

RESEARCH

Open Access



Crop management shapes the diversity and activity of DNA and RNA viruses in the rhizosphere

George Muscatt¹, Sally Hilton¹, Sebastien Raguideau^{1,2}, Graham Teakle¹, Ian D. E. A. Lidbury^{1,3}, Elizabeth M. H. Wellington¹, Christopher Quince^{1,2}, Andrew Millard^{4*}, Gary D. Bending¹ and Eleanor Jameson^{1,5*}

Abstract

Background: The rhizosphere is a hotspot for microbial activity and contributes to ecosystem services including plant health and biogeochemical cycling. The activity of microbial viruses, and their influence on plant-microbe interactions in the rhizosphere, remains undetermined. Given the impact of viruses on the ecology and evolution of their host communities, determining how soil viruses influence microbiome dynamics is crucial to build a holistic understanding of rhizosphere functions.

Results: Here, we aimed to investigate the influence of crop management on the composition and activity of bulk soil, rhizosphere soil, and root viral communities. We combined viromics, metagenomics, and metatranscriptomics on soil samples collected from a 3-year crop rotation field trial of oilseed rape (*Brassica napus* L.). By recovering 1059 dsDNA viral populations and 16,541 ssRNA bacteriophage populations, we expanded the number of underexplored *Leviviricetes* genomes by > 5 times. Through detection of viral activity in metatranscriptomes, we uncovered evidence of “Kill-the-Winner” dynamics, implicating soil bacteriophages in driving bacterial community succession. Moreover, we found the activity of viruses increased with proximity to crop roots, and identified that soil viruses may influence plant-microbe interactions through the reprogramming of bacterial host metabolism. We have provided the first evidence of crop rotation-driven impacts on soil microbial communities extending to viruses. To this aim, we present the novel principal of “viral priming,” which describes how the consecutive growth of the same crop species primes viral activity in the rhizosphere through local adaptation.

Conclusions: Overall, we reveal unprecedented spatial and temporal diversity in viral community composition and activity across root, rhizosphere soil, and bulk soil compartments. Our work demonstrates that the roles of soil viruses need greater consideration to exploit the rhizosphere microbiome for food security, food safety, and environmental sustainability.

Keywords: Bacteriophage, Crop rotation, Kill-the-Winner, Metagenomics, Metatranscriptomics, Rhizosphere, Roots, Soil viruses, Viral priming, Viromics

Background

Soils harbor organisms from multiple kingdoms of life and provide ecosystems for > 25% of Earth's biodiversity [1]. Viruses, the smallest microorganisms in terrestrial ecosystems, often exceed the number of co-existing bacteria [2], with up to 10¹⁰ virus-like particles per gram of soil [3]. Of particular interest are the viruses of microbes,

*Correspondence: adm39@le.ac.uk; E.jameson@bangor.ac.uk

⁴ Department of Genetics and Genome Biology, University of Leicester, Leicester, UK

⁵ School of Natural Sciences, Bangor University, Bangor, UK

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

whose lytic activity provides top-down control of microbial host populations, and whose expression of viral encoded auxiliary metabolic genes (AMGs) modulates host metabolism [4–7]. In marine ecosystems, viruses have been estimated to turnover ~ 20% of microbial biomass each day [8], resulting in drastic impacts on ocean carbon and nutrient cycling [9, 10]. Given that there is an estimated 70 times more terrestrial biomass than marine biomass [11], and that viral infection rates are speculated to be greater in soils than oceans [12], there is a significant interest in unearthing the importance of viruses in terrestrial ecosystems [13, 14]. The physical structure of soil, however, hinders the extraction and subsequent cultivation of soil viruses, resulting in the current knowledge gap surrounding the ecological roles of viruses in soils [15].

Circumventing the requirement to culture viruses and their microbial hosts, metagenomics and viral size-fractionated metagenomics (viromics) have facilitated the estimation of total viral community composition and diversity across Earth's ecosystems [16–20]. Moreover, the recent optimization of soil viromic protocols [15, 21, 22] and de novo viral prediction tools [23–25] have enabled the systematic characterization of soil viral communities. However, conventional DNA approaches are unable to reveal the activity of recovered viruses. To overcome this, metatranscriptomics can be applied to characterize gene expression through the quantification of sequenced messenger RNA transcripts [26]. Given that viruses require host cell machinery for the transcription of their genes, viral activity can be used to indicate host infection. Additionally, RNA viral genomes can be assembled from metatranscriptomes, which has revealed the abundance and activity of single-stranded RNA (ssRNA) bacteriophages (phages) in both non-terrestrial [27–29] and terrestrial ecosystems alike [30–33]. For example, the discovery and emergent role of a disproportionately understudied class of ssRNA soil phages in terrestrial biogeochemistry, named *Leviviricetes* [30, 31]. Despite the advantages of combining metatranscriptomics with metagenomics to simultaneously investigate the composition and activity of DNA and RNA viral communities, there has been no such implementation in previous soil viromics studies.

Plants release ~ 20% of the carbon assimilated during photosynthesis into the soil through root exudates [34]. This provides labile nutrients and energy to microorganisms in the soil adjacent to the root system, known as the rhizosphere. Subsequently, the rhizosphere soil compartment contains greater microbial density and activity than surrounding bulk soil [35] and contributes to ecosystem services including plant health and biogeochemical cycling [36–38]. While growing evidence implicates soil

viruses in contributing to terrestrial carbon and nutrient cycling [16, 39–43], viruses remain a black box in soil and rhizosphere ecology. It is unclear whether the rhizosphere is a zone of high viral density and activity, and the effects of viral activity on plant-microbe interactions remain undetermined.

Agricultural management utilizes a variety of strategies to maintain soil fertility and productivity. The impacts of these on soil and rhizosphere microbiomes have been intensively studied, with the focus on prokaryote and eukaryote communities [44, 45], while interactions with viral communities have received little or no attention. Crop rotation is a widespread practice in which different crop plant species are grown sequentially to improve soil fertility and reduce pest and pathogen pressures [46, 47]. Subsequently, crop rotation has been associated with shifts in bacterial, fungal, and archaeal community compositions, with resulting benefits to crop health and yield [48–51]. However, there is no understanding of the effects of rotation on viral communities or the associated interactions with microbial communities. Given the impact of viruses on the ecology and evolution of their host communities in non-soil systems [52, 53], determining the roles of soil viruses in moderating microbiome dynamics is crucial to build a holistic understanding of rhizosphere functions [54].

Thus, we aimed to investigate the influence of crop rotation on the composition and activity of bulk soil, rhizosphere soil, and root viral communities. Combining viromics, metagenomics, and metatranscriptomics, we recovered novel double-stranded DNA (dsDNA) and ssRNA viral operational taxonomic units (vOTUs), expanding the number of known *Leviviricetes* genomes by > 5 times. Next, we simultaneously estimated the relative contributions of compartment and crop rotation in shaping the composition of DNA and RNA soil viral communities, relative to bacterial communities. Lastly, we characterize the spatiotemporal activity of DNA viral communities across three stages of crop growth, revealing dynamic viral-host interactions across the root-associated microbiomes.

Methods

Field site

The field site was established in 2014 at the University of Warwick Crop Centre in Wellesbourne, UK, following conventional management as previously described [51]. Eight plots of 24 m × 6 m were set up as shown in Fig. S1, allowing for four replicate samples of the two crop management practices. Two crop growth strategies of oilseed rape (*Brassica napus* L.) were adopted: continuous cropping, whereby oilseed rape was grown for three consecutive years; and virgin rotation, whereby oilseed

rape was grown following two preceding years of winter wheat (*Triticum aestivum*). The soil was a sandy-loam of the Wick series, with 73% sand, 12% silt, 14% clay, a pH of 6.5, and organic carbon content of 0.8% [55].

Sample collection

Samples were collected from each plot at three time points (November 2016, March 2017, June 2017) during the growing season of year 3 (2016/2017). For each sample, eight plants were taken from the plot by sampling ~ 1 m into the plot to avoid the edge. Loosely adhered soil was removed from the roots by tapping. The roots from all eight plants (or for large roots, six ~ 5 cm root sections were used per plant) were transferred to a 50-mL tube containing 20-mL autoclaved Milli-Q water and shaken for 20 s (first wash). The roots were transferred to a second tube and washing was repeated (second wash). The first and second washes were combined and frozen in liquid nitrogen (rhizosphere soil samples). The roots were washed a final time in 20-mL water, transferred to an empty 50-mL tube, and frozen in liquid nitrogen (root samples). This whole process was performed in the field in < 5 min. The bulk soil was sampled from each plot by selecting areas between plants ~ 50 cm into the plot to avoid the edge. Two to 3 mm of surface soil was removed, and an auger was used to collect the soil to a depth of 15 cm. Eight soil cores were sampled per plot, combined, and added to a falcon tube containing 30-mL water to take the total volume to 45-mL. The tubes were shaken for 40 s and frozen in liquid nitrogen (bulk soil samples). All samples were stored at -80 °C. The rhizosphere soil and bulk soil samples were subsequently freeze-dried. The root samples were homogenized under liquid nitrogen using a mortar and pestle.

RNA and DNA extractions

RNA extractions were performed on all root and soil samples from the three time points. RNA was extracted from 1 g of homogenized root or 2 g of soil (rhizosphere soil or bulk soil) using the RNeasy PowerSoil Total RNA Isolation Kit (Qiagen, Hilden, Germany) with two homogenizations in a Fastprep machine (MP Biomedicals) at 5.5 m/s for 30 s, resting on ice for 5 min between runs. RNA was eluted in a 50- μ L elution buffer, and 46- μ L was subsequently DNase-treated (DNase MaxTM) according to the manufacturer's instructions. The DNase was then removed using the DNase MaxTM Removal Resin. The RNA was checked for residual contaminating DNA using 16S rRNA gene universal primers. DNA-free RNA was then purified using RNAClean XP Beads (New England Biolabs) according to the manufacturer's instructions. The RNA was quantified using Qubit RNA BR kit on a Qubit[®] fluorometer (Invitrogen, CA, USA).

DNA extractions for total metagenomes and size-fractionated metagenomes (DNA viromes) were only performed on soil samples from the second time point (March 2017). For these samples, a total DNA was eluted from the same column as the RNA extractions using the RNeasy PowerSoil DNA Elution Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. The DNA was quantified using Qubit DNA HS and its purity profile checked using a Nanodrop 2.

