W95-W101.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7**: 203-214.

Zhao, Z.B., He, J.Z., Geisen, S., Han, L.L., Wang, J.T., Shen, J.P. et al. (2019) Protist communities are more sensitive to nitrogen fertilization than other microorganisms in diverse agricultural soils. *Microbiome* **7**: 33.

Zhou, Y., Liang, Y.J., Lynch, K.H., Dennis, J.J., and Wishart, D.S. (2011) PHAST: A Fast Phage Search Tool. *Nucleic Acids Res* **39**: W347-W352.

Table 1 Overview of sequences of eight soil viromes.

Sample	Number of clean reads	Total size of clean reads (Gbp)	GC content	hit (%)	Number of contigs (>10 kb) identified by Virsorter	Longest contig (bp)
QJB1	25,236,158	3.66	59.14%	0.35	3	65,802
QJB3	23,711,163	3.44	58.75%	0.52	9	65,802
QJB4	9,550,902	1.40	57.61%	1.07	13	324,666
QJB7	16,726,092	2.48	57.41%	1.51	31	324,666
QJR1	24,175,567	3.50	55.73%	0.26	8	59,010
QJR3	15,735,058	2.32	49.39%	0.59	5	49,603
QJR4	13,110,016	1.93	50.40%	1.54	14	112,923
QJR7	17,937,585	2.61	45.63%	1.18	11	64,301

Table 2 Significance tests of the effects of treatments and soil habitats on viral community composition using three different statistical approaches (PERMANOVA, ANOSIM and MRPP). Values in bold indicate a significant difference between the treatments or soil habitats ($P < 0.05$).

Treatments	PERMANOVA		ANOSIM		MRPP	
	F	*P*	*R*	*P*	δ	*P*
Bulk vs. Rhizosphere	**2.767**	**0.030**	**0.521**	**0.032**	**0.532**	**0.029**
N vs. No-N	1.991	0.104	0.260	0.113	0.552	0.105
Straw vs. No-straw	0.870	0.415	0.125	0.270	0.625	0.789
W12 vs. No-W12	0.776	0.622	0.000	0.467	0.624	0.773

N: samples with nitrogen fertilization, No-N: samples without nitrogen fertilizer, Straw: samples with straws, No-straw: samples without straw, W-12: samples with W12, No-W12: samples without W12. In case of Bulk vs. Rhizosphere, there were 4 viromes vs. 4 viromes. As for the rest of comparisons, there were 2 viromes vs. 6 viromes.

Figure Legends

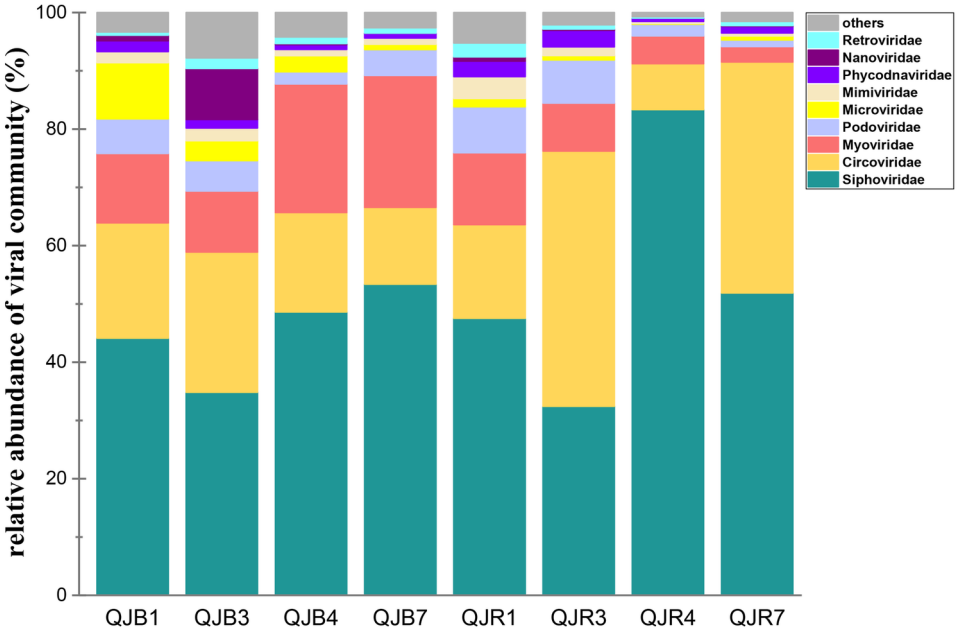
Fig. 1 Taxonomic composition of viromes in QJ bulk and rhizosphere soils, assessed for all virus-associated reads at the family level. 48 different vial families (abundance <1%) make up "Other" category. The library construction was performed with the Accel-NGS 1S Plus DNA Library Kit.

