Eco-evolutionary response of soil viral communities to land use changes

Hu Liao
Institute of Urban Environment
Hu Li
Institute of Urban Environment, Chinese Academy of Sciences
Chen-song Duan
Chinese Academy of Sciences
Xin-Yuan Zhou
Chinese Academy of Sciences
Qiu-Ping Luo
Chinese Academy of Sciences
Xinli An
Institute of Urban Environment, Chinese Academy of Sciences
Yong-guan Zhu
Institute of Urban Environment, Chinese Academy of Sciences
Jian-Qiang Su (jqsu@iue.ac.cn)
Chinese Academy of Sciences

Article

Keywords:

Posted Date: March 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1415921/v1

Version of Record: A version of this preprint was published at Nature Communications on October 12th, 2022. See the published version at https://doi.org/10.1038/s41467-022-33771-2.

Abstract

Viruses are distributed across the globe, but the soil virome remained relatively understudied compared to the aquatic ecosystems. Land use changes are profoundly affecting the diversity of soil microbes that are vital for soil health, but the response of soil viral communities has not been systematically explored. Here we investigated the ecological patterns of soil viral communities across various land use types encompassing forest, agricultural, and urban soil. We recovered 59,626 viral populations (VPs) via size-fractioned viromic approach with additional mitomycin treatment to induce virus release from bacterial fraction. Significantly different profiles of viral communities were observed among land use types rather than geographic locations. Soil pH was one of the major determinants of viral community structures, and was strongly associated with the changes of *in-silico* predicted host ranges of soil VPs. Habitat disturbance and variation of soil moisture potentially contributed to the dynamics of lysogenic VPs. Microdiversity analysis implied diverse origins of the ubiquitous VPs across land use types, while, selection act on soil VPs, resulting in strong speciation of soil viral communities upon land use types. These findings provide mechanistic understandings of the ecology and evolution of soil viral communities and their associated ecosystem services in changing environments.

Introduction

Soils are intrinsically diverse partially due to their wide compositional spectrum and spatial heterogeneity in terms of physicochemical properties[1], supporting a high diversity of interacting microbes that are pivotal in ecosystem services including global C and N biogeochemical cycles[2]. Viruses are extremely abundant and diverse biological entities on earth, playing vital roles in affecting soil microbiota and functions[3,4] via regulating microbial community dynamics[5], reprogramming host metabolism during infection through the expression of auxiliary metabolic genes (AMGs) [6,7], and serving as vectors of horizontal gene transfer[3]. Viral metagenomics has greatly expanded the viral ecology of aquatic ecosystems and mammalian guts, yet, knowledge about soil viruses lag behind. There were approximately a hundred of reported soil viromes derived from Arctic permafrost[8], agricultural soils[9,10], mangrove sediment[4], peat[11] and Antarctic soils[12], providing insights into the niche partitioning of viral populations[11], their response to climate change and environmental stress[8], and phage mediated horizontal gene transfer[3]. However, none of these viromes was from forest, urban park and road verge soil, the macrodiversity and microdiverisity of viral communities in soils with varied land use types have not been fully explored.

Human activities increasingly change the earth's landscape. It was estimated that nearly a third of global land area were affected by land use change in sixty years[13]. Rapid urbanization often involves changes in land use types, including deforestation, arable farming, and conversion of natural habitat to urban land uses. The habitat disturbance and land use change are profoundly affecting soil microbiome due to the change of vegetation, loss of biodiversity, and input of man-made chemicals. Many studies have documented impacts of land use changes on soil microbial communities, functional traits and soil

functions[14–16]. Since bacteria are the host of phages that dominate soil virome[17], it is expected that the distribution of soil viral communities could considerably differ across various land use types.

Viruses are classified as lytic or lysogenic according to their strategies of life[18]. There were conflicting studies about the proportion of lysogenic and lytic viruses across different habitats[19]. Some studies showed that coinfection would lead to frequent lysogenization, especially in highly productive environments due to higher microbial density and higher phage adsorption rate[20,21]. While other studies observed that lysogenic phages were at lower abundance in response to the eutrophic conditions [22]. These results suggest that host associated and abiotic factors strongly influence the switch between lytic and lysogenic strategies[23]. The changes of environmental factors will also exert selection pressures on soil viral populations (VPs), resulting in variant microdiversity. Recent investigations into the viral microdiversity have provided insights into the evolutionary trends of viral populations in response to geographic distance and depth in marine environments[24,25], while microdiversity of soil VPs has not been explored.

Anthropogenic uses of land have converted and fragmented ecosystems, decreased biodiversity[26,27], disturbed soil biogeochemical cycles [28], and caused soil pollution, leading to significant changes in soil biotic and abiotic environmental factors. However, the ecology and evolution of soil viral communities and interactions with their host in response to such stressors induced by land use changes remained largely unknown. To address these knowledge gaps, we conducted an in-depth characterization of soil viral composition and spatial distribution across five land use types including forest, paddy field, vegetable field, urban park, and road verge through viromic analysis. Host-linked interactions, lysogenicity, and microdiversity were further investigated to illustrate the ecological and evolutionary adaptation of soil virome to land use changes.

Materials And Methods

Sample collection

Samples were collected from soil with various types of uses, including forest, paddy field and vegetable field, urban park and road verge in Xiamen, China on Jul 2020, which were further grouped into three land use zones as forest zone (FO), agricultural zone (AG), and urban green space zone (UG)**.** These soils represent the major land use types, and are closely related to anthropogenic activities. Each land use type was consisted of five locations, and five replicate topsoil (0-20 cm) samples were collected from each location (**Table S1 and Figure S1),** resulting in a total of 25 soil samples.

Soil samples were sieved (~2 mm) to remove stones for measure of soil properties. Soil pH (1:2.5 soil:water) was measured using a pH meter (IS126C Icon). Total carbon (TC), total nitrogen (TN), and total sulfur (TS) were determined with combustion method using a Vario MAX (elementar analysensysteme GmbH, Germany), and 49 metal elements were determined using an X-ray photoelectron spectrometry (Axis Supra, Kratos/Shimadzu). Soil moisture content was determined

according to the method described by Emerson *et al*[8]. The compositions of soil were detected using a laser particle size analyzer, the major soil components were sand and silt (Table S1). The concentration of NH_4^+-N, NO_3^--N, and NO_2^--N were determined using an AA3 analyzer after extraction with 10-fold (weight/volume) of 2M $CaCl_2$. The concentration of total organic carbon (TOC) (combustion method) and dissolved organic carbon (DOC) (0.5M K_2SO_4 extracts) were detected using TOC-L CPH (Shimadzu, Japan).

Soil DNA and viral DNA extraction

Soil viral DNA was extracted in accordance to a previous study[29] with some modification. Briefly, soil sample (300 g) was extracted with PBS (900 ml) by shaking at 200 rpm 24 °C for 45 min, the supernatant was obtained by centrifugation at 3000×g for 15 min at 4 ℃, which was sequentially filtered through 5.0, 0.45, 0.22 μm cellulose membrane. The filtrate was used for the extraction of extracellular virus-like particles (eVLPs). To obtain intracellular VLPs (iVLPs), the membranes with microbes were transferred to 100 mL of PBS with 1 μg/mL of mitomycin C for induced release of iVLPs from the host. The mixture was filtered using a 0.22 μm cellulose membrane after incubation overnight (> 8 h) for 150 rpm in dark at 30 ℃[30] (**Figure S1**). The filtrates were concentrated to ~ 250 μL using 100 kDa Amicon Ultra centrifugal filter units (Millipore), which was treated with 20U DNase I (50 min). VLPs concentrates were filtered using a sterile 0.22 μm Millex-GP filter (Millipore) before viral DNA extraction using a TIANamp Virus DNA/RNA Kit (TIANGEN DP315, Beijing, China).

To extract total soil DNA for nontargeted metagenome sequencing, approximately 0.25 g soil was extracted with PowerSoil DNA isolation kit (Mo Bio Laboratories, Inc. Carlsbad, CA) according to the manufacturer's instructions. A NanoDrop ND-2000 spectrophotometer (NanoDrop, Wilmington, DE, USA) was used to determine the concentration and quality (A260/A280) of extracted DNA.

Library construction, sequencing, and reads processing

Sequencing libraries were prepared using the ALFA-SEQ DNA Library Prep kit (mCHIP, China) following the manufacturer's recommendations and the index codes were added. Paired-end sequencing (150 bp) of total DNA and viral DNA were performed by MagiGene Co. Ltd. (Guangzhou in China) on the Illumina Novaseq 6000 platform respectively. The quality of raw sequences was assessed using FastQC v0.11.5[31]. Clean reads were obtained after quality filtering, trimming, and adaptor removing using cutadpat 2.11 [32]and trimmomatic v0.39[33].