Size-fractionated DNA for DNA viromes was extracted from ~ 5 g of soil. Briefly, soil was mixed into a total volume of 50-mL of sterile PBS and shaken vigorously for ~ 5 min, before being gently agitated on a tube roller for 1 h. Following centrifugation at 500 g to pellet large material, the supernatant was removed, sequentially filtered through 0.44 μ m and 0.22 μ m pore size filters and concentrated using Amicon 100 kDa columns as previously described. The sample was DNase I treated (1 U/ μ L) for 60 min at room temperature to remove free contaminating DNA. Viral fraction DNA was extracted through sequential rounds of phenol: chloroform as previously described [56].

Library construction and sequencing

RNA sequencing was performed by the Earlham Institute, Norwich, UK. Libraries were made using the Illumina TruSeq RNA library (HT, non-directional) kit, and all libraries were run across two lanes of the Illumina HiSeq 2500 platform (2 x 150 bp). Following sequencing, Trimmomatic v0.36 [57] was used to remove any TruSeq adapters from the sequences. SortmeRNA [58] was then used to separate and retain the rRNA reads. The forward reads (R1) were quality filtered using VSEARCH with a fast-maxee of 1 and a minimum length of 100 nt. This dataset was used as the raw metatranscriptome for read mapping.

The 16S rRNA gene operational taxonomic unit (OTU) table was generated by first assigning taxonomy to rRNA reads using QIIME and the SILVA database (version 132) at 99% identity. Then, only reads assigned to bacteria (representing 16S rRNA gene transcripts) were retained, while reads assigned to mitochondria or chloroplasts were removed.

Libraries for total metagenome sequencing were prepared and sequenced by Novogene Ltd on an Illumina HiSeq (2 x 150 bp).

Libraries for DNA virome sequencing were prepared using 1 ng of input DNA for the NexteraXT library preparation, following the manufacturer's instructions. Libraries were sequenced on an Illumina MiSeq in 2 flow cells using v3 chemistry (2 x 300 bp).

Read processing and assembly

Metatranscriptome reads were quality-filtered and trimmed with `trim_galore v0.5.0_dev` [59], and then assembled with `SPAdes v3.14.0` [60, 61] using the script `rna.spades.py` and default settings. Total metagenome reads were quality-filtered and trimmed with `trim_galore v0.5.0_dev` [59] and then assembled with `MEGAHIT v1.2.9` [62, 63] using `-kmer` steps of “27,37,47,57,67,77,87,97,107,117,127,137,141”. DNA virome reads were quality-filtered and trimmed with `sickle v1.33`. The viral DNA libraries were then assembled with `MEGAHIT` using `-kmer` steps of “21,41,61,81,101,121,141,161,181,201,221,241,249”.

Recovery of viral populations

dsDNA viral contigs were predicted from the pooled assembled reads from all soil samples, independently for each of the three libraries i.e., DNA virome, total metagenome, and metatranscriptome. For the DNA virome and metatranscriptome, viral contigs were predicted with `DeepVirFinder v1.0` [25] and filtered for $q < 0.05$ (estimated for false discovery rate of 0.1) and contig length ≥ 10 kb, rather than using a viral score threshold. For the total metagenome, viral contigs were predicted with `VIBRANT v1.0.1` [24] and filtered for contig length ≥ 10 kb, with proviral sequences > 5 kb retained. In using `NexeraXT` for DNA virome and total metagenome sequencing, our recovery approach will have targeted viruses with dsDNA genomes. dsDNA viral contigs predicted from the three libraries were combined and de-duplicated at 95% nucleotide identity across 95% of the contig length using `CD-HIT v4.6` [64] to define 1059 non-redundant vOTUs, representing approximately species-level dsDNA viral populations, in accordance with benchmarking [65]. The quality of recovered vOTUs was estimated with `CheckV v0.8.1` [66]. To determine whether any recovered vOTUs represented previously isolated phage species, we computed the pairwise `MinHash` genome distances (D) to a custom database of all complete phage genomes that were available at the time (May 2020) [67] using `MASH v2.0` [68]. Average nucleotide identity (ANI) was estimated by $1 - D$, and two genomes with ANI values $\geq 95\%$ were considered to represent the same species.

Positive-sense ssRNA phage contigs were predicted from the pooled assembled metatranscriptome reads for each soil sample [27]. The resulting contigs were de-duplicated at 100% global identity using `CD-HIT` to identify 187,588 non-redundant ssRNA phage contigs, representing 16,541 ssRNA phage vOTUs (containing three core genes in any order). 11,222 vOTUs were assumed to represent near-complete genomes, given the presence of three full-length core genes [27].

Characterization of viral populations

All dsDNA vOTUs were annotated with `Prokka v1.14.6` [69] using the Prokaryotic Virus Remote Homologous Groups (PHROGs) database [70] and the metagenome flag. Additional annotations were provided with `eggNOG-mapper v2` [71, 72] with default settings. Genes putatively involved in metabolism were identified by clusters of orthologous groups (COGs): C, E, F, G, H, I, P, and Q.

Taxonomic assessment of vOTUs was achieved with `vConTACT2 v0.9.13` using “`-rel-mode` `Diamond`,” “`-vcs-mode` `ClusterONE`,” and a custom phage genome database (May 2020) [67] with all other settings set to default. The resultant genome network was visualized in R v4.0.5 using `ggnet2` from `GGally v2.1.2` [73] and the Fruchterman-Reingold force-directed algorithm. vOTUs were assigned into viral clusters (VCs) when clustering was significant ($p < 0.05$) and classified as outliers to the VC when clustering was non-significant. All unclustered vOTUs were classified as singletons.

ssRNA phages were classified into orders and families based on core protein isoforms [74], while genera and species were estimated using previously established RdRp gene clustering thresholds [27]. Phylogenetic assessment was performed on the concatenated core protein sequences aligned with `MAFFT v7.271` [75]. Phylogenetic trees were constructed with `FastTree v2.1.8` [76] using default settings and visualized in R using `ggtree v2.5.3` [77–79].

Putative temperate phages were identified using previously described methods [80, 81]. Briefly, this identified temperate vOTUs encoding a protein associated with lysogeny (transposase, integrase, excisionase, resolvase, and recombinase) by searching for the Pfam domains: PF07508, PF00589, PF01609, PF03184, PF02914, PF01797, PF04986, PF00665, PF07825, PF00239, PF13009, PF16795, PF01526, PF03400, PF01610, PF03050, PF04693, PF07592, PF12762, PF13359, PF13586, PF13610, PF13612, PF13701, PF13737, PF13751, PF13808, PF13843, and PF13358. Additionally, vOTUs clustered with a known temperate phage during `vConTACT2` clustering or representing proviral sequences were assigned as temperate. Non-temperate vOTUs were assigned as lytic. Bacterial hosts were predicted using `WIsH v1.0` [82] and a null model trained against 9620 bacterial genomes as previously described [80]. Host predictions were filtered for $p < 0.05$ and were presented at the genus level.

vOTU abundance and viral gene activity

DNA vOTU abundance was estimated by mapping DNA virome and total metagenome reads against a database of viral genomes (including non-redundant dsDNA

vOTUs recovered in this study and all complete phage genomes in the custom phage database) using BMAP within BBTools [83] with “minid = 0.9.” vOTUs were only considered present in a sample if $\geq 75\%$ of the contig length was covered $\geq 1\times$ by reads, as recommended [65, 84]. Given that the DNA virome and total metagenome libraries were only constructed in March 2017, DNA viral community compositions were only investigated at the stem extension growth stage. For detection of ssRNA phage vOTUs in RNA libraries, we used the above method and thresholds with the additional flag “ambig = random.” ssRNA phage community compositions were compared across seedling, stem extension, and pre-harvest growth stages. For the detection of DNA vOTU gene transcripts in RNA libraries, BAM files were sorted and indexed with SAMtools v1.10 [85]. BEDtools v2.26.0 [86] was used to extract read counts for each gene loci. Resulting read counts were filtered for ≥ 4 gene reads mapped across the replicates of each soil sample, with those < 4 converted to zero in each sample replicate. DNA vOTUs were identified as active when metatranscriptome reads mapped to ≥ 1 gene per 10 kb of the genome, as others have used previously [16].

Data analysis and visualization

All statistical analyses were conducted using R v4.0.5 [87]. Relative vOTU abundance values (counts per kilobase million, CPM) were computed by normalizing read counts by genome length and library sequencing depth. The median of CPM values derived from the DNA virome and total metagenome libraries was computed to generate one abundance value per vOTU per sample. Viral community alpha diversity was described with Shannon's H index computed on vOTU CPM profiles with phyloseq v1.34.0 [88]. Viral community beta diversity was described by computing a Bray-Curtis dissimilarity matrix from square root-transformed vOTU CPM profiles using vegan v2.5-7 [89] and subsequently visualized with non-metric multidimensional scaling (NMDS) ordination using vegan. Similarly, relative gene abundance values (transcripts per kilobase million, TPM) were computed by normalizing read counts by gene length and library sequencing depth. Beta diversity in viral community activity was described in the same way as viral community composition. Two-way analysis of variance (ANOVA) tests and Tukey's honestly significant differences (HSDs) were computed with stats v4.0.5. Permutational multivariate analysis of variance (PERMANOVA) tests and Mantel tests using Pearson's product-moment correlation were performed on Bray-Curtis dissimilarity matrices using vegan. Linear mixed effect models were implemented using lmerTest v3.1-3 [90]. Differential abundance analysis was performed on raw

read counts with DESeq2 v1.30.1 [91]. Plots were generated with ggplot2 v3.3.3 [92].