Fig. 2 Diversity of viral taxonomic composition in bulk and rhizosphere soils. Non-metric multidimensional scaling (NMDS) analysis of viral communities between bulk and rhizosphere soil. Bray-Curtis dissimilarity distances were calculated based on the relative abundance of species-level taxonomic assignment (A). Linear discriminant analysis effect size (LEfSe) showing differentially abundant viruses (from order level to genus level) between bulk and rhizosphere soils, based on $P < 0.05$ and an LDA score > 2.0 (B).

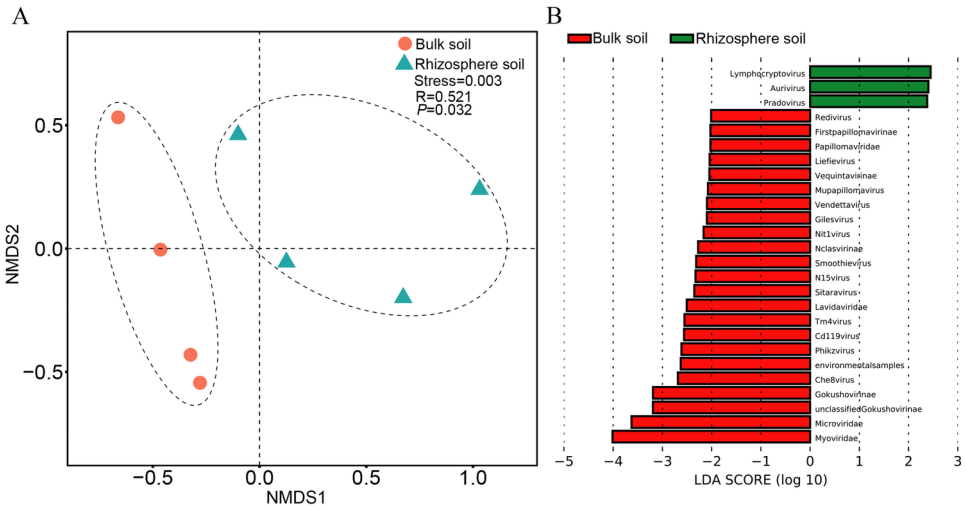
Fig. 3 Maximum likelihood tree of the terminase large subunit protein (*terL*) of *Caudovirales*. The tree was bootstrapped with 500 sub-replicates, and bootstrap scores > 50% are flagged with circles. Contigs of bulk and rhizosphere soil are colored in red and blue respectively. Reference sequences are colored in black.

Fig. 4 Functional profile (SEED level 1 subsystem) of eight viromes (A). Occurrence of significant variations in relative abundances of the SEED subsystem level 2 annotations between the bulk and rhizosphere soils (B).