Analysis of metagenomes

The taxonomy of remaining high-quality reads were classified with Kraken2 v2.0.7-beta[34], the package Bracken[35] was used to estimate the relative abundances within a specific sample from Kraken2 classification results. The alpha- (Shannon's index) and beta- (Bray-Curtis dissimilarity) diversity of microbial community structures were performed using vegan[36] in R.

Viral contig assembly, identification, and dereplication

The qualified sequences of each individual viromes were assembled using metaspades v3.13.0[37]. Contigs larger than 1.5kb were piped to VIBRANT v1.2.1[38]. Contigs > 10 kb or circular contigs < 10 kb were obtained, then the ORFs were predicted by prodigal[39] and were searched against the viral protein family (VPFs) dataset[17] containing 25,281 viral protein families using hmmsearch (e-value ≤ 1e-5). Contigs containing more than 3 ORFs with VPF hits were kept for further analysis[40].

Contigs were grouped into a viral population if they shared >95% nucleotide identity with more than 70% coverage of the shorter contigs using Perl script Cluster_genomes_5.1.pl (https://github.com/simroux/ClusterGenomes) based on nucmer v4.0.0beta5[41]. The longest contig was selected as representative genome of the viral population. Subsequently, eVLPs and iVLPs datasets were merged and dereplicated at the population level to construct a total soil viral dataset across all viromes. The final dataset was named as LVD (land use virome dataset). The circular and temperate viral genomes in the LVD were identified through package VIBRANT[38]. The quality of genome was assessed using package VIBRANT and CheckV v0.7.0[42].

Taxonomy assignment and host prediction of VPs

VPs were clustered with viral genomes from NCBI RefSeq Release 201 using package vConTACT2[43], enabling assignment of VPs to a known viral taxa at family or genus level[43]. For VPs that could not be assigned through vConTACT2, family-level taxonomic annotations were conducted using Demovir script (https://github.com/feargalr/Demovir) with default parameters and database. This script performs a search for amino acid sequence homologies between proteins encoded by contigs in question to a viral subset of the TrEMBL database[44], then uses a voting approach to determine taxonomic assignment[45].

Microbial hosts of each viral population was predicted using VirHostMatcher-Net with short-contigs mode[46]. This software provides prediction of the prokaryotic host based on genomes of bacteria and archaea using previously developed CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats) and WIsH models[47,48]. The predicted host with the highest score and accuracy > 90% was selected for further analysis.

Estimation of viral population relative abundances

Since the minimum reads depth was 31M, a threshold of 30M reads was selected and the data was randomly subsampled without replacement across all viromes using Seqtk v1.3 (https://github.com/lh3/seqtk). To estimate the relative abundance of VPs, virome reads were mapped to representative genomes of LVD dataset using bowtie2[49] with parameter –very-sensitive. The reads with nucleotide identity <95% and coverage <70% were removed and the remaining reads were used to calculate the RPKM (Reads Per Kilobase per Million mapped reads) value of each VP using package CoverM v0.5.0 (https://github.com/wwood/CoverM). For the Macrodiversity calculations, the RPKM

values of each viral population was normalized by total RPKM value per virome, which was used as a proxy for relative abundance. Each VP can be flagged by different features (e.g. lifestyle, host or VC). The species accumulation curve was calculated through function Specaccum in R and the result indicated that the all VPs of LVD can be detected in subsampled viromes (Figure S2).

Macrodiversity of viromes

The alpha- (Shannon's index) and beta- (Bray-Curtis dissimilarity) diversity statistics of viral communities were performed using vegan in R[36]. The difference among viral communities among land use zones was evaluated using a PERMANOVA test (function "anosim" and "adonis") and the confidence intervals were plotted using function "ordiellipse" at the confidence limits of 95% and 97.5% using the standard deviation method.

The correlations between environmental variables and all the PCoA dimensions were evaluated using mantel test (function mantel; permutations = 9999 and method = "spearman") after scaling (function scale) and calculating their distance matrices (function dist; method "bray" and na.rm = TRUE). The network of co-occurrence among environmental factor, bacterial class and viral abundance summed up according to predicted host at class level was calculated using the Spearman coefficient and visualized by Gephi v0.9.2[50]. The relationship between the factors, pH and longitude, and the host-linked viral relative abundance at class level were assessed using linear model in R, respectively.

Proportion of shared VPs between samples

To identify shared VPs between zones or within a zone, the viral population abundance table was transformed into a binary presence-absence matrix in R, where a relative abundance of 0.0001 was used as a threshold to determine the presence of a viral population. The viral population presence-absence data of paired extracellular and intracellular viromes were merged. The proportion of shared VPs that were present in different samples was calculated through the equation:

Proportion of shared VPs = ((Sn/a) + (Sn/b))/2

Where a represents the numbers of VPs presented in one sample, b represents the number of VPs presented in another samples, and the Sn represented the numbers of shared VPs between the samples a and b. The heatmap of the proportion of shared VPs was constructed in R v.4.0 using the package pheatmap, which was hierarchically clustered using method average. Wilcox.test was used to test the difference of proportion of shared VPs between different zones.

To explore the relative abundance of shared viral clusters (VCs) based on vContact2 among the different zones, we define the "all zone shared" as the VCs presented in all zones; "FO and AG shared" as the VCs only presented in zone FO and AG, "UG and AG shared" as the VCs only presented in zone UG and AG; "FO and UG shared" as the VCs only presented in zone FO and UG; "special" as the VCs exclusively presented in each zone.

Classifying multi-zonal, regional, and local VPs

VPs and bacterial species were evaluated for their distributions across the three zones of land use and plotted using the VennDiagram package in R. With reference to a previous criteria[25], VPs were designated as “multi-zonal” if they were observed in >1 zone of land use, or zone-specific if they were observed in only one zone. Zone-specific VPs were further divided into local (observed only in 1 site) and regional (observed in ≥ 2 sites).

Microdiversity of VPs

VPs with an RPKM > 5 across 70% of their representative genome in at least one sample in the datasets were flagged for subsequent microdiversity analyses. Nucleotide diversity (π) of flagged contigs, pN/pS for each gene of the flagged VPs were calculated using package inStrain v1.5.3 according to Nei *et al*[51]. The inStrain can identify and annotate biallelic and multiallelic SNVs and their frequencies at the positions where phred30 quality-filtered reads differ from the reference sequence and where multiple bases are simultaneously detected at levels above the expected sequencing error rate[52]. Genes were considered under positive selection if pN/pS was > 1.

To assess the microdiversity of VPs per zone, samples were randomly subsampled without replacement. Within each sample, π values of 10, 20, and 30 VPs of each distribution of zones (multi-zonal, regional, and local, respectively) were randomly selected and averaged. Within samples that lacked enough deeply-sequenced VPs, all the VPs were selected and averaged with particular range of land use. Similarly, within each sample, pN/pS value of 10, 20, and 30 genes located on VPs with different distribution of zones (multi-zonal, regional, and local) were randomly selected and averaged respectively. This was repeated 1000x and the average of the all 1000 subsamplings was used as the final microdiversity and pN/pS value for each sample.

Gene annotation

All genes were predicted using prodigal. ORFs were clustered at 95% identity over 70% contig length using CD-HIT v4.6[53] to reduce redundancy. The resultant ORFs were annotated by searching for matches against the InterPro protein signature database using InterProScan v5[54] with parament -appl Pfam.

Availability of data and materials

Reads from Illumina viromes sequencing were submitted to the NCBI under the project PRJNA691683. Reads from Illumina metagenomes sequencing were submitted to the NCBI under the project PRJNA746419. The database LVD have been deposited in figshare (10.6084/m9.figshare.19108391).

Results

Characteristics of LVD dataset and assembled VPs

The land use virome dataset LVD was derived from 2.6 billion paired clean reads of sequences across 50 viromes of 25 samples with five types of land uses. A total of 6,442,065 contigs > 1,500 bp were yielded, then 764,466 (11.8%) putative viral contigs were identified through VIBRANT, in which 27,951 and 48,936 bona fide viral genomes were retained from 25 iVLPs and 25 eVLPs viromes after removing putative false positive viral genomes, respectively. These genomes were clustered into 25,941 and 45,152 VPs for iVLPs and eVLPs viromes, respectively, in which the iVLPs and eVLPs viromes shared 11,467 (19.2%) VPs. Subsequently, they were merged and dereplicated, resulting in 59,626 VPs (**Table S2**) for the following analysis. A total of 8,112 (13.6%) VPs were obtained with at least one complete viral genome, in which the median length of all and circular VPs were 25,183 bp and 45,511 bp, respectively (Figure 1a). In addition, 15.3% (9,172 VPs) lysogenic VPs were detected (Table S2). Furtherly, 4,844 (8.1%) completed genomes, 6,475 (10.8%) high-quality genomes and 15,156 medium-quality genomes (25.4%) were distributed into separate VPs through checkV (Figure 1b), only 133 genomes (0.2%) were identified as not-determined.