Results

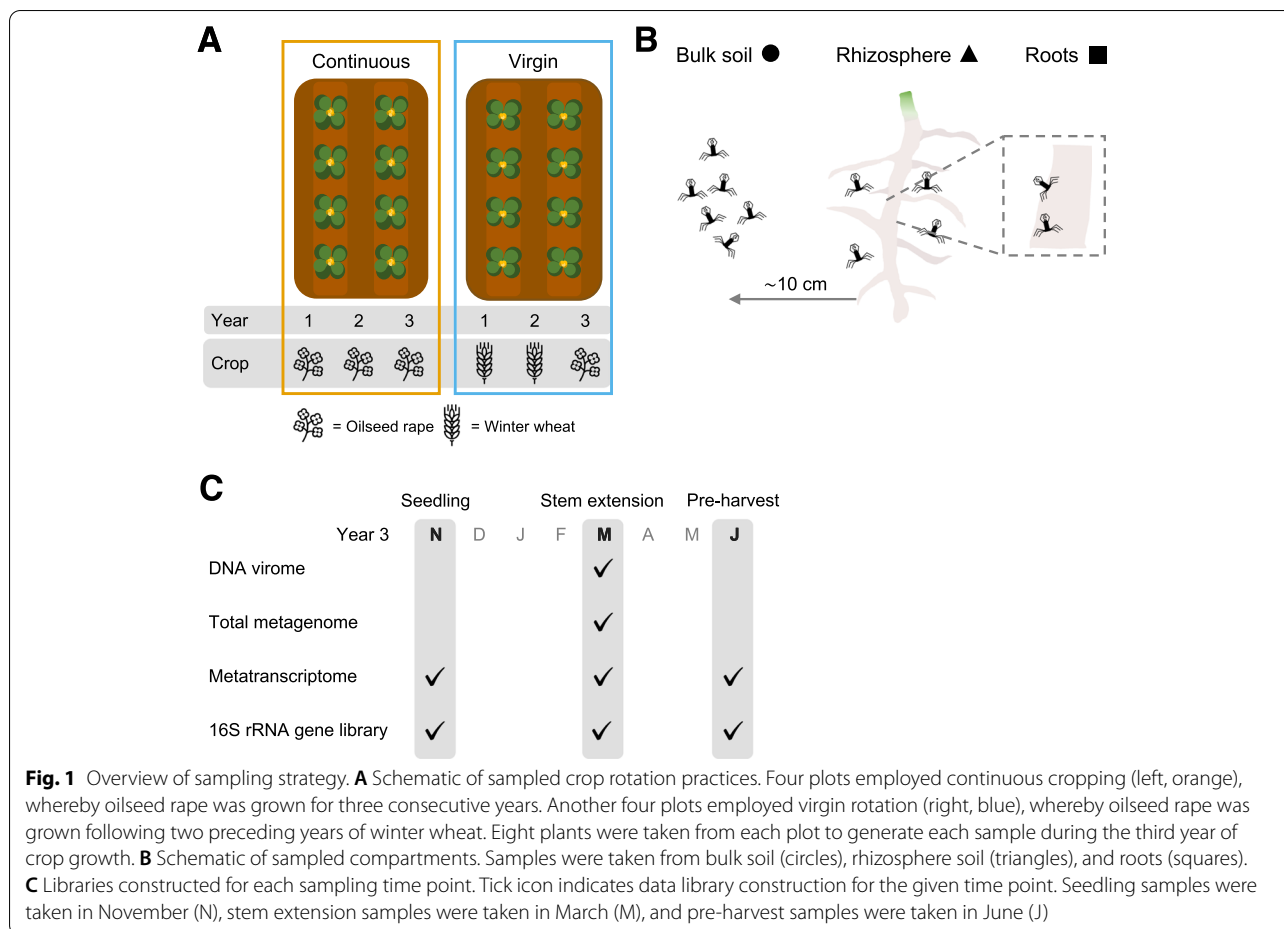
Significant expansion of plant root-associated viruses identified from field-grown oilseed rape

To determine viral community composition across root/soil compartments and crop rotation practices, we recovered vOTUs from samples outlined in Fig. 1. A total of 1059 non-redundant dsDNA vOTUs were recovered from the size-fractionated metagenome (DNA virome), total metagenome, and metatranscriptome libraries, with only one vOTU belonging to a previously isolated phage species (Table S2). The reconstruction of viral sequences from the metatranscriptome yielded 521 (49.8% of total) dsDNA vOTUs, which were not assembled from either of the DNA libraries (i.e., DNA virome and total metagenome). Additionally, a total of 16,541 non-redundant ssRNA phage vOTUs were recovered from the metatranscriptome, with 11,222 of these vOTUs representing near-complete ssRNA phage genomes.

Next, we performed shared protein-based classification to investigate the similarity of recovered vOTUs with all currently available phage genomes, using vCONTACT2 [93]. The resultant network contained viral clusters (VCs) representing roughly genus-level taxonomic groups (Fig. 2A); 262 (24.7% of total) dsDNA vOTUs and 7677 (46.4% of total) ssRNA phage vOTUs formed 95 and 884 VCs, respectively (Table S2; Table S3). However, only 10 of these VCs contained phage genomes that had been previously isolated, demonstrating the undiscovered viral diversity found in this study. The proportion of dsDNA vOTUs forming genus-level VCs was similar across each library used to assemble vOTUs and consistently lower than ssRNA phage vOTUs (Fig. 2B). Using previously established criteria [27], ssRNA phage vOTUs were resolved into 909 genera and 2440 species within the class *Leviviricetes* (Table S3). This included 683 (75.1% of total) new genera and 2379 (97.5% of total) new species, further highlighting the vast novel taxonomic diversity in the ssRNA phage vOTUs.

Novel ssRNA phage diversity was further interrogated by constructing a phylogeny of 11,222 near-complete ssRNA phage vOTUs and all currently available *Leviviricetes* genomes (Fig. S2). 6217 (55.4% of total) near-complete ssRNA phage vOTUs were resolved into 557 new genera, across all five existing *Leviviricetes* families (Table S3). This revealed the extension on existing *Leviviricetes* diversity found in other ecosystems [27–29, 31] and the expansion of the known number of *Leviviricetes* genomes by > 5 times.

To understand the potential ecological roles of soil viruses, we predicted the lifestyles and hosts of recovered



vOTUs. Only 105 (9.9% of total) dsDNA vOTUs were predicted to represent temperate phages, indicating that the majority were likely to be obligately lytic (Table S2). In contrast, it was assumed that none of the ssRNA phage vOTUs were temperate, given that there has been no reported lysogeny among *Leviviricetes* phages. Bacterial hosts were predicted de novo for 518 (48.9% of total) dsDNA vOTUs and 1691 (10.2% of total) ssRNA phage vOTUs using a probabilistic model [82]. The most common host taxonomic class varied depending on the library that the vOTUs were assembled from (Table S2; Table S3); gammaproteobacterial hosts were the most common among DNA virome-assembled vOTUs (61.2% of assigned hosts), actinobacterial hosts were the most common among total metagenome-assembled vOTUs (37.7% of assigned hosts), and betaproteobacterial hosts were the most common among metatranscriptome-assembled vOTUs (68.7% of assigned hosts) (Fig. 2C). A single, uncultured alphaproteobacterial host genus was the most common among ssRNA phage vOTUs (91.5% of assigned hosts) (Fig. 2C). 68/85 (80.0%) of the bacterial genera putatively infected by dsDNA vOTUs were

detected in our soil samples, while only 3/12 (25.0%) of the bacterial genera putatively infected by ssRNA phage vOTUs were detected (Table S4).

The prevalence of the vOTUs recovered in this study was compared by detecting vOTU sequences through read mapping, from the DNA libraries (for dsDNA vOTUs) and metatranscriptomes (for ssRNA vOTUs). Despite sampling not achieving a richness asymptote, 698 DNA vOTUs were detected in at least one sample (Fig. S3A). This included 382/420 (91.0%) DNA virome-assembled vOTUs, 116/129 (89.9%) total metagenome-assembled vOTUs, 215/527 (40.8%) metatranscriptome-assembled vOTUs, and one previously isolated ssDNA phage genome (Table S2). DNA virome-assembled vOTUs were detected in a mean of 1.60 samples and represented 80.9% of all DNA vOTUs present in only one sample. Total metagenome-assembled vOTUs were detected in a mean of 2.83 samples, while metatranscriptome-assembled vOTUs were detected in a mean of 4.31 samples and represented 70.6% of the vOTUs detected in at least half of the samples. As with the dsDNA vOTUs, the sampling of ssRNA