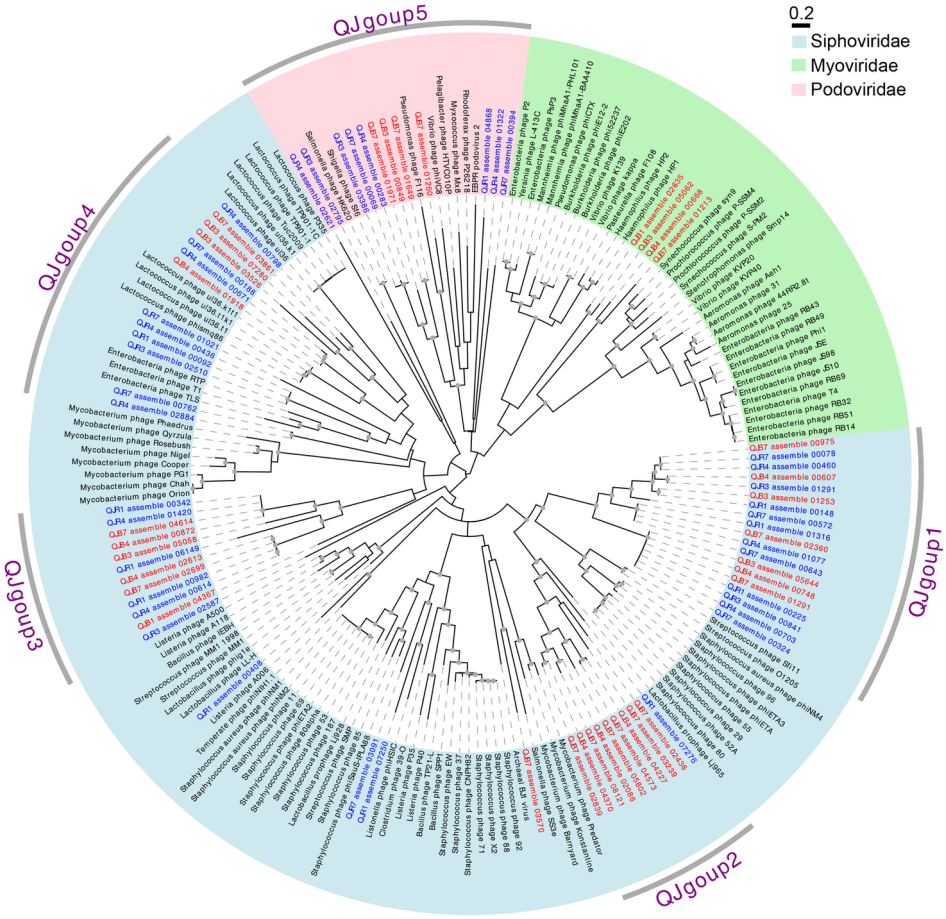
Fig. 5 Annotated auxiliary CAZymes from QJ bulk and rhizosphere soils.