To explore the taxonomic affiliation of VPs in family and genus-level, a gene sharing network consist of 59,626 VP genomes from this study and 3,502 reference phage genomes (from NCBI Viral RefSeq version 201) revealed 6,009 VCs comprising of 37,224 VPs, of which 34,417 VPs were from LVD, besides 2,794 singletons (2,653 from LVD dataset), 16,056 outliers (15,833 from LVD) and 8,492 overlaps (8,061 from LVD) were detected (**Table S3**). Of these, only 157 VCs contained genomes from both the RefSeq and LVD dataset (1,864 viral genomes) (**Table S3**). Most of VCs (1,837, 30.4%) included only two members.

At the family-level, most of VPs were classified into *Siphoviridae* (712 by Vcontact2 and 29,671 (50.9%) by Demovir, tailed dsDNA), *Podoviridae* (610 by Vcontact2 and 9,923 (17.6 %) by Demovir, tailed dsDNA), *Myoviridae* (485 by Vcontact2 and 5,445 (9.9%) by Demovir, tailed dsDNA), *Tectiviridae* (50 by Vcontact2 and 10 (0.10%) by Demovir, non-tailed dsDNA) (**Figure 1c**). Besides, the Eukaryotic viruses *Herpesviridae* (159 by Demovir, 0.26%, dsDNA), *Phycodnaviridae* (120 (0.20%) by Demovir, dsDNA); the Virophage Family *Lavidaviridae* (15 (0.03%) by Demovir) were detected as well, but a majority of VPs were unclassified in genus-level.

Viral community structures differ across land use types

Bray-Curtis dissimilarity of viral communities (median 0.9951) showed strong heterogeneity of viral communities among different sample sites (**Figure 2a**). While, the Bray-Curtis dissimilarity (median: 0.5109) between paired viral communities of iVLPs and eVLPs from each site have a significant lower heterogeneity than inter-sites (Wilcox.test, $p < 0.001$) (**Figure 2a**). Similarly, viral communities of paired iVLPs and eVLPs were grouped together for each site and were well separated from the other sites **(anosim.test, R=0.01, $p > 0.05$)** (**Figure 2b**), paired iVLPs and eVLPs from each site were merged for subsequently viral community analysis.

PCoA based on Bray–Curtis dissimilarity of viral communities indicated the viral communities of the 25 samples from five land use types were clustered into three land use zones (**Figure 2b**). We designated these three emergent land use zones as the agricultural zone (AG) including paddy and vegetable plot, urban green spatial zone (UG) including park and road verge, and forest zones (FO) respectively **(anoism.test, R=0.72, *p*=1e-4).** PERMANOVA indicated that the best predictor of β-diversity of viral community composition was the land use zones compare with other environmental parameters, which explained 10.4% of the variance ($p < 0.0001$). The proportion of shared VPs within land use zones was significantly greater than those between zones (Wilcox.test, $p<0.0001$) (**Figure 3a)**. The heatmap indicated that the viral communities were clustered according to the land use zones **(Figure 3b)** except sample T5 rather than geographic distance **(Figure S3)**. Furthermore, at the VCs level, the multi-zonal VCs were dominant and accounted for 24.7% to 45.2% relative abundance (**Figure 3c**), suggesting that it revealed that the strong VPs boundary among different zones of land use presented a lesser pronounced at VCs level, such as the zone FO and AG did not share any VPs, but they shared 34% VCs (**Figure 3d**). We also observed that UG and AG shared the most VCs (47%), followed by UG and FO (42%), and FO and AG (32%) (**Figure 3d**).

The diversity of viral communities of three land use zones were comparable as indicated by Shannon (6.1–7.0) and Simpson (0.9878–0.9977) indexes. The compositions of viral communities were similar, dominated by tailed dsDNA bacteriophages assigned to the family *Siphoviridae*, followed by *Podoviridae*, *Myoviridae*, *Herpesviridae*, and *Tectiviridae* (**Figure S4**). Mantel tests revealed that viral community structures at population-level significantly correlated with 22 environmental parameters, of which the pH played the most important role in driving viral community structure (**Figure 2c**).

The relative abundances of lysogenic phages vary among different land uses

The relative abundance of lysogenic phage in individual viromes was significantly different among land use zones (**Figure 4a**), such as the relative abundance of lysogenic VPs of the AG (from 4.3% to 17.4%) was significantly lower than those in the UG (from 7.5% to 31.0%) and FO (from 24.5% to 47.8%) (**Figure 4a**). Likewise, the alpha-diversity of lysogenic phage showed a significant difference among the three zones (Figure S5). Besides, the random sample analysis in the relative abundance matrix revealed that the mitomycin C treatment facilitate the acquiring of lysogenic VPs (**Figure 4b**). The relative abundances of lysogenic VPs were negative correlated with soil moisture (**Figure 4c**), whereas a positive correlation with DOC was shown (**Figure 4c**).

Host-linked viral community compositions

The variations of prokaryotic community composition of different land uses were characterized based on taxonomic profiles derived from corresponding metagenomes. The results showed that the dominant class was *Actinobacteria*, *Bacilli*, *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, and *Deltaproteobacteria* (**Figure 5a**). There was a significant difference among the five land use types (anoim.test R=0.44, $p<0.001$) (**Figure S5a**), whereas the alpha-diversity were similar (**Figure S5b**).

Next, the relative abundances of VPs were summed up according to their assigned hosts at class level (**Figure 5b, Table S4**). The results indicated obvious variations in community composition according to zones (**Figure 5b**) compared with overall soil bacterial compositions (**Figure 5a**). Overall, the dominant predicted host were *Actinobacteria*, *Alphaproteobacteria*, *Gammaproteobacteria*, *Betaproteobacteria*, and *Acidobacteriia* **(Figure 5b)**. The zone UG occupied the highest relative abundance (from 17.5% to 52.8%, wilcox.test: $p<0.01$) of *Actinobacteria*-linked phages; the AG occupied the highest relative abundance (from 6.8% to 26.9%, wilcox.test: $p<0.05$) of *Betaproteobacteria*-linked phages **(Figure 5b).** We explored the host of 8,910 VPs with significant differences among the three zones (**Table S5**). Phages infected *Gammaproteobacteria* were enriched in zone FO, and the phages infected *Actinobacteria* and *Alphaproteobacteria* were enriched in zone UG, whereas the zone AG enriched phages that infected *Acidobacteriia* and *Betaproteobacteria* **(Figure S6)**.

Mantel tests revealed that the compositions of bacterial communities were significantly related to the environmental factors DC/DN, C/N, DN, NO_3, Se and longitude **(Figure 5c)**. The host-linked viral communities were significantly related to bacterial communities (mantel statistic r=0.44, $p<0.001$), and they were significantly sensitive to the changes of geographic distance (longitude), soil pH, and concentration of metal element Ba, Ni, As, Mo, La, Nd, and Zn (**Figure 5c**). The network based on the Spearman correlation coefficient indicated that pH, longitude, P, Zn, and moisture have a higher connection degree compared to other environmental factors **(Figure S7**). pH was strongly correlated to the phages infecting *Actinobacteriota*, *Acidobacteriia, Gammaproteobacteria*, but not to the abundance of these taxa in soil metagenome (**Figure S8**). The relative abundances of *Actinobacteria*-linked phages were positive correlated with the elevated pH, whereas the relative abundances of *Acidobacteria* and *Gammaproteobacteria*-linked phages were negative correlated with pH (Figure S8).

Shared VPs and microdiversity

The shared VPs based on their distribution within and between land use zones were classified as local, regional, and multi-zonal VPs (**Figure 6a**). In total, 44,534 VPs (74.6%) were classified as “local”, and 12,314 VPs (20.6%) were classified as “regional”. Only 2,771 VPs (4.6%) were present in at least two zones and were defined as “multi-zonal”, of which zone FO and AG only shared 103 VPs, whereas the 1,479 VPs were shared by zone FO and UG. In addition, only 76 VPs were detected in all zones, whereas there were 7,398 bacterial species (64.4% of all) presented in all zones **(Figure 6a)**.

The microdiversity of VPs increased from local, regional to multi-zonal, despite that in FO, significant difference in the microdiversity between regional and multi-zonal VPs was not observed **(Figure 6b).** The microdiversity of multi-zonal VPs in FO were significantly lower than UG and AG (**Figure 6c**).

The connection number of node (viral population) in the gene-sharing network constructed by vContact2 were calculated. We found that the number of connections (median 47.00) of multi-zonal VPs with other VPs was significantly greater than that of regional one (median, 38), and regional one (medium, 38) was significantly greater than that of local (median, 31) (**figure 6d**). Subsequently, 2,771 multi-zonal VPs were

observed in 1,416 VCs. The results showed that the size (median, 10) of VCs included multi-zonal VPs is significantly higher than the total (median, 4), the multi-zonal VPs prefer to exist in the large-sized VCs **(figure 6e)**. These results demonstrated that the multi-zonal VPs shared genetic information with more VPs than local and regional in soil.