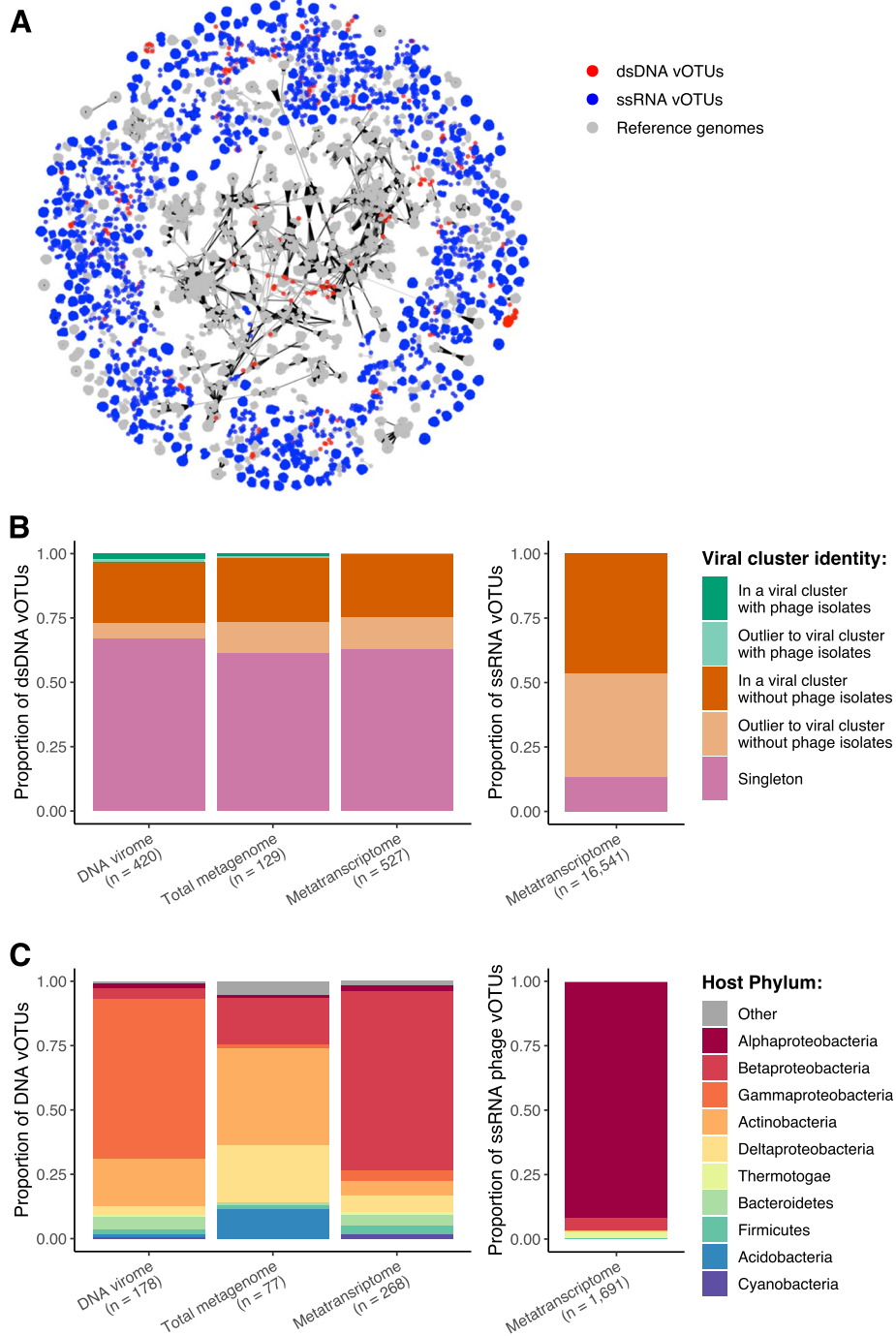


Fig. 2 Taxonomic diversity and predicted hosts for recovered viral populations. **A** Shared protein content of recovered vOTUs with all currently available phage genomes, as determined by vConTACT2. Network graph visualization includes 262 clustered dsDNA vOTUs (red nodes), 7677 clustered ssRNA phage vOTUs (blue nodes), and 12,586 clustered reference DNA and RNA phage genomes (gray nodes). **B** Formation of genus-level viral clusters by recovery library. Relative proportion of vOTUs that formed viral clusters with previously isolated phage genomes (green), without previously isolated phage genomes (orange), and singletons i.e., those that did not form viral clusters (pink) for dsDNA vOTUs (left) and ssRNA phage vOTUs (right). **C** Putative bacterial host phyla of vOTUs by recovery library. Relative proportion of vOTUs predicted to infect bacterial host phyla for dsDNA vOTUs (left) and ssRNA phage vOTUs (right). vOTUs with unknown host genera are excluded. Bar fill color indicates bacterial host phylum for the top ten most common host phyla. Proteobacteria are separated into classes. “Other” represents remaining host phyla

phage vOTUs did not reach a richness asymptote, with 12,162/16,541 (73.5%) vOTUs detected in at least one sample metatranscriptome (Fig. S3B).

By mapping metatranscriptome reads to vOTU gene sequences, we identified 827 (78.1% of total) active dsDNA vOTUs, including 296/420 (70.5%) DNA virome-assembled vOTUs, 104/129 (80.6%) total metagenome-assembled vOTUs, and 444/527 (84.3%) metatranscriptome-assembled vOTUs (Table S5). Additionally, 63 previously isolated dsDNA and ssDNA phage genomes were identified as active in at least one sample. The median relative activity of dsDNA vOTUs assembled from the metatranscriptome was 10.2 and 11.6 times greater than vOTUs assembled from the total metagenome and DNA virome, respectively.

Plant root association and crop rotation shapes both DNA and RNA viral community composition

The alpha diversity of viral communities was compared across root/soil compartments, using the Shannon's H diversity index computed on the relative vOTU abundances for each sample (Fig. 3A). Two-way ANOVA tests were performed, revealing the significant effect of compartment in driving the diversity of both DNA viral communities ($F = 5.116$, $df = 1$, $p = 0.0431$) and ssRNA phage communities ($F = 101.344$, $df = 2$, $p < 0.0001$). For subsequent analyses of ssRNA phage communities across compartments, we chose to exclude the roots given that their very low richness and diversity inflated overall compartmental differences.

Next, we investigated the dissimilarities between viral communities through NMDS ordinations (Fig. 3B). PERMANOVA tests identified that both compartment (8.6–14.7% variance) and crop rotation (10.3–19.2% variance) had significant contributions to the differences in viral community composition at each growth stage (Fig. 3C; Table S6). Despite compartment contributing to > 2 times the variation in co-existing bacterial community composition, the contribution of crop rotation was similar for viruses and bacteria. Subsequently, Mantel tests revealed significant correlations between

bacterial community composition and both DNA viral communities ($r = 0.3301$, $p = 0.0039$) and ssRNA phage communities ($r = 0.4642$, $p = 0.0001$).

Given the vast richness of *Leviviricetes* found in this study, we interrogated compartmental differences further by describing the composition of *Leviviricetes* families across root, rhizosphere soil, and bulk soil compartments (Fig. S4). This indicated a high degree of spatial structuring among ssRNA phage communities, even at the family level, with additional smaller effects of crop rotation and growth stage.

Continuous cropping drives the emergence of distinct, active DNA viruses in seedling rhizospheres

To uncover the potential drivers of viral activity, we first explored the number of active DNA vOTUs detected in metatranscriptomes across root/soil compartments over time (Fig. S5). A two-way ANOVA test was performed, revealing significant effects of compartment ($F = 218.546$, $df = 2$, $p < 0.0001$), crop rotation ($F = 139.185$, $df = 1$, $p < 0.0001$), and growth stage ($F = 508.088$, $df = 2$, $p < 0.0001$) on active vOTU prevalence. It took until stem extension for differences between rotation practices to be observed in the bulk soil and roots, while differences in rhizosphere soil were apparent from the seedling stage (Fig. S5). In fact, there were 196 active vOTUs detected in the seedling rhizosphere under continuous cropping which were absent in the seedling rhizosphere under virgin rotation. The relative activity of these vOTUs increased over time ($F = 51.764$, $df = 2$, $p < 0.0001$), particularly in rhizosphere soil under continuous cropping (Fig. S6).

To investigate the ecological consequences of active vOTUs on their hosts, we trained linear mixed effect models, using compartment as a random effect. This revealed significant linear relationships between active vOTUs and both bacterial host abundance ($b = -0.039$, $p < 0.0001$, Table S7; Fig. S7A) and bacterial community alpha diversity ($b = 0.001$, $p = 0.0147$, Table S8; Fig. S7B).

(See figure on next page.)

Fig. 3 Diversity in viral community composition. **A** Alpha diversity of DNA viral community composition and ssRNA phage community composition. Mean alpha diversity indexes (Shannon's H) for each viral community composition across compartments, at each crop growth stage. Shapes are colored based on field crop rotation strategy: continuous cropping (orange) and virgin rotation (blue). Shapes indicate compartment: bulk soil (circles), rhizosphere soil (triangles), and roots (squares). Error bars denote a 95% confidence interval around the mean. **B** Beta diversity of DNA viral community composition and ssRNA phage community composition. Non-metric multidimensional scaling (NMDS) ordination plots, representing the dissimilarities between community compositions, for each growth stage. Ordinations represent community compositions containing ssRNA phages at seedling ($n = 8125$), ssRNA phages at stem extension ($n = 6936$), DNA viruses at stem extension ($n = 698$), and ssRNA phages at pre-harvest ($n = 10,998$). Shapes are colored based on field crop rotation strategy: continuous cropping (orange) and virgin rotation (blue). Shapes indicate compartment: bulk soil (triangles) and rhizosphere soil (circles). Stress values associated with two-dimensional ordination are reported for each plot. **C** Variation in community composition explained by soil compartment and crop rotation. PERMANOVA results describe the variance in community composition explained by soil compartment and crop rotation, respectively, for each growth stage. Points are colored based on community: ssRNA phage community (blue), DNA viral community (red), and bacterial community (gray)