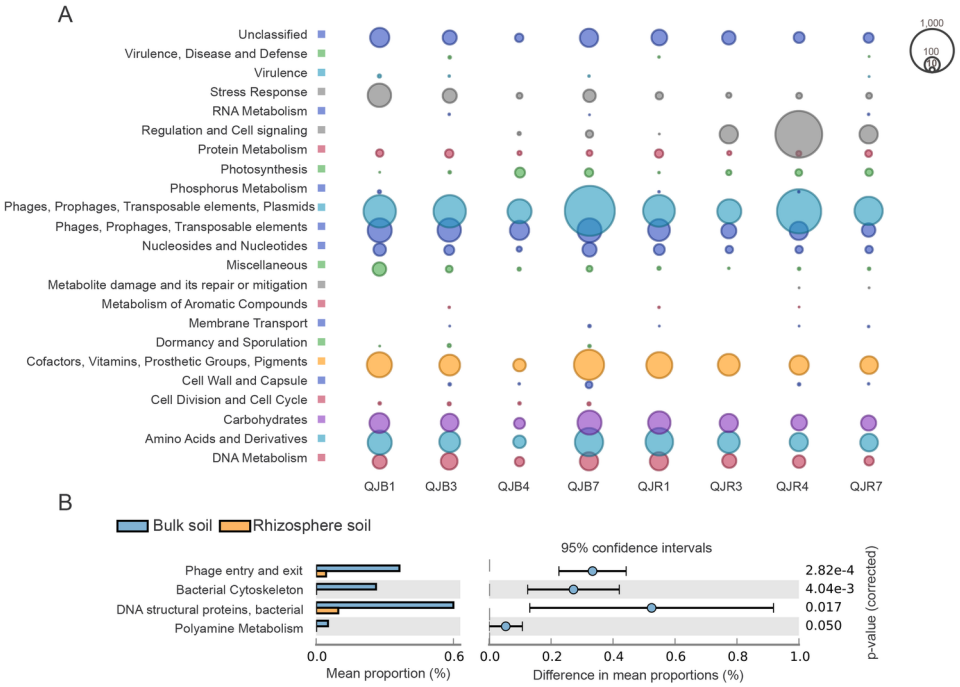
EMI_15010_Figure_1.tif



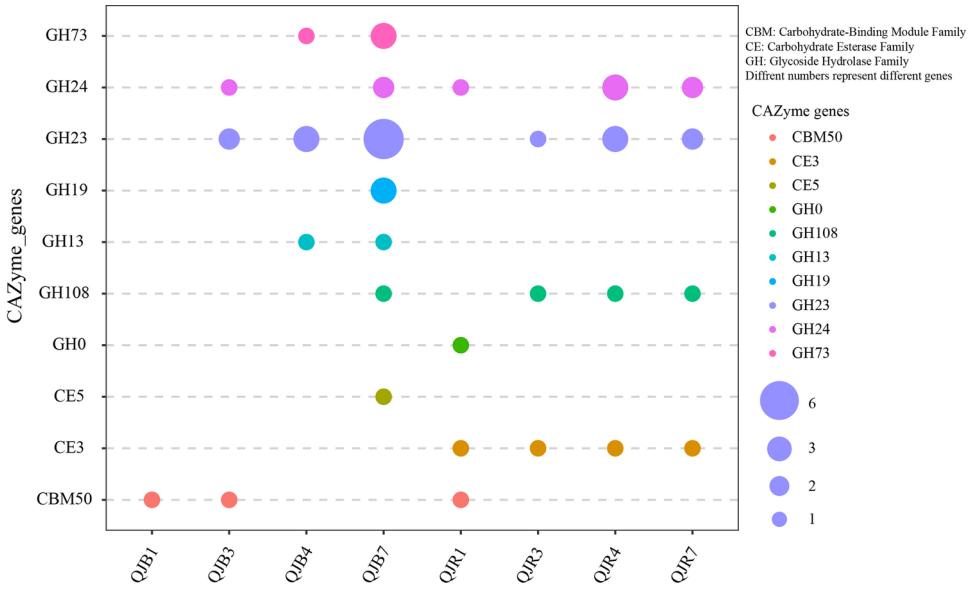
EMI_15010_Figure_2.tif



EMI_15010_Figure_3.tif



EMI_15010_Figure_4.tif



EMI_15010_Figure_5.tif

Author/s:
Bi, L; Yu, D-T; Du, S; Zhang, L-M; Zhang, L-Y; Wu, C-F; Xiong, C; Han, L-L; He, J-Z

Title:
Diversity and potential biogeochemical impacts of viruses in bulk and rhizosphere soils

Date:
2020-04-16

Citation:
Bi, L., Yu, D. -T., Du, S., Zhang, L. -M., Zhang, L. -Y., Wu, C. -F., Xiong, C., Han, L. -L. & He, J. -Z. (2020). Diversity and potential biogeochemical impacts of viruses in bulk and rhizosphere soils. Environmental Microbiology, 23 (2), pp.588-599. https://doi.org/10.1111/1462-2920.15010.