We identified genes whether under positive selection by evaluating the ratio of non-synonymous to synonymous mutations observed in gene sequences using the pN/pS equation. Of 726,331 genes tested from populations with sufficient coverage (> 5 mean coverage depth), 39,118 genes were identified as being under positive selection in at least one sample, and most of them were predicated responsible for structure and DNA metabolism (**Table S6**). The genes carried by local VPs yielded higher pN/pS values than those carried by regional and multi-zonal VPs (**Figure 6f**).

Discussion

Here, we conducted a comprehensive viromic examination of viral communities in soils various land use types including forest, vegetable and paddy field, and urban park and road verge. To improve our understanding of soil viral diversity, we extracted VLPs from soil and perfomed metagenomic analysis of viral profiles. This size-fractioned viromic approach was far more prominent in the recovery of VPs than viral sequence mining from total soil metagenome[11]. A large size of soil (300 g) was subjected to viral particles purification and viral DNA extraction, providing sufficient DNA amount for direct viromic library construction and sequencing that could reduce the bias due to the multiple displacement amplification (MDA) of low viral DNA yield. In addition to the extraction of eVLPs, we further adopted mitomycin C treatment of bacterial fraction to induce the release of iVLPs from their hosts. Sequencing and viromic analysis of eVLPs and iVLPs were conducted separately, with 25,941 and 45,152 VPs recovered from iVLPs and eVLPs viromes using a validated de novo virus-discovery approach[45], respectively. The iVLPs and eVLPs viromes shared 11,467 VPs, accounting for only 19.2% of total assembled VPs. We thereby advocate the implement of mitomycin C treatment to promote the recovery of VPs in future studies for comprehensive profiling of soil viral communities. Through these efforts, we obtained a by far the largest soil viral genomic catalog (LVD) consisting of 59,626 VPs, of which 8,112 VPs were recovered with complete genome, expanding approximately 3 fold of currently available complete viral genomes from soil metagenomes[42].

Most of the recovered VPs were classified into dsDNA viruses and bacteriophages. The dominant families across all samples were *Siphoviridae*, *Myoviridae*, *Podoviridae*, and a number of unassigned *Caudovirales*, which was different from Antarctic soils including a fraction of nucleocytoplasmic large DNA virus (NCLDV)[12], mangrove sediment including ssDNA *Microviridae*, whereas consistent with those in permafrost[8]. Despite most of the recovered VPs can be taxonomically assigned to a family using a voting approach[45], over 95% of VPs obtained in our study did not clustered with reference genomes in the gene-sharing network, suggesting that this research cast light on part of the huge dark matter hidden in the soil, whereas more viral diversity remained uncovered.

The constructed LVD facilitated genome enabled analysis of macrodiversity and microdiversity of VPs across soils with various land use types. Soil samples were collected from five types of land uses, which were classified into three ecological zones, i.e. forest zone (FO), agricultural zone (AG) including paddy and vegetable fields, and urban green space zone (UG) including urban park and road verge. The profiles viral communities significantly varied across land use types and ecological zones (Fig. 2b), as well as bacterial community profiles (**Figure S5**). Geographically distinct viral community compositions and structures have been well documented in natural ecosystems from global[55] to a very local scale within a 18m^2 agricultural field[10]. We also compared the VPs in the LVD with those recovered from agricultural soil of Dezhou in our previous study[56], which locate in the North of China with more than 2000 km distance from Xiamen, only 23 VPs were shared between both datasets, indicating strong geographic isolation and existence of endemism for viral communities. However, in this study, the viral communities of soil samples clustered according to the ecological zones with different land use types rather than spatial proximity (Fig. 3b and S3). The samples from the same zones shared considerably more VPs compared to that of among different zones (Fig. 3a). These results indicated a stronger effect of land use types on the speciation of soil viral communities than spatial distance. The significant correlation between viral and bacterial communities indicated shift in bacterial community structures induced by land use changes contributed to the shift of viral communities (Mantel test r = 0.24, p = 0.02). Land use conversion was often accompanied with changes in soil geochemical properties. We further explored the environmental factors associated with the change of viral community, pH exhibited the strongest correlation (Fig. 3c), suggesting the potential role of pH in shaping viral communities[57].

The infection dynamics of recovered VPs for specific host populations grouped at class level was assessed across three land use zones. Viral hosts were predicted via a CRISPR-based approach, a total of 59,570 VPs (99.9%) were assigned to host, providing robust analysis of virus-host interactions. A stronger correlation between the host-linked viral communities and host communities (Mantel test r = 0.39, p < 1e-04) was observed than that between viral communities and total bacterial communities. The major bacterial host were *Actinobacteria*, *Alphaproteobacteria*, *Gammaproteobacteria*, and *Acidobacteria.* Actinophages was the dominant, which has been reported as the largest identifiable group in aquatic ecosystems[58,55] but poorly studied in soil. The compositions of bacterial communities and viral communities were similar across land use zones (Fig. 5a and S3), however, the lineage-specific virus abundance significantly varied among habitats (Fig. 5b), with the abundances of major lineage-specific viruses correlating with geochemical parameters, particularly with soil pH. The abundances of *Actinobacteria*, *Alphaproteobacteria*, *Gammaproteobacteria*, and *Acidobacteria* remained relative stable across pH ranges, while their corresponding lineage-specific virus abundances significantly differed and correlated with pH (**Figure S8**). Soil pH have been demonstrated as the major factor in shaping bacterial community compositions in global topsoil[2], it is expected that pH would also affect the viral communities. Nevertheless, this result imply that the impact of pH on the host ranges of viruses contributed to the structuring of viral communities other than indirect effect through changing host communities.

The relative abundance of lysogenic viruses increased from zone AG to UG to FO (Fig. 4a), indicating that the life strategies of viruses could be determined by land use types. Generally, the agricultural field bear the strongest stress of habitat disturbance, and that of the forest soil was the lowest. Both computational and experimental analysis demonstrated spatial structures select for lower virulence[59], with disturbed environment select lytic lifestyle partially due to the availability of susceptible host cells and creation of empty sites for lytic infection[60,61]. Thus, land use change associated soil disturbance could be responsible for the decline of lysogenic viruses from forest soil to agricultural field. This was further supported by the significantly negative correlation between soil moisture and lysogenic viruses (Fig. 4c), since higher soil moisture is associated with more connectivity and dispersal, resulting in lower spatial structuring, and thus facilitating the movement of extracellular viruses and propagation of lytic viruses. In addition, DOC exhibited a significantly positive effect on the lysogenic virus abundance (Fig. 4c), indicating soil nutrient level is a major determinant for the life strategy of viruses, which has been well documented in previous studies[22].

We next profiled the microdiversity and calculated pN/pS across viral genomes for fully understanding the structures of viral populations and their eco-evolutional process upon land use changes. The recovered VPs were classified as local, regional, and multi-zonal VPs based on their distribution across land use types, with only 4.6% multi-zonal VPs detected. The dispersal potential of VPs was weaker than bacterial species, and the drift of VPs was rare among different habitats (Fig. 6a and 3d), which is consistent with previous studies[62]. The microdiversity of regional and multi-zonal VPs were significantly higher than local VPs (Fig. 6b), suggesting a more diverse pool of viruses for regional and multi-zonal VPs[52]. Similar result was observed for multi-zonal VPs in zones AG and UG (Fig. 6c) that may acquire viruses from more sources. The size of VCs containing multi-zonal VPs were the largest (Fig. 6d, e), further demonstrated that the multi-zonal VPs share genetic information with more VPs than local and regional VPs. Higher microdiversity of multi-zonal VPs may increase the probability of possessing an adaptive genotype as land use changes. Nevertheless, genes of regional and multi-zonal VPs have a significantly lower pN/pS ratio than those of local VPs, indicating that soil viruses are likely under negative selection that could cause the extinction of most drifted viral populations and thus leading to strong niche partition of soil viral communities when land use changes.

In conclusion, this study provides a comprehensive database of high-quality viral genomes for a multitude analysis of soil virome, and a systematic investigation of patterns and drivers of soil viral macro- and microdiverisy. Land use types show a stronger effect on the speciation of soil viral communities than spatial distance, contrasting known biogeographic pattern. In addition, land use changes significantly shape the life strategy, host interaction and microdiverisity of soil viral communities. Shift in the bacterial communities and geochemical factors, in particular, pH, moisture and nutrient level associated with land use changes, are the major determinants. With massive urbanization and changing climate, these advances provide insights into the response of soil viral communities to habitat disturbance and land management. Such understanding of soil viruses, and further interrogation

the linking between viral communities and complex ecosystem processes in soil, are critical for broader inclusion of viruses in ecosystem models in the Anthropocene.

Declarations

Acknowledgement

This study was financially supported by the National Natural Science Foundation of China (42177097, 42021005) and the Alliance of International Science Organizations (ANSO-PA-2020-18).