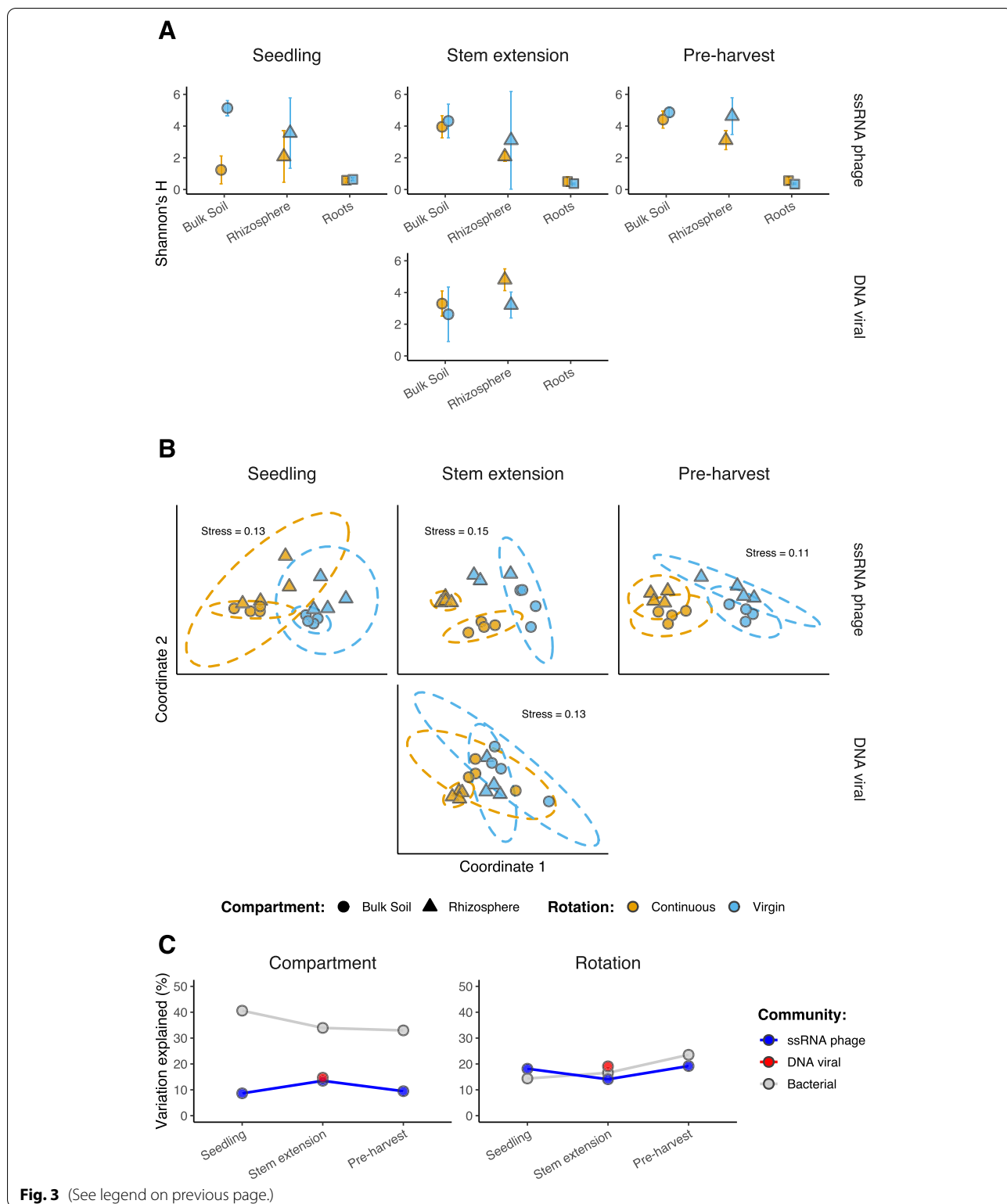
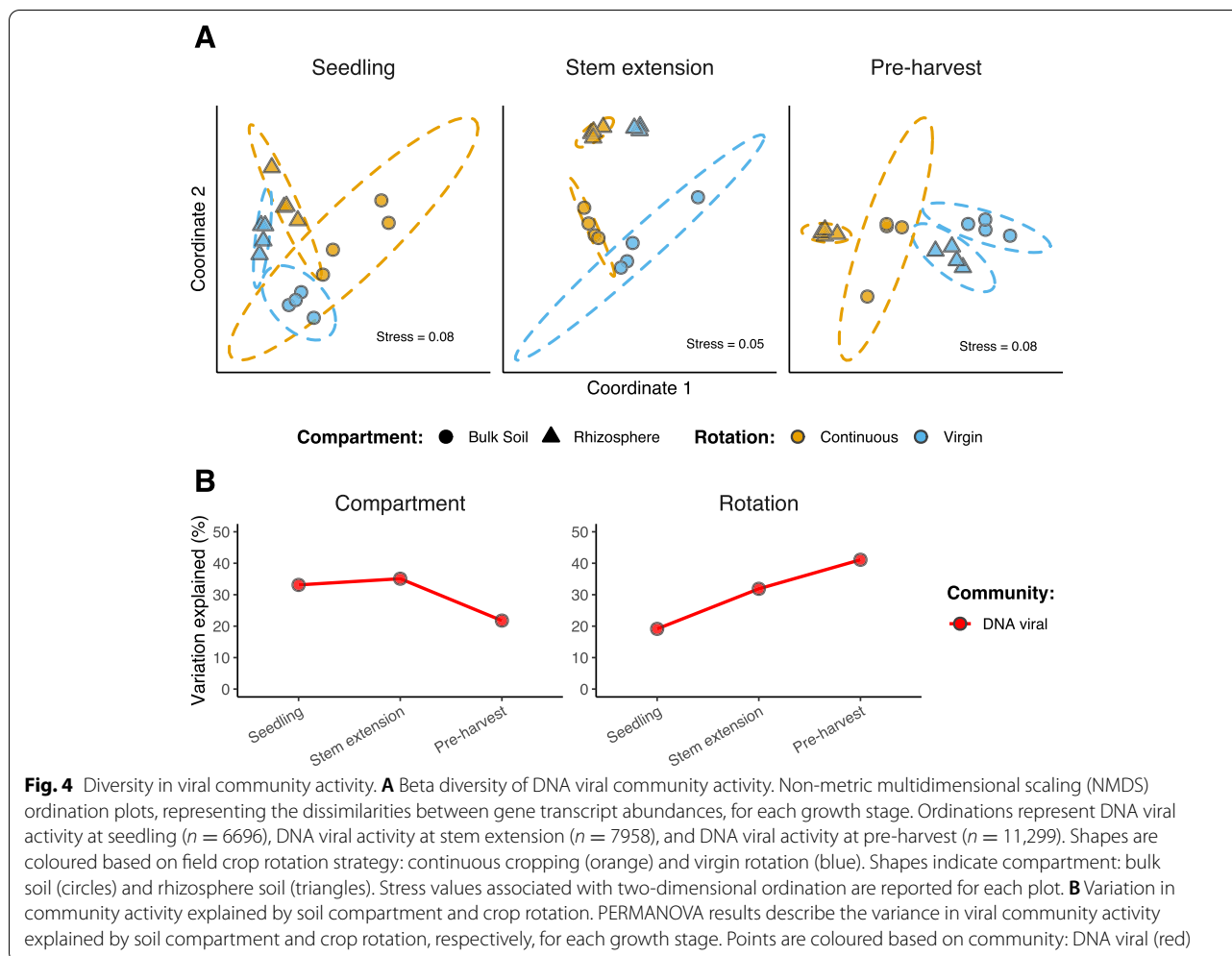


Fig. 3 (See legend on previous page.)

Rhizosphere enrichment of DNA viral activity displays a spatial gradient

The dissimilarity between total viral community activity

at each growth stage was investigated with NMDS ordinations (Fig. 4A). PERMANOVA tests revealed the significant and dynamic contributions of both



compartment (22.5–35.8% variance) and crop rotation (19.1–41.0% variance), such that the effect of crop rotation increased over time, while the effect of compartment decreased (Fig. 4B; Table S9). Comparing these effects on the activity of vOTUs assembled from each library independently revealed that compartmental differences were greatest among metatranscriptome-assembled vOTUs (Fig. S8; Table S10).

To further interrogate compartment-specific viral activity, we first identified differentially active viral genes in either rhizosphere soil or bulk soil. This found ~ 14 times more genes (3589 vs. 250) with significantly greater activity in the bulk soil (relative to rhizosphere soil) than in rhizosphere soil (relative to the bulk soil). We then compared the total activity of these genes across all compartments, revealing that > 78% of viral community activity was soil compartment enriched (Fig. S9). Furthermore, rhizosphere-enrichment of viral activity displayed a spatial gradient, representing

a greater proportion of total community activity at the roots than in rhizosphere soil.

Given that previous research has associated viruses with modulating their hosts' metabolism, we investigated the proportion of viral community activity encoding metabolic functions (Fig. S10). A two-way ANOVA was performed, identifying significant effects of both compartment ($F = 67.682$, $df = 2$, $p < 0.0001$) and growth stage ($F = 11.098$, $df = 2$, $p < 0.0001$) on viral-encoded metabolic activity.

Discussion

Novel, diverse, and active ssRNA phages in plant root-associated ecosystems

In the present study, we have demonstrated that ssRNA phages were both abundant and active across root, rhizosphere soil, and bulk soil compartments. In doing so, we have expanded the number of *Leviviricetes* genomes by > 5 times and identified 683 new genera and 2379 new species (Table S3). The existing phylogeny defined from

a variety of ecosystems remained stable with the addition of the new viral sequences we recovered [27–29, 31, 74] (Fig. S2). During the review of this work, a further 65,000 putative *Leviviricetes* sequences were reported in a preprint from > 1000 metatranscriptomes assembled from soil [94]. It is likely some of these sequences overlap with the diversity obtained within our study. In addition to uncovering novel diversity, we discovered compartmental differences in ssRNA phage communities (Fig. 3C), such that the composition of *Leviviricetes* families varied with proximity to crop roots (Fig. S4). This is the first time that ssRNA phage communities have been investigated at the root surface, and the first evidence of plant roots shaping their community composition. In combination with recent investigations of *Leviviricetes* in terrestrial ecosystems [30, 31], our discovery emphasises the underappreciation of RNA viruses in soils, relative to DNA viruses.

We hypothesize that the trend of *Leviviricetes* across compartment mirrors host populations, given the significant correlation observed between phage and host communities, and that phages require their hosts to replicate. This phenomenon is indicative of predator-prey dynamics, which have been previously demonstrated to link phage and bacterial population abundances in soil crust [95]. Furthermore, the dynamic changes in the relative abundances of ssRNA phages is likely to represent substantial viral reproduction, indicating the active infection of bacterial hosts. ssRNA phages have been established to infect *Pseudomonadota* [96, 97], a highly abundant soil phylum including key participants in the cycling of carbon, nitrogen, and sulphur [98]. Therefore, by driving the turnover of bacteria, the diverse, abundant, and active ssRNA phages recovered in this study are expected to impact terrestrial biogeochemical cycling. Our de novo host predictions suggest that ssRNA phages infect additional bacteria outside of the current *Alphaproteobacteria* and *Gammaproteobacteria* model systems [96, 97] (Fig. 2C). This highlights the underestimated potential of ssRNA phages in the turnover of host populations across root-associated ecosystems. Thus, the development of further model systems, involving cultivation of both phage and its host, is imperative to investigate the impacts of ssRNA phages on rhizosphere ecology.

Viral-host interactions across root-associated microbiomes

Through their detection across root, rhizosphere soil, and bulk soil compartments, we were able to implicate active DNA viruses in shaping co-existing bacterial communities. The majority of the vOTUs recovered in this study were likely to represent lytic dsDNA phages, given almost half of the dsDNA vOTUs had predicted bacterial hosts (Fig. 2C) and the prevalence of lysogeny was low (Table S2). Lytic viral activity was evidenced by a negative

association between active vOTUs and co-existing bacterial host abundance (Figure S7A; Table S7). Furthermore, active vOTU prevalence was positively associated with bacterial community diversity (Figure S7B; Table S8), potentially implicating viral activity in driving bacterial diversity (or vice versa). These are features of the “Kill-the-Winner” hypothesis, which predicts that the population growth of dominant bacterial species is limited by viral lysis [99, 100]. In driving the turnover of host populations, phages are likely to contribute to bacterial community succession and the maintenance of community diversity. Despite there being very limited previous evidence of “Kill-the-Winner” dynamics occurring in soils [95], other predator-prey dynamics, notably bacterial predation by protists and nematodes, has long been associated with changes to microbial community composition [101]. But given that viruses are more selective in their infection of specific host taxa, viral host interactions are likely to be more variable in space and time. Accordingly, the increasing prevalence of active vOTUs observed across the growing season (Fig. S5) suggests that “Kill-the-Winner” dynamics exhibits temporal variation in soil. This implies that the impact of soil viral activity may increase over the crop growth season, as the root-associated microbiome matures.

By comparing the activity of viral communities between soil compartments, we observed that most viral activity was enriched in either rhizosphere soil or bulk soil (Fig. S9). Moreover, we revealed a spatial gradient of viral activity across the root-associated microbiomes, indicating that viruses could have soil niche-specific functions. Viral infection can result in the modulation of microbial host metabolism through the expression of viral-encoded AMGs, as evidenced in marine ecosystems [4–7]. The previous identification of AMGs relating to carbon acquisition and processing has implicated soil viruses in terrestrial carbon and nutrient cycling [16, 39, 42, 43]. Here, we have extended these previous efforts by characterizing the activity of viral-encoded metabolic genes, whose combined relative activity increased with proximity to crop roots (Fig. S10). This parallels the enrichment of microbial activity in the root-associated microbiomes [35]. Thus, soil viruses may contribute to rhizosphere ecology and function through the augmented reprogramming of host metabolism, which could act either antagonistically or synergistically with plants in the control of their root-associated microbiomes [54].

Viral priming in the crop rhizosphere

Previous studies have demonstrated the impacts of crop management practices on bacterial, fungal, and nematode communities in the rhizosphere [49, 50]. By comparing viral communities associated with continuous

cropping and virgin rotation, we have provided the first evidence of crop rotation-driven impacts on soil microbial communities extending to viruses. This is likely, in part, the result of crop rotation impacts on bacterial host communities (Fig. 3C) which were significantly correlated with those of viruses. Nonetheless, differences in viral community composition (Fig. 3C) and activity (Fig. 4B) were sometimes greater between crop rotation practices than between soil compartments. Furthermore, we detected significantly more active vOTUs in the seedling rhizosphere under continuous cropping than in the seedling rhizosphere of the crop grown in virgin rotation (Fig. S5). We propose the principal of “viral priming” to explain this observation; viruses remaining in the soil from the previous growing season were adapted to infect hosts colonizing the juvenile rhizosphere in the current season, only under continuous cropping. Put simply, a greater number of viruses were actively infecting hosts to which they had been previously exposed to. Subsequent differences between rotation practices at the seedling stage were limited to the rhizosphere given the similarity of bacterial communities observed in the bulk soil (Fig. S11). The required persistence of viruses between growth seasons could have been facilitated by (i) the continuous presence of susceptible hosts, (ii) clay particles providing protection from degradation [102], and/or (iii) low soil temperatures preventing viral inactivation [103].

The local adaptation that allows priming viruses to infect hosts colonizing the rhizosphere can be explained by antagonistic coevolution. A study in soil microcosms demonstrated that the fitness cost of phage resistance among bacteria limited their resistance to include only co-occurring phages [104]. Meanwhile, among soil phages, a high level of local adaptation has been shown to result in greater infection rates of co-existing rhizobia strains, as compared to geographically distant strains [105]. Therefore, given the elevated fitness cost of resistance, specific to the rhizosphere [54], newly colonising bacteria are likely to be susceptible to primed viruses. Patterns of phage-bacteria coevolution have previously been observed on the centimetre scale within soils [106], indicating the feasibility for viral priming to occur specifically in the rhizosphere. In contrast, under virgin rotation, viruses remaining in the soil following the harvest of wheat were maladapted to infect the distinct bacterial community that colonized the seedling rhizosphere of the new crop (Fig. S11). Continuous crop growth has been used to explain the accumulation of plant fungal pathogens in rhizosphere soil, which were shown to result in crop yield decline [51]. We speculate that greater viral activity under continuous cropping, due to viral priming, could play a role in regulating both deleterious and beneficial plant-microbe interactions, thus impacting

plant health and yield. Moreover, given that the activity of primed viruses increased across growth stages (Fig. S6), there is likely to be a significant and increasing impact of viral priming on the root-associated microbiomes throughout the growing season. While the net positive or negative consequences of viral priming are yet to be elucidated, we have provided evidence that crop rotation mitigates viral priming activity in the rhizosphere.

Combining metatranscriptomics with metagenomics and viromics to study soil viral communities

We also demonstrate that integrating metatranscriptomics with conventional DNA-based omics approaches mitigates any potential failure to capture ecologically significant viral communities. To describe viral populations, we simultaneously recovered viral genomes from a DNA virome, total metagenome, and metatranscriptome. Remarkably, almost half of the dsDNA vOTUs presented here were assembled from the metatranscriptome alone (Table S2), despite there being no precedent for this recovery method in previous viral ecology studies. Different vOTUs can be recovered between DNA libraries as a result of the library preparation method used, particularly size-filtration to obtain the DNA virome, which has been confirmed to underrepresent viruses with larger capsid sizes [107]. However, this is the first time that DNA viral genomes have been simultaneously recovered from DNA and RNA libraries using the same viral prediction tools and thresholds.

The average prevalence of metatranscriptome-assembled vOTUs was greater than those assembled from DNA libraries, which may have been responsible for our ability to observe greater compartmental differences among these vOTUs (Fig. S8). Previously, total metagenomes have been shown to bias towards the most persistent viruses, capable of infecting the most abundant host organisms [108]. However, many of the highly prevalent metatranscriptome-assembled vOTUs eluded recovery from the total metagenome. Recently, it has become apparent that the hyper-modification of phage DNA prevents the sequencing of certain phage genomes [109, 110]. Subsequently, many phage genomes remain absent in DNA metagenomic samples prepared using transposon-based library methods, as used in this study. Transcriptomics has previously been used to assemble the genome of phage YerA41 from phage-infected cells, thus overcoming the unknown DNA modification that prevented DNA sequencing [111]. The assembly of phage genomes eluding standard DNA sequencing methods, in addition to differences in library sizes, could explain why so many dsDNA vOTUs were exclusively recovered from the metatranscriptome.

Upon further investigation of the vOTUs assembled from each library, we observed consistently high taxonomic novelty (Fig. 2B), but large shifts in the most common bacterial host phyla (Fig. 2C). This indicates possible ecological differences in the viruses accessed by each method, highlighting the value of their combination for describing viral ecology. In addition to its role in reconstructing viral genomes, we implemented metatranscriptomics to detect active vOTUs and characterise viral community activity. Subsequently, we were able to distinguish the ecologically active viral fraction from the “banked” viruses remaining dormant in viral communities at the time of sampling [15, 112]. Furthermore, we have demonstrated that the metatranscriptome accessed the most active viruses, which are vital for investigations of viral ecology, given that viral activity implies the presence and susceptibility of co-existing host organisms. In fact, almost half of the rhizosphere priming viruses were exclusively accessed by the metatranscriptome. This presents the metatranscriptome as a useful, yet underutilized, tool to study soil viral communities.

Conclusions

In summary, we aimed to investigate the influence of crop rotation on the composition and activity of bulk soil, rhizosphere soil, and root viral communities. Combining viromics, metagenomics, and metatranscriptomics, we recovered 1059 dsDNA vOTUs, with almost half of them assembled from the metatranscriptome alone. We also recovered thousands of ssRNA phage vOTUs, including 683 new genera and 2379 new species, and expanding the number of *Leviviricetes* genomes by > 5 times. By describing ssRNA phage communities at the root surface for the first time, we emphasize their underappreciation in soil, as compared to DNA viruses. Furthermore, we revealed spatiotemporal viral activity indicative of “Kill-the-Winner” dynamics, and postulate that viral reprogramming of host metabolism is greater in the rhizosphere than in bulk soil. We also provided the first evidence of crop rotation-driven impacts on soil microbial communities extending to viruses, proposing the novel principal of “viral priming” in the rhizosphere. Our work demonstrates that the roles of soil viruses need greater consideration to exploit the rhizosphere microbiome for food security, food safety, and environmental sustainability. Future studies should continue to investigate soil viral activity with relation to rhizosphere ecology to provide a framework by which we can manage viral communities within agricultural ecosystems. Critically, viruses should be universally included in plant microbiome studies, particularly where these microbiomes have implications for agricultural productivity.

Abbreviations

AMG: Auxiliary metabolic gene; ANI: Average nucleotide identity; ANOVA: Analysis of variance; bp: Base pairs; CPM: Counts per million; dsDNA: Double-stranded DNA; HSD: Honestly significant differences; kb: Kilobases; NMDS: Non-metric multi-dimensional scaling; OTU: Operational taxonomic unit; PERMANOVA: Permutational multivariate analysis of variance; ssRNA: Single-stranded RNA; VC: Viral cluster; vOTU: Viral operational taxonomic unit.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-022-01371-3>.

Additional file 1: Supplementary Figure S1. Overview of field site. **Supplementary Figure S2.** Phylogenetic assessment of ssRNA phage vOTUs. **Supplementary Figure S3.** Accumulation curves of vOTUs. **Supplementary Figure S4.** Summed mean relative abundance of *Leviviricetes* families. **Supplementary Figure S5.** Detection of active vOTUs. **Supplementary Figure S6.** Summed mean relative activity of rhizosphere-priming vOTUs. **Supplementary Figure S7.** Linear relationships between the number of active vOTUs 3 detected. **Supplementary Figure S8.** Variation in dsDNA vOTU activity explained by crop rotation 4 and soil compartment by recovery library. **Supplementary Figure S9.** Summed mean relative compartment-enriched viral activity. **Supplementary Figure S10:** Summed mean relative viral metabolic activity. **Supplementary Figure S11.** Beta diversity in bacterial community composition.

Additional file 2: Supplementary Table S1. Sample metadata for all sample libraries. **Supplementary Table S2.** DNA viral population read counts. **Supplementary Table S3.** ssRNA viral population read counts. **Supplementary Table S4.** 16S rRNA OTU read counts. **Supplementary Table S5.** Viral gene read counts. **Supplementary Table S6.** PERMANOVA testing of contribution of soil compartment, crop rotation strategy and growth stage on viral and bacterial community composition. **Supplementary Table S7.** Mixed effect model output for linear relationship between the number of active vOTUs detected and relative host abundance. **Supplementary Table S8.** Mixed effect model output for linear relationship between the number of active vOTUs detected and bacterial community alpha diversity. **Supplementary Table S9.** Mixed effect model output for linear relationship between the number of active vOTUs detected and bacterial community alpha diversity. **Supplementary Table S10.** PERMANOVA testing of contribution of soil compartment and crop rotation strategy on viral community activity of viral fractions.