References

1. Dion, M. B., Oechslin, F. & Moineau, S. Phage diversity, genomics and phylogeny. Nat Rev Microbiol **18**, 125–138, doi:10.1038/s41579-019-0311-5 (2020).
2. Bahram, M. *et al.* Structure and function of the global topsoil microbiome. Nature **560**, 233–237, doi:10.1038/s41586-018-0386-6 (2018).
3. Chevallereau, A., Pons, B. J., van Houte, S. & Westra, E. R. Interactions between bacterial and phage communities in natural environments. Nat Rev Microbiol, doi:10.1038/s41579-021-00602-y (2021).
4. Jin, M. *et al.* Diversities and potential biogeochemical impacts of mangrove soil viruses. Microbiome **7**, 58, doi:10.1186/s40168-019-0675-9 (2019).
5. Martinez Arbas, S. *et al.* Roles of bacteriophages, plasmids and CRISPR immunity in microbial community dynamics revealed using time-series integrated meta-omics. Nat Microbiol **6**, 123–135, doi:10.1038/s41564-020-00794-8 (2021).
6. Mageeney, C. M. *et al.* Mycobacterium Phage Butters-Encoded Proteins Contribute to Host Defense against Viral Attack. *mSystems* **5**, doi:10.1128/mSystems.00534-20 (2020).
7. Moon, K. *et al.* Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. Microbiome **8**, 75, doi:10.1186/s40168-020-00863-4 (2020).
8. Emerson, J. B. *et al.* Host-linked soil viral ecology along a permafrost thaw gradient. Nat Microbiol **3**, 870–880 (2018).
9. Bi, L. *et al.* Diversity and potential biogeochemical impacts of viruses in bulk and rhizosphere soils. Environ Microbiol **23**, 588–599, doi:https://doi.org/10.1111/1462-2920.15010 (2021).
10. Santos-Medellin, C. *et al.* Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. ISME J, doi:10.1038/s41396-021-00897-y (2021).
11. Ter Horst, A. M. *et al.* Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global viral populations. Microbiome **9**, 233, doi:10.1186/s40168-021-01156-0 (2021).
12. Adriaenssens, E. M. *et al.* Environmental drivers of viral community composition in Antarctic soils identified by viromics. Microbiome **5**, 83, doi:10.1186/s40168-017-0301-7 (2017).

13. Winkler, K., Fuchs, R., Rounsevell, M. & Herold, M. Global land use changes are four times greater than previously estimated. Nat Commun **12** (2021).
14. Dong *et al.* Land Use Influences Antibiotic Resistance in the Microbiome of Soil Collembolans Orchesellides sinensis. Environ Sci Technol (2018).
15. Qian, X. *et al.* Metagenomic analysis reveals the shared and distinct features of the soil resistome across tundra, temperate prairie, and tropical ecosystems. Microbiome **9**, 108, doi:10.1186/s40168-021-01047-4 (2021).
16. Schaufler, G. *et al.* Greenhouse gas emissions from European soils under different land use: effects of soil moisture and temperature. Euro J Soil Sci **61**, 683–696, doi:https://doi.org/10.1111/j.1365-2389.2010.01277.x (2010).
17. Paezespino, D. *et al.* Uncovering Earth's virome. Nature **536**, 425–430 (2016).
18. Mavrich, T. N. & Hatfull, G. F. Bacteriophage evolution differs by host, lifestyle and genome. Nat Microbiol **2**, 17112, doi:10.1038/nmicrobiol.2017.112 (2017).
19. Knowles, B. *et al.* Lytic to temperate switching of viral communities. Nature **531**, 466–470, doi:10.1038/nature17193 (2016).
20. Chen, X., Weinbauer, M., Jiao, N. & Zhang, R. Revisiting marine lytic and lysogenic virus-host interactions: Kill-the-Winner and Piggyback-the-Winner. Science Bulletin, doi:10.1016/j.scib.2020.12.014 (2020).
21. Luque, A. & Silveira, C. B. Quantification of Lysogeny Caused by Phage Coinfections in Microbial Communities from Biophysical Principles. *mSystems* **5**, e00353-00320, doi:10.1128/mSystems.00353-20 (2020).
22. Silpe, J. E. & Bassler, B. L. A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. *Cell* **176**, 268–280 e213, doi:10.1016/j.cell.2018.10.059 (2019).
23. Correa, A. M. S. *et al.* Revisiting the rules of life for viruses of microorganisms. Nat Rev Microbiol **19**, 501–513, doi:10.1038/s41579-021-00530-x (2021).
24. Coutinho, F. H., Rosselli, R. & Rodríguez-Valera, F. Trends of Microdiversity Reveal Depth-Dependent Evolutionary Strategies of Viruses in the Mediterranean. mSystems **4**, e00554-00519, doi:10.1128/mSystems.00554-19 (2019).
25. Gregory, A. C. *et al.* Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell* **177**, 1109–1123 e1114, doi:10.1016/j.cell.2019.03.040 (2019).
26. Marques, A. *et al.* Increasing impacts of land use on biodiversity and carbon sequestration driven by population and economic growth. Nat Ecol Evol **3**, 628–637, doi:10.1038/s41559-019-0824-3 (2019).
27. Newbold, T. *et al.* Global effects of land use on local terrestrial biodiversity. Nature **520**, 45–50, doi:10.1038/nature14324 (2015).
28. Hong, C. *et al.* Global and regional drivers of land-use emissions in 1961–2017. Nature **589**, 554–561, doi:10.1038/s41586-020-03138-y (2021).

29. Prestel, E., Salamitou, S. & DuBow, M. S. An examination of the bacteriophages and bacteria of the Namib desert. The Journal of Microbiology **46**, 364, doi:10.1007/s12275-008-0007-4 (2008).
30. Pan, D. *et al.* Correlation between viral production and carbon mineralization under nitrate-reducing conditions in aquifer sediment. ISME J **8**, 1691–1703, doi:10.1038/ismej.2014.38 (2014).
31. Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2014).
32. Martin, M. CUTADAPT removes adapter sequences from high-throughput sequencing reads. EMBnet.journal **17**, doi:10.14806/ej.17.1.200 (2011).
33. Bolger, A. M., Marc, L. & Bjoern, U. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 2114–2120 (2014).
34. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. Genome Biol **20**, 257, doi:10.1186/s13059-019-1891-0 (2019).
35. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. PeerJ Comput Sci **3**, e104, doi:10.7717/peerj-cs.104 (2017).
36. Oksanen, J. *et al.* vegan: Community Ecology Package. CRAN R package. (2015).
37. Nurk, S. *et al.* 158–170 (Springer Berlin Heidelberg).
38. 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 **8**, 90, doi:10.1186/s40168-020-00867-0 (2020).
39. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**, 119–119 (2010).
40. Paez-Espino, D., Pavlopoulos, G. A., Ivanova, N. N. & Kyrpides, N. C. Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data. Nat Protoc **12**, 1673–1682, doi:10.1038/nprot.2017.063 (2017).
41. Gregory, A. C. *et al.* The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. Cell Host Microbe, doi:10.1016/j.chom.2020.08.003 (2020).
42. Nayfach, S. *et al.* CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol, doi:10.1038/s41587-020-00774-7 (2020).
43. Bin Jang, H. *et al.* Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol **37**, 632–639, doi:10.1038/s41587-019-0100-8 (2019).
44. Boeckmann, B. *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucl Acids Res **31**, 365–370, doi:10.1093/nar/gkg095 (2003).
45. Shkoporov, A. N. *et al.* The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. *Cell Host Microbe* **26**, 527–541 e525, doi:10.1016/j.chom.2019.09.009 (2019).
46. Wang, W. *et al.* A network-based integrated framework for predicting virus-prokaryote interactions. NAR Genom Bioinform **2**, lqaa044, doi:10.1093/nargab/lqaa044 (2020).