Acknowledgements

We acknowledge the use of MRC-CLIMB for the provision of high-performance servers, without which this work would not be possible. We would also like to thank Howard Hilton for kindly providing the photograph used in Figure S1.

Authors' contributions

E.M.H.W., C.Q., A.M., G.D.B., and E.J. conceived and designed the experiment. G.T. and G.D.B. designed and managed the field experiment. S.H. and I.D.E.A.L. collected the samples and extracted the nucleic acids. G.M. and S.R. performed read processing and preparation. G.M. carried out the bioinformatic analyses, generated the R scripts, interpreted the data, prepared the figures, and produced the first draft of the manuscript. I.D.E.A.L., A.M., G.D.B., and E.J. provided edits and additional contributions to the manuscript. The authors read and approved the final submitted manuscript.

Funding

G.M. was funded by the EPSRC & BBSRC Centre for Doctoral Training in Synthetic Biology grant EP/L016494/1. A.M. was funded by MRC grants MR/L015080/1 and MR/T030062/1. G.B. was funded by BBSRC grant BB/L025892/1. E.J. was funded by Warwick Integrative Synthetic Biology (WISB), supported jointly by BBSRC & EPSRC, grant BB/M017982/1.

Availability of data and materials

Post-QC reads are available from the European Nucleotide Archive (ENA) under the Study Accession PRJEB49313. Sample Accession information is

included in Table S1. dsDNA vOTU genome sequences were deposited to ENA under Sample Accession SAMEA12363644. ssRNA phage vOTU genome sequences were deposited to ENA under Sample Accession SAMEA11777518. FASTA nucleotide files containing vOTU genomes, FASTA amino acid files containing vOTU genes, vOTU gene annotations, FASTA amino acid files containing ssRNA phage core protein sequences, Newick tree file containing ssRNA phage phylogeny, and vCONTACT2 network input and output files are available from figshare (<https://figshare.com/s/db3360ed8a3c25b4f9ae>). The custom R script used to generate the figures and tables, along with required input files, are available from GitHub (<https://github.com/GeorgeMuscott/RhizosphereVirome>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Life Sciences, University of Warwick, Coventry, UK. ²Earlham Institute, Norwich Research Park, Norwich, UK. ³Plants, Photosynthesis and Soil, School of Biosciences, University of Sheffield, Sheffield, UK. ⁴Department of Genetics and Genome Biology, University of Leicester, Leicester, UK. ⁵School of Natural Sciences, Bangor University, Bangor, UK.

Received: 22 April 2022 Accepted: 18 August 2022

Published online: 24 October 2022

References

- FAOstat. The state of knowledge of soil biodiversity. 2020.
- Williamson KE, Radosevich M, Wommack KE. Abundance and diversity of viruses in six Delaware soils. *Appl Environ Microbiol*. 2005;71:3119–25.
- Williamson KE, Fuhrmann JJ, Wommack KE, Radosevich M. Viruses in soil ecosystems: an unknown quantity within an unexplored territory. *Annu Rev Virol*. 2017;4:201–19.
- Aylward FO, Boeuf D, Mende DR, Wood-Charlson EM, Vislova A, Eppley JM, et al. Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proc Natl Acad Sci*. 2017;114:11446–51.
- Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, et al. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature*. 2007;449:83–6.
- Puxty RJ, Evans DJ, Millard AD, Scanlan DJ. Energy limitation of cyanophage development: implications for marine carbon cycling. *ISME J*. 2018;12:1273–86.
- Zeng Q, Chisholm SW. Marine viruses exploit their host's two-component regulatory system in response to resource limitation. *Curr Biol*. 2012;22:124–8.
- Suttle CA. Marine viruses — major players in the global ecosystem. *Nat Rev Microbiol*. 2007;5:801–12.
- Breitbart M, Bonnain C, Malki K, Sawaya NA. Phage puppet masters of the marine microbial realm. *Nat Microbiol*. 2018;3:754–66.
- Jover LF, Effler TC, Buchan A, Wilhelm SW, Weitz JS. The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nat Rev Microbiol*. 2014;12:519–28.
- Bar-On YM, Phillips R, Milo R. The biomass distribution on earth. *Proc Natl Acad Sci*. 2018;115:6506–11.
- Kuz'yakov Y, Mason-Jones K. Viruses in soil: nano-scale undead drivers of microbial life, biogeochemical turnover and ecosystem functions. *Soil Biol Biochem*. 2018;127:305–17.
- Emerson JB. Soil viruses: a new hope. *mSystems*. 2019;4:e00120–19 / msystems/4/3/msys.00120-19.atom.
- Pratama AA, van Elsas JD. The 'neglected' soil virome – potential role and impact. *Trends Microbiol*. 2018;26:649–62.
- Trubl G, Hyman P, Roux S, Abedon ST. Coming-of-age characterization of soil viruses: a user's guide to virus isolation, detection within metagenomes, and viromics. *Soil Syst*. 2020;4:23.
- Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, et al. Host-linked soil viral ecology along a permafrost thaw gradient. *Nat Microbiol*. 2018;3:870–80.
- Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, et al. Marine DNA viral macro- and microdiversity from pole to pole. *Cell*. 2019;177:1109–1123.e14.
- Paez-Espino D, Eloë-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. *Nature*. 2016;536:425–30.
- Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*. 2016;537:689–93.
- Schulz F, Roux S, Paez-Espino D, Jungbluth S, Walsh DA, Denev VJ, et al. Giant virus diversity and host interactions through global metagenomics. *Nature*. 2020;578:432–6.
- Göller PC, Haro-Moreno JM, Rodríguez-Valera F, Loessner MJ, Gómez-Sanz E. Uncovering a hidden diversity: optimized protocols for the extraction of dsDNA bacteriophages from soil. *Microbiome*. 2020;8:17.
- Trubl G, Roux S, Solonenko N, Li Y-F, Bolduc B, Rodríguez-Ramos J, et al. Towards optimized viral metagenomes for double-stranded and single-stranded DNA viruses from challenging soils. *PeerJ*. 2019;7:e7265.
- Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*. 2021;9:37.
- Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome*. 2020;8:90.
- Ren J, Song K, Deng C, Ahlgren NA, Fuhrman JA, Li Y, et al. Identifying viruses from metagenomic data using deep learning. *Quant Biol*. 2020;8:64–77.
- Helbling DE, Ackermann M, Fenner K, Kohler H-PE, Johnson DR. The activity level of a microbial community function can be predicted from its metatranscriptome. *ISME J*. 2012;6:902–4.
- Callanan J, Stockdale SR, Shkoporov A, Draper LA, Ross RP, Hill C. Expansion of known ssRNA phage genomes: from tens to over a thousand. *Sci Adv*. 2020;6:eayy5981.
- Krishnamurthy SR, Janowski AB, Zhao G, Barouch D, Wang D. Hyperexpansion of RNA bacteriophage diversity. *PLoS Biol*. 2016;14:e1002409.
- Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X, et al. Redefining the invertebrate RNA virosphere. *Nature*. 2016;540:539–43.
- Hillary LS, Adriaenssens EM, Jones DL, McDonald JE. RNA-viromics reveals diverse communities of soil RNA viruses with the potential to affect grassland ecosystems across multiple trophic levels. *ISME Commun*. 2022;2:34.
- Starr EP, Nuccio EE, Pett-Ridge J, Banfield JF, Firestone MK. Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. *Proc Natl Acad Sci*. 2019;116:25900–8.
- Stough JMA, Kolton M, Kostka JE, Weston DJ, Pelletier DA, Wilhelm SW. Diversity of active viral infections within the *Sphagnum* microbiome. *Appl Environ Microbiol*. 2018;84:16.
- Wu R, Davison MR, Gao Y, Nicora CD, McDermott JE, Burnum-Johnson KE, et al. Moisture modulates soil reservoirs of active DNA and RNA viruses. *Commun Biol*. 2021;4:992.
- el Haichar FZ, Heulin T, Guyonnet JP, Achouak W. Stable isotope probing of carbon flow in the plant holobiont. *Curr Opin Biotechnol*. 2016;41:9–13.
- Reinhold-Hurek B, Büniger W, Burbano CS, Sabale M, Hurek T. Roots shaping their microbiome: global hotspots for microbial activity. *Annu Rev Phytopathol*. 2015;53:403–24.
- Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev Microbiol*. 2017;15:579–90.
- Lambers H, Mougel C, Jaillard B, Hinsinger P. Plant-microbe-soil interactions in the rhizosphere: an evolutionary perspective. *Plant Soil*. 2009;321:83–115.

38. Mendes R, Garbeva P, Raaijmakers JM. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol Rev.* 2013;37:634–63.
39. Bi L, Yu D, Du S, Zhang L, Zhang L, Wu C, et al. Diversity and potential biogeochemical impacts of viruses in bulk and rhizosphere soils. *Environ Microbiol.* 2021;23:588–99.
40. Jin M, Guo X, Zhang R, Qu W, Gao B, Zeng R. Diversities and potential biogeochemical impacts of mangrove soil viruses. *Microbiome.* 2019;7:58.
41. Starr EP, Shi S, Blazewicz SJ, Koch BJ, Probst AJ, Hungate BA, et al. Stable-isotope-informed, genome-resolved metagenomics uncovers potential cross-kingdom interactions in rhizosphere soil. *mSphere.* 2021;6:e00085–21.
42. ter Horst AM, Santos-Medellín C, Sorensen JW, Zinke LA, Wilson RM, Johnston ER, et al. Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global viral populations. *Microbiome.* 2021;9:233.
43. Trubl G, Jang HB, Roux S, Emerson JB, Solonenko N, Vik DR, et al. Soil viruses are underexplored players in ecosystem carbon processing. *mSystems.* 2018;3:e00076–18.
44. Hunter PJ, Teakle GR, Bending GD. Root traits and microbial community interactions in relation to phosphorus availability and acquisition, with particular reference to *Brassica*. *Front. Plant Sci.* 2014;5:27.
45. Picot E, Hale CC, Hilton S, Teakle G, Schäfer H, Huang Y-J, et al. Contrasting responses of rhizosphere bacterial, fungal, protist, and nematode communities to nitrogen fertilization and crop genotype in field grown oilseed rape (*Brassica napus*). *Front Sustain Food Syst.* 2021;5:613269.
46. Bennett AJ, Bending GD, Chandler D, Hilton S, Mills P. Meeting the demand for crop production: the challenge of yield decline in crops grown in short rotations. *Biol Rev.* 2012;87:52–71.
47. Karlen DL, Varvel GE, Bullock DG, Cruse RM. Crop rotations for the 21st century. In: *Advances in Agronomy*: Elsevier; 1994. p. 1–45.
48. Edwards JA, Santos-Medellín CM, Liechty ZS, Nguyen B, Lurie E, Eason S, et al. Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice. *PLoS Biol.* 2018;16:e2003862.
49. Hilton S, Picot E, Schreiter S, Bass D, Norman K, Oliver AE, et al. Identification of microbial signatures linked to oilseed rape yield decline at the landscape scale. *Microbiome.* 2021;9:19.
50. Hilton S, Bennett AJ, Chandler D, Mills P, Bending GD. Preceding crop and seasonal effects influence fungal, bacterial and nematode diversity in wheat and oilseed rape rhizosphere and soil. *Appl Soil Ecol.* 2018;126:34–46.
51. Hilton S, Bennett AJ, Keane G, Bending GD, Chandler D, Stobart R, et al. Impact of shortened crop rotation of oilseed rape on soil and rhizosphere microbial diversity in relation to yield decline. *PLoS One.* 2013;8:e59859.
52. Bohannan BJM, Lenski RE. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecol Lett.* 2000;3:362–77.
53. Koskella B, Brockhurst MA. Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev.* 2014;38:916–31.
54. Koskella B, Taylor TB. Multifaceted impacts of bacteriophages in the plant microbiome. *Annu Rev Phytopathol.* 2018;56:361–80.
55. Whitfield WAD. The soils of the National Vegetable Research Station, Wellesbourne. In: *Report of the National Vegetable Research Station for 1973; 1974*. p. 21–30.
56. Rihlman B, Meaden S, Clokie MRJ, Koskella B, Millard AD. Assessing Illumina technology for the high-throughput sequencing of bacteriophage genomes. *PeerJ.* 2016;4:e2055.
57. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–20.
58. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics.* 2012;28:3211–7.
59. Krueger F, James F, Ewels P, Afyounian E, Schuster-Boeckler B. FelixKrueger/TrimGalore: v0.6.7 - DOI via zenodo: Zenodo; 2021. <https://zenodo.org/record/5127899#.YtluC0RpQI>.
60. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. maSPAdes: a de novo transcriptome assembler and its application to RNA-seq data. *GigaScience.* 2019;8:giz100.
61. Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, et al. Assembling genomes and mini-metagenomes from highly chimeric reads. In: Deng M, Jiang R, Sun F, Zhang X, editors. *Research in Computational Molecular Biology*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 158–70.
62. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods.* 2016;102:3–11.
63. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct De Bruijn graph. *Bioinformatics.* 2015;31:1674–6.
64. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012;28:3150–2.
65. Roux S, Emerson JB, Eloe-Fadrosch EA, Sullivan MB. Benchmarking viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ.* 2017;5:e3817.
66. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosch E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol.* 2021;39:578–85.
67. Cook R, Brown N, Redgwell T, Rihlman B, Barnes M, Clokie M, et al. Infrastructure for a PHAge REference database: identification of large-scale biases in the current collection of cultured phage genomes. *PHAGE.* 2021;2:214–23.
68. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 2016;17:132.
69. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30:2068–9.
70. Terzian P, Olo Ndela E, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, et al. PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genom Bioinform.* 2021;3:lqab067.
71. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol.* 2017;34:2115–22.
72. Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, et al. EggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.* 2012;40:D284–9.
73. Schloerke B, Cook D, Larmarange J, Briatte F, Marbach M, Thoen E, et al. GGally: extension to “ggplot2”. 2021.
74. Callanan J, Stockdale SR, Adriaenssens EM, Kuhn JH, Pallen M, Rumnieks J, et al. Rename one class (*Leviviricetes* - formerly *Allasoviricetes*), rename one order (*Norzivirales* - formerly *Levivirales*), create one new order (*Timlovirales*), and expand the class to a total of six families, 420 genera and 883 species. 2020.
75. Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30:3059–66.
76. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One.* 2010;5:e9490.
77. Yu G. Using ggtree to visualize data on tree-like structures. *Curr Protoc Bioinformatics.* 2020;69:e96.
78. Yu G, Lam TT-Y, Zhu H, Guan Y. Two methods for mapping and visualizing associated data on phylogeny using ggtree. *Mol Biol Evol.* 2018;35:3041–3.
79. Yu G, Smith DK, Zhu H, Guan Y, Lam TT. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol.* 2017;8:28–36.
80. Babenko VV, Millard A, Kulikov EE, Spasskaya NN, Letarova MA, Konanov DN, et al. The ecogenomics of dsDNA bacteriophages in feces of stabled and feral horses. *Comput Struct Biotechnol J.* 2020;18:3457–67.
81. Cook R, Hooton S, Trivedi U, King L, Dodd CER, Hobman JL, et al. Hybrid assembly of an agricultural slurry virome reveals a diverse and stable community with the potential to alter the metabolism and virulence of veterinary pathogens. *Microbiome.* 2021;9:65.
82. Galiez C, Siebert M, Enault F, Vincent J, Söding J. WISH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics.* 2017;33:3113–4.
83. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. Berkeley: Lawrence Berkeley National Lab.(LBNL); 2014.
84. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, et al. Minimum information about an uncultivated virus genome (MIUViG). *Nat Biotechnol.* 2019;37:29–37.

85. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
86. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
87. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2021.
88. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8:e61217.
89. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. *vegan*: community ecology package. 2020.
90. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. *J Stat Softw*. 2017;82:1–26.
91. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
92. Wickham H. *ggplot2*: elegant graphics for data analysis. New York: Springer-Verlag; 2016.
93. Bolduc B, Jang HB, Doulcier G, You Z-Q, Roux S, Sullivan MB. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect archaea and bacteria. *PeerJ*. 2017;5:e3243.
94. Neri U, Wolf YI, Roux S, Camargo AP, Lee BD, Kazlauskas D, et al. A five-fold expansion of the global RNA virome reveals multiple new clades of RNA bacteriophages. preprint. *Genomics*. 2022.
95. Van Goethem MW, Swenson TL, Trubl G, Roux S, Northen TR. Characteristics of wetting-induced bacteriophage blooms in biological soil crust. *mBio*. 2019;10:e02287–19/mbio/10/6/mBio.02287-19.atom.
96. Kazaks A, Voronkova T, Rumnieks J, Dishlers A, Tars K. Genome structure of *Caulobacter* phage phiCb5. *J Virol*. 2011;85:4628–31.
97. Klovins J, Overbeek GP, van den Worm SHE, Ackermann H-W, van Duin J. Nucleotide sequence of a ssRNA phage from *Acinetobacter*: kinship to coliphages. *J Gen Virol*. 2002;83:1523–33.
98. Spain AM, Krumholz LR, Elshahed MS. Abundance, composition, diversity and novelty of soil *Proteobacteria*. *ISME J*. 2009;3:992–1000.
99. Thingstad T, Lignell R. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat Microb Ecol*. 1997;13:19–27.
100. Winter C, Bouvier T, Weinbauer MG, Thingstad TF. Trade-offs between competition and defense specialists among unicellular planktonic organisms: the “killing the winner” hypothesis revisited. *Microbiol Mol Biol Rev*. 2010;74:42–57.
101. Trap J, Bonkowski M, Plassard C, Villenave C, Blanchart E. Ecological importance of soil bacterivores for ecosystem functions. *Plant Soil*. 2016;398:1–24.
102. Stotzky G. Influence of soil mineral colloids on metabolic processes, growth, adhesion, and ecology of microbes and viruses. In: Huang PM, Schnitzer M, editors. *Interactions of Soil Minerals with Natural Organics and Microbes*. Madison: Soil Science Society of America; 1986. p. 305–428.
103. Jończyk E, Kłak M, Międzybrodzki R, Górski A. The influence of external factors on bacteriophages—review. *Folia Microbiol (Praha)*. 2011;56:191–200.
104. Gomez P, Buckling A. Bacteria-phage antagonistic coevolution in soil. *Science*. 2011;332:106–9.
105. Van Cauwenbergh J, Santamaría RI, Bustos P, Juárez S, Ducci MA, Figueroa Fleming T, et al. Spatial patterns in phage-*Rhizobium* coevolutionary interactions across regions of common bean domestication. *ISME J*. 2021. <https://doi.org/10.1038/s41396-021-00907-z>.
106. Vos M, Birkett PJ, Birch E, Griffiths RI, Buckling A. Local adaptation of bacteriophages to their bacterial hosts in soil. *Science*. 2009;325:833.
107. Hingamp P, Grimsley N, Acinas SG, Clerissi C, Subirana L, Poulain J, et al. Exploring nucleocytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J*. 2013;7:1678–95.
108. Santos-Medellín C, Zinke LA, ter Horst AM, Gelardi DL, Parikh SJ, Emerson JB. Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. *ISME J*. 2021. <https://doi.org/10.1038/s41396-021-00897-y>.
109. Korn AM, Hillhouse AE, Sun L, Gill JJ. Comparative genomics of three novel jumbo bacteriophages infecting *Staphylococcus aureus*. *J Virol*. 2021;95:e02391–20.
110. Rihtman B, Puxty RJ, Hapeshi A, Lee Y-J, Zhan Y, Michniewski S, et al. A new family of globally distributed lytic roseophages with unusual deoxythymidine to deoxyuridine substitution. *Curr Biol*. 2021;31:3199–3206.e4.
111. Leskinen K, Pajunen MI, Vilanova MVG-R, Kiljunen S, Nelson A, Smith D, et al. YerA41, a *Yersinia ruckeri* bacteriophage: determination of a non-sequencable DNA bacteriophage genome via RNA-sequencing. *Viruses*. 2020;12:620.
112. Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol*. 2005;13:278–84.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