47. Butina, T. *et al.* Extended Evaluation of Viral Diversity in Lake Baikal through Metagenomics. Microorganisms **9**, 760, doi:10.3390/microorganisms9040760 (2021).
48. Galiez, C., Siebert, M., Enault, F., Vincent, J. & Söding, J. WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. Bioinformatics **33**, 3113–3114, doi:10.1093/bioinformatics/btx383 (2017).
49. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat Methods **9**, 357–359, doi:10.1038/nmeth.1923 (2012).
50. Bastian, M., Heymann, S. & Jacomy, M. *Gephi: An Open Source Software for Exploring and Manipulating Networks*. (2009).
51. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. PNAS **76**, 5269–5273, doi:10.1073/pnas.76.10.5269 (1979).
52. Olm, M. R. *et al.* inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nature Biotechnol, doi:10.1038/s41587-020-00797-0 (2021).
53. Fu, L., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics (Oxford, England) **28**, doi:10.1093/bioinformatics/bts565 (2012).
54. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. Bioinformatics **30**, 1236–1240, doi:10.1093/bioinformatics/btu031 (2014).
55. Sakowski, E. G. *et al.* Interaction dynamics and virus–host range for estuarine actinophages captured by epicPCR. Nat Microbiol, doi:10.1038/s41564-021-00873-4 (2021).
56. Chen, M.-L. *et al.* Viral Community and Virus-Associated Antibiotic Resistance Genes in Soils Amended with Organic Fertilizers. Environ Sci Technol, doi:10.1021/acs.est.1c03847 (2021).
57. Sungeun, L. *et al.* Soil pH influences the structure of virus communities at local and global scales. Soil Biol Biochem **166**, 108569, doi:10.1016/j.soilbio.2022.108569 (2022).
58. Kavagutti, V. S., Andrei, A. S., Mehrshad, M., Salcher, M. M. & Ghai, R. Phage-centric ecological interactions in aquatic ecosystems revealed through ultra-deep metagenomics. Microbiome **7**, 135, doi:10.1186/s40168-019-0752-0 (2019).
59. Lion, S. & Gandon, S. Evolution of spatially structured host–parasite interactions. J Evol Biol **28**, 10 (2015).
60. Berngruber, T., Lion, S. & Gandon, S. Spatial Structure, Transmission Modes and the Evolution of Viral Exploitation Strategies. PLoS pathogens **11**, e1004810, doi:10.1371/journal.ppat.1004810 (2015).
61. Gandon, S. Why Be Temperate: Lessons from Bacteriophage λ. Trends Microbiol **24**, doi:10.1016/j.tim.2016.02.008 (2016).
62. Jian, H. *et al.* Diversity and distribution of viruses inhabiting the deepest ocean on Earth. ISME J, doi:10.1038/s41396-021-00994-y (2021).

Figures

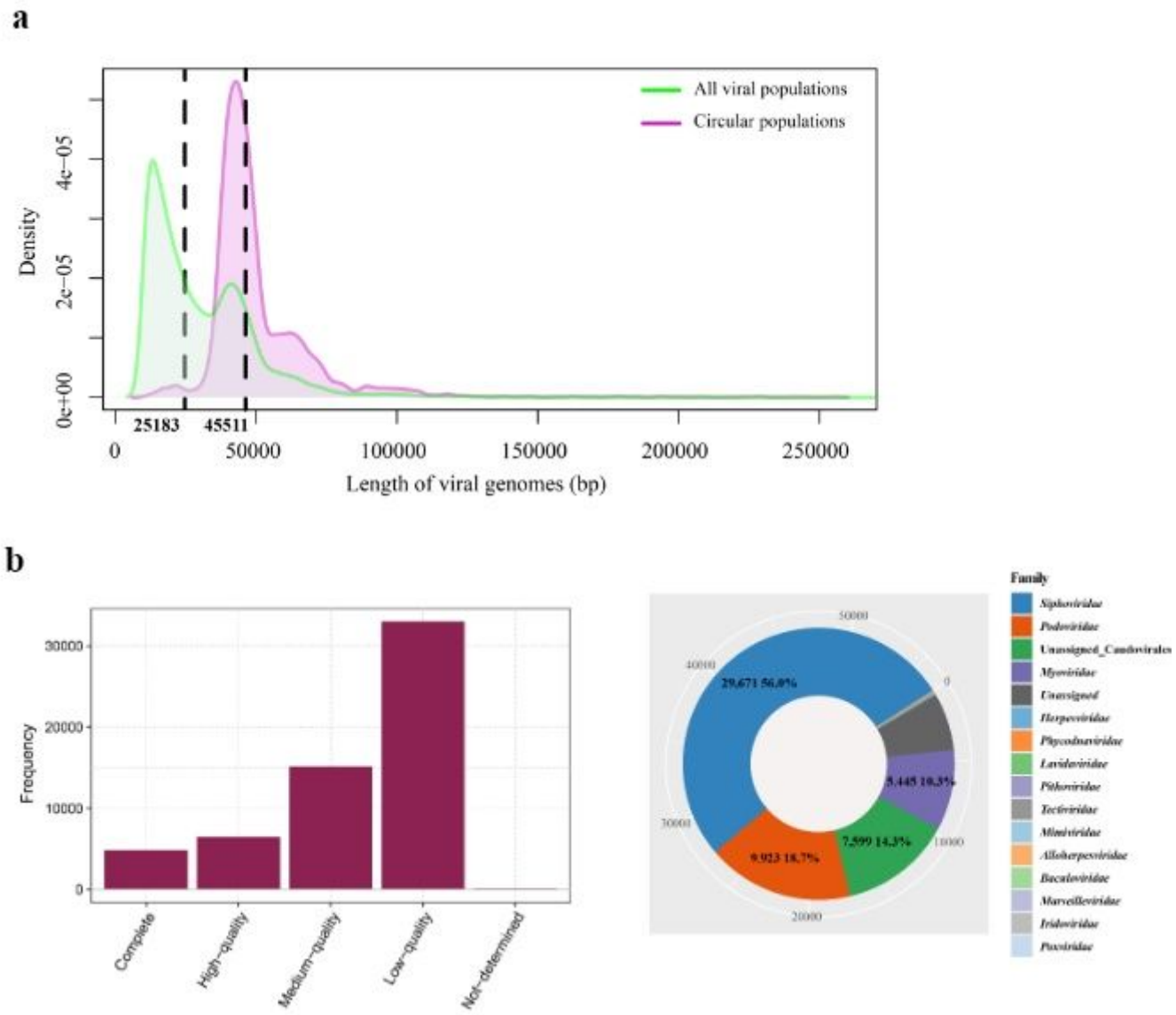

Figure 1

(a) Genomic length distribution of the viral dataset including 59,626 VPs (blue line) and 8,112 VPs with completed circular genomes (purple line) identified by VIBRANT. (b) Quality assessment of 59,626 VPs by checkV analysis. (c) Pie charts showing affiliation of 56,870 VPs at family level assigned by script Demovir.sh.

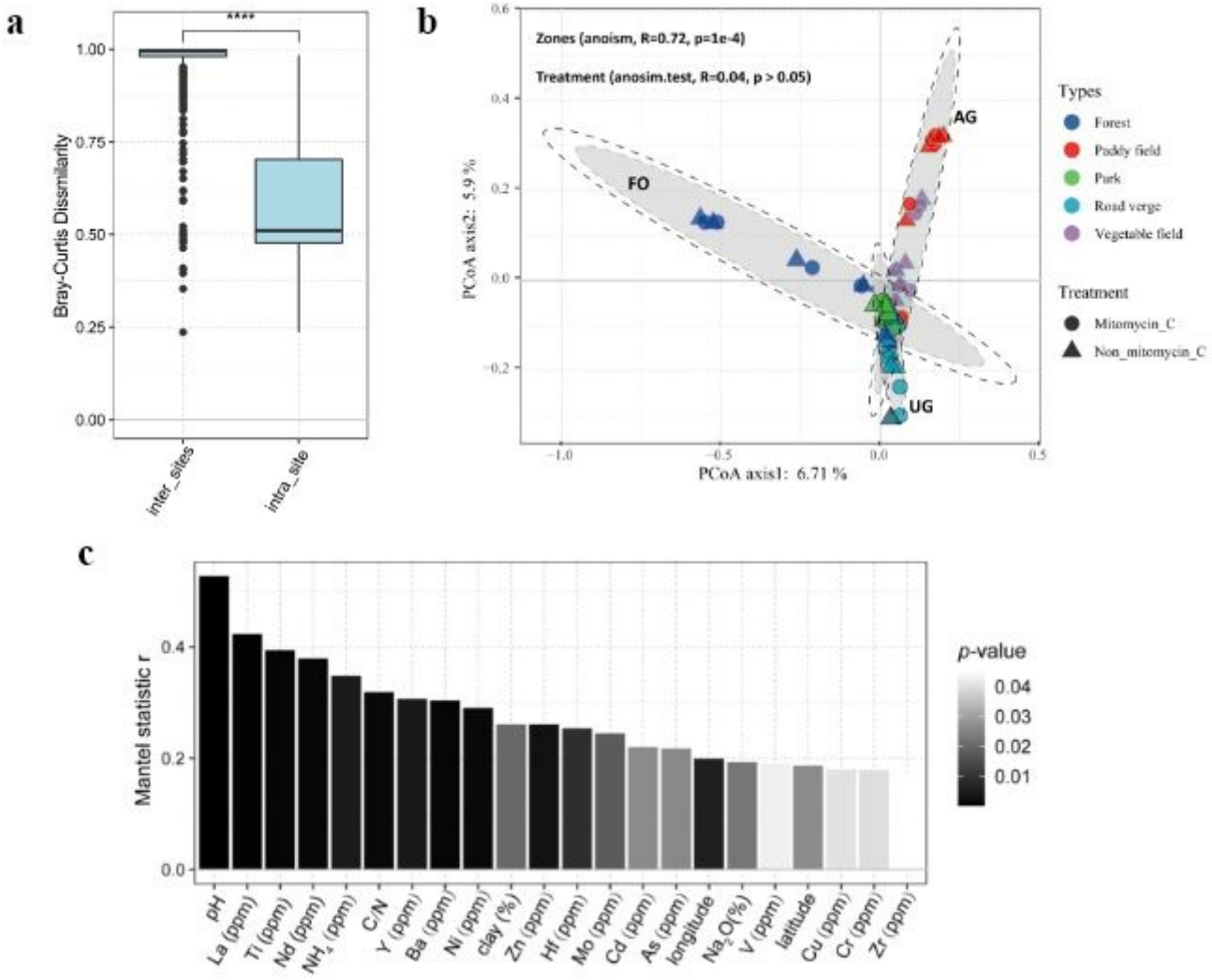

Figure 2

(a) Boxplot showing Bray-Curtis dissimilarity of viral communities of intra-sites (between the corresponding community of iVLPs and eVLPs) and inter-sites (between different sample sites). (b) Principal coordinates analysis (PCoA) of viral community structures, as derived from reads mapping to 59,626 viral populations and Bray–Curtis dissimilarities; each point is one sample, triangle indicated extra-cellular (Non_mitomycin_C) and circles indicated intra-cellular (mitomycin_C treatment) viral community. The analysis of similarity (ANOSIM) statistics considered viral community composition grouped by habitat and treatment. (c) Mantel test showing the correlation between viral communities and geochemical parameters. *** and **** represents $p < 0.001$ and $p < 0.0001$ respectively.

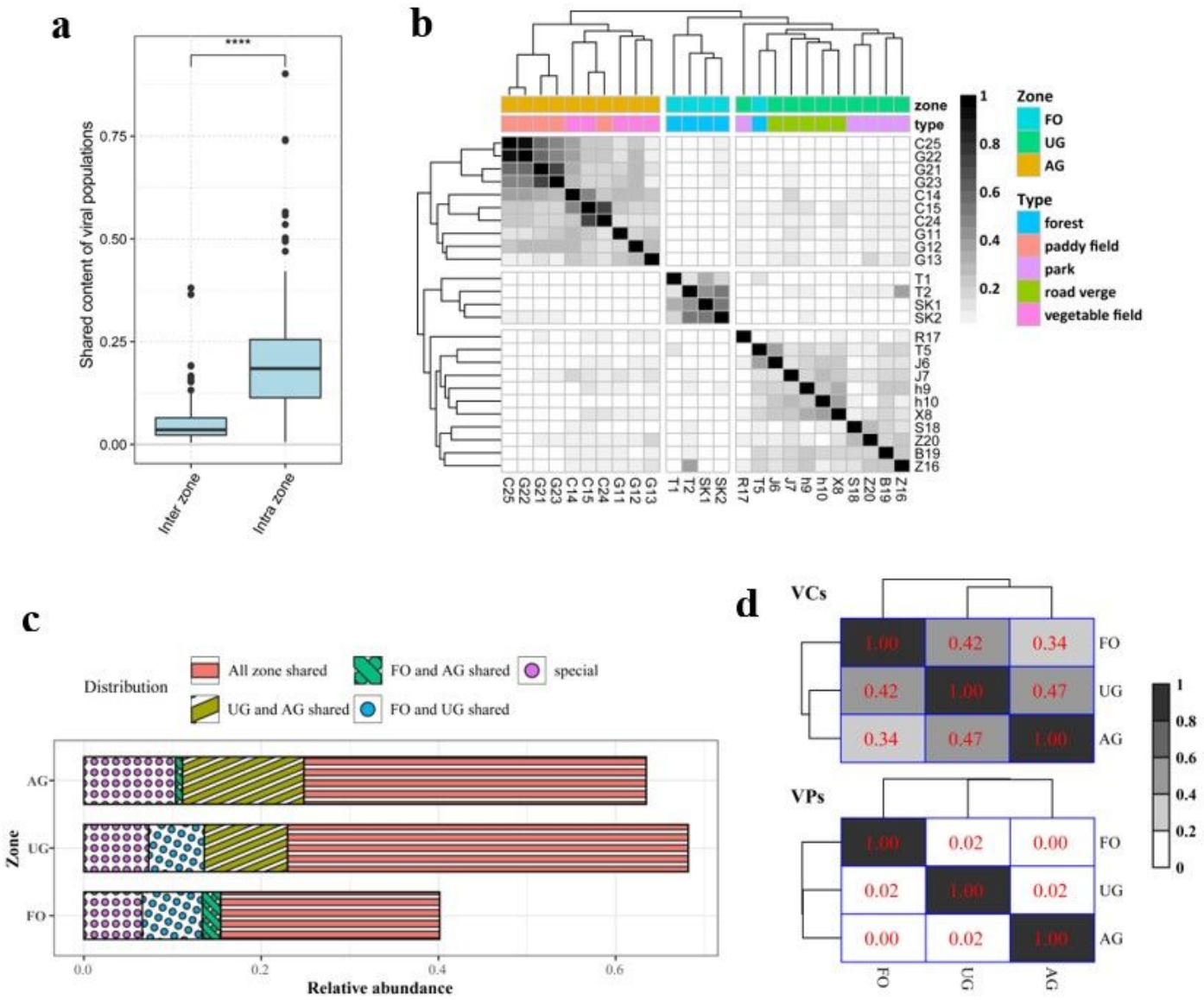

Figure 3

(a) The shared VPs of inter-zones and intra-zones (**** Wilcox test, $p < 0.001$). (b) Heatmap showed the shared viral content between different samples. (c) The relative abundances of VCs according to their distribution. (d) The proportion of shared VCs (above) and VPs (below) between different zones.

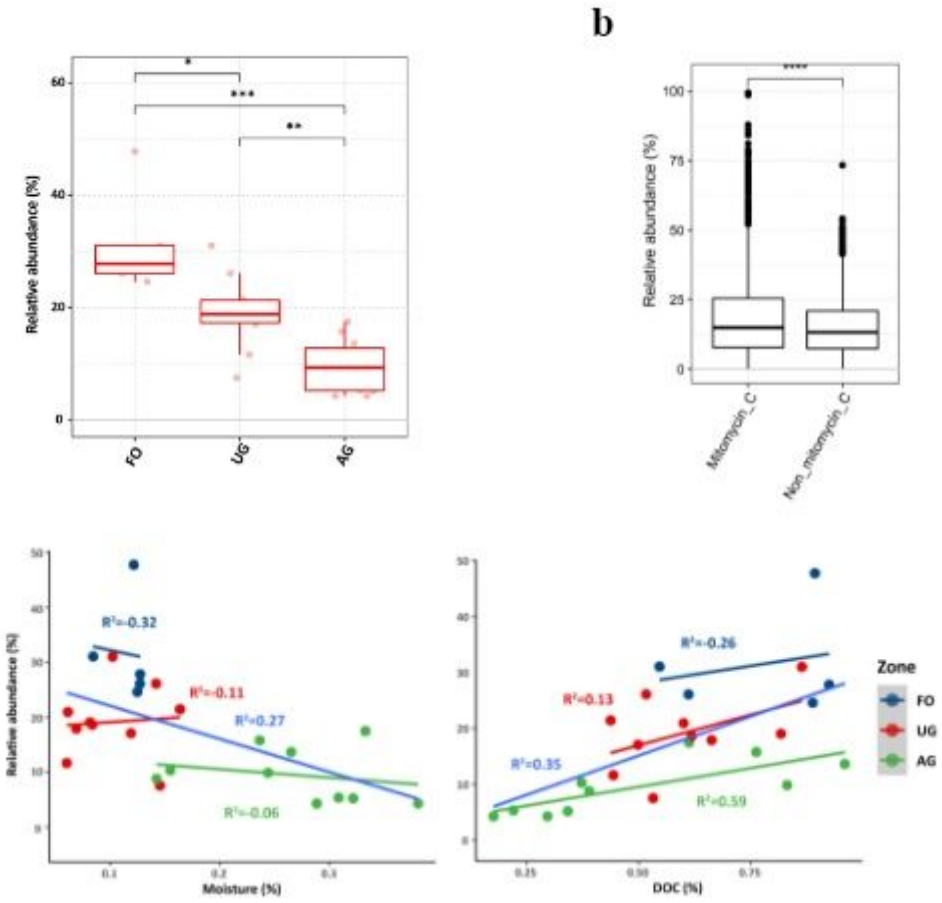

Figure 4

The proportions of lysogenic phage in zones FO, AG, UG (a), and in eVLPs and iVLPs, all samples were subsampled 100 times with a size of 5,000 (b), the median and quartiles were shown the difference between different zones was tested using the Wilcox.test, **** and * represents $p < 0.0001$ and $p < 0.05$ respectively. (c) Relationship between the relative abundance of lysogenic phages and moisture and DOC.

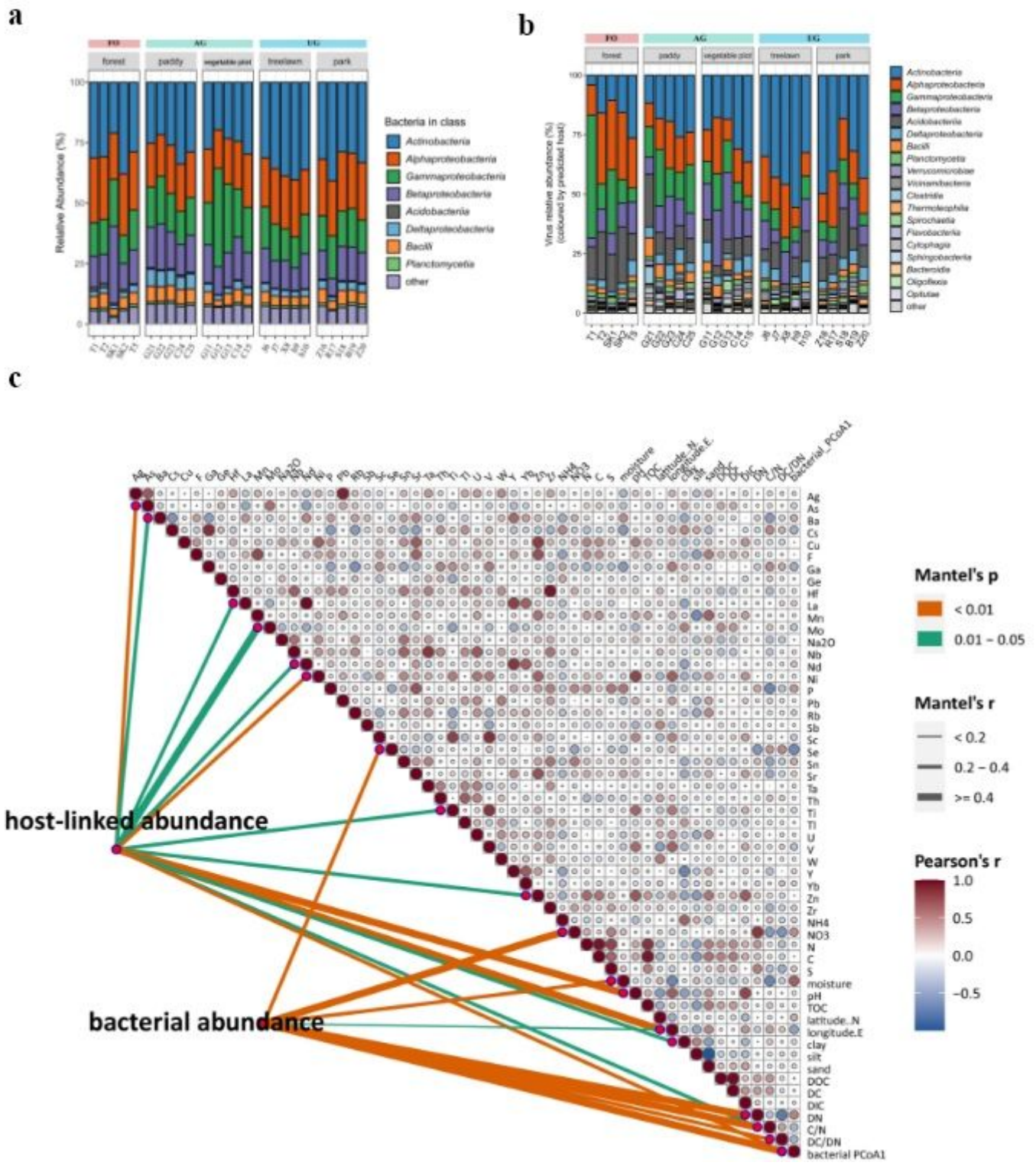

Figure 5

(a) Class distribution of bacterial compositions of metagenomes from various land use types (b) Relative abundances of lineage specific phages grouped at the host class level (c) Mantel test showed the correlation between environmental factors and lineage specific phages (top node, host-linked abundance) and bacterial community (bottom node, bacterial abundance).

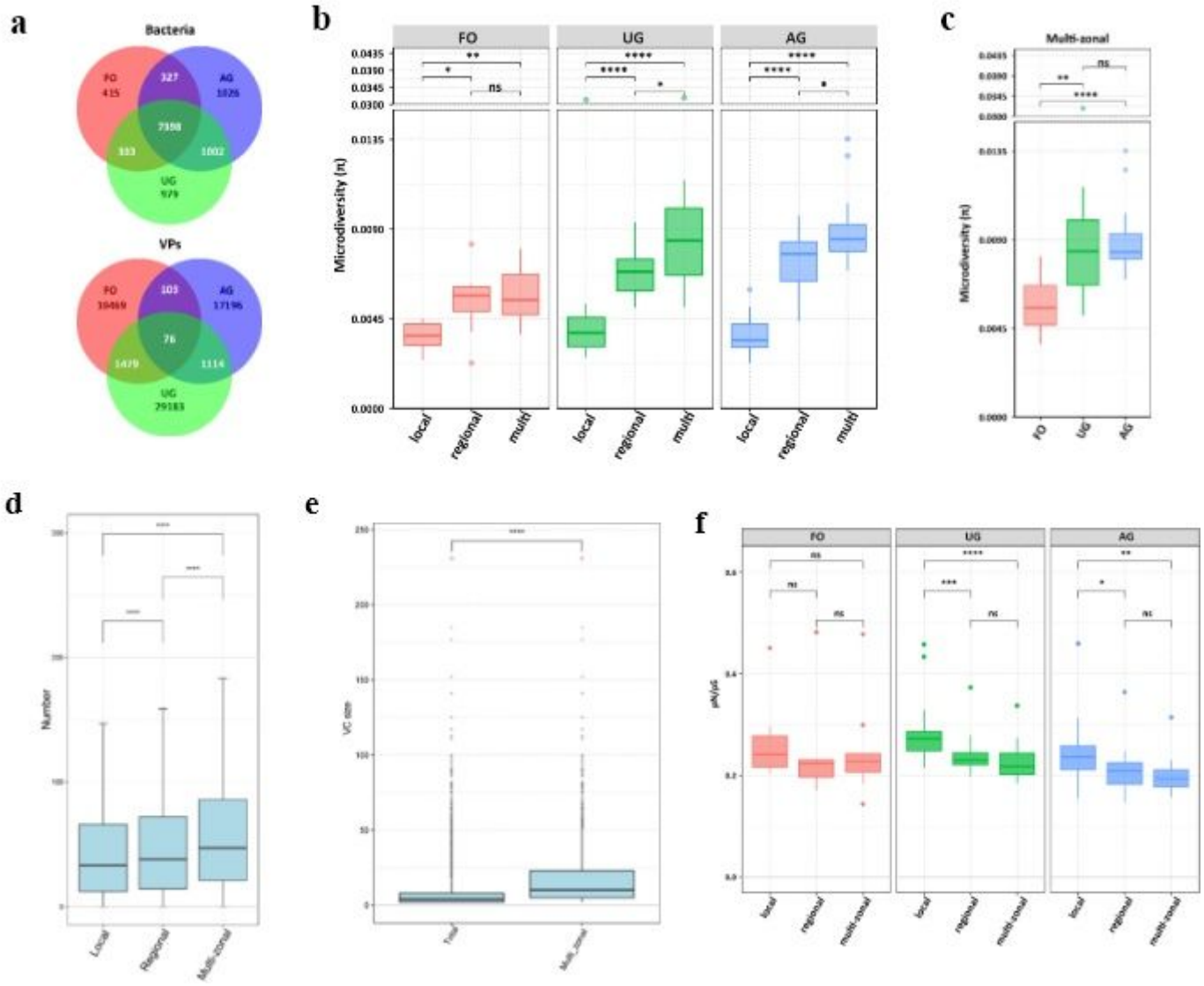

Figure 6

(a) Venn diagram showing the shared VPs (top) and bacterial species (bottom) among three land use zones. (b) The microdiversity of local, regional, and multi-zonal VPs. (c) The microdiversity of multi-zonal VPs among three land use zones. (d) The connection number of local, regional, and multi-zonal VPs in gene-sharing network. (e) The size of total VCs and multi-zonal VCs. (f) pN/pS values of local, regional, and multi-zonal VPs. For boxplots, median and quartiles were shown.

Supplementary Files

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

- Supplementaryfigure.docx
- TableS1env.xlsx
- TableS2populationsinfo.xlsx

- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx
- TableS6.xlsx
- nrreportingsummaryNCOMMS2208495.pdf
- scriptNCOMMS2208495.zip