Soil viral community dynamics over seven years of heat disturbance: spatial variation exceeds temporal in annually sampled soils

Samuel E Barnett, Ashley Shade

Soil viral community dynamics over seven years of heat disturbance: spatial variation exceeds temporal in annually sampled soils

Samuel E. Barnett[a]* and Ashley Shade[b]

[a] Department of Microbiology, Genetics, and Immunology, Michigan State University, East Lansing, MI 48824, USA, barne424@msu.edu

[b] Universite Claude Bernard Lyon 1, CNRS, INRAE, VetAgro Sup, Laboratoire d'Ecologie Microbienne LEM, CNRS UMR5557, INRAE UMR1418, Villeurbanne, F-69100 France, ashley.shade@cnrs.fr

* Corresponding author

Abstract

Viruses are important components of the soil microbiome, influencing microbial population dynamics and the functions of their hosts. However, the relationships and feedbacks between virus dynamics, microbial host dynamics, and environmental disturbance are not understood. Centralia, PA, USA, is the site of an underground coal seam fire that has been burning for over 60 years. As the fire moves along the coal seam, previously heated soils cool to ambient temperature, creating a gradient of heat disturbance intensity and recovery. We examined annual soil viral population dynamics over seven consecutive years in Centralia using untargeted metagenome sequencing. Viral communities changed over time and were distinct between fire-affected and reference sites. Dissimilarity in viral communities was greater across sites (space) than within a site across years (time), and cumulative viral diversity more rapidly stabilized within a site across years than within a year across sites. There also were changes in number of CRISPR arrays per genome as soils cooled, corresponding to shifts in viral diversity. Finally, there were also differences in viral-encoded auxiliary metabolic genes between fire-affected and reference sites. Thus, despite high site-to-site soil viral diversity, there was surprising viral community consistency within a site over the years and shifting host-viral interactions in soils recovering from disturbance. Overall, this work provides insights into the interannual dynamics of soil viruses and their host communities, as well as how they collectively respond to long-term warming.

1 Introduction

Viruses are ubiquitous in the environment and form complex and dynamic relationships with their host populations. Viruses that infect bacteria or archaea are of particular interest because they influence the dynamics and functions of their host populations that in turn drive important ecosystem processes such as biogeochemical cycling (Weinbauer and Rassoulzadegan, 2004; Wigington et al., 2016; Liang et al., 2020; Roy et al., 2020; Jansson, 2023). Recent studies have begun to illuminate the roles phages play in complex natural habitats such as soils (Van Goethem et al., 2019; Trubl et al., 2020, 2021; Starr et al., 2021; Chevallereau et al., 2022; Muscatt et al., 2023). For example, phages have been shown to influence soil carbon and nitrogen cycling through mechanisms such as host population control, biomass turnover, and auxiliary metabolic genes (AMG) (Kuzyakov and Mason-Jones, 2018; Trubl et al., 2018, 2023; Starr et al., 2019; Braga et al., 2020; Barnett and Buckley, 2023; Tong et al., 2023). While it is clear that viruses are essential components of ecosystems, we still have much to learn about how disturbances influence viral communities, especially disturbances related to human activity and climate change (Jansson and Wu, 2022; Liao et al., 2022; Santos-Medellín et al., 2023).

Disturbances can affect the population dynamics of soil viruses and their host microorganisms (Voigt et al., 2021; Jansson and Wu, 2022). Pulse disturbances, which impact in the short-term within a single generation (e.g., soil re-wetting), can produce drastic shifts in the viral community structure, including increased viral abundances, particularly viruses linked to hosts undergoing population growth (Van Goethem et al., 2019; Santos-Medellín et al., 2023). Press disturbances, which are long-term or continuous events that impact multiple generations (e.g., land use change), also can alter soil viral communities. For example, across land use types, differences in soil pH are important determinants of viral composition (Liao et al., 2022). As a

virus' population dynamics feedback on those of its hosts, viral dynamics during and following disturbances could provide insights a microbiome's overall recovery and stability. However, there are knowledge gaps in how viral communities shift during soil recovery from disturbances, particularly press disturbances (Jansson and Wu, 2022). The complex lifestyle of viruses can make it difficult to predict their dynamics under changing environmental conditions. For example, viruses can remain dormant either as infectious particles or in a lysogenic lifestyle for extended periods. This lifestyle is expected in soils, as viruses have been experimentally shown to persist for weeks in soils without a host (DiPietro et al., 2023).

We used the unique temperature disturbance gradient in Centralia, Pennsylvania, USA, to examine viral community diversity and interannual dynamics over seven years as soils recover from a decades-old underground coal seam fire (Barnett and Shade, 2024a). The fires under Centralia have been burning along two roughly linear coal seams since 1962, elevating surface soil temperatures and releasing combustion products through soil vents (Elick, 2011). A disturbance gradient has formed as the fires advance along the coal seams. Undisturbed soils (reference) are in front of and alongside the fire front (Fig. S1A), heated soils (fire-affected) are directly above the active fire (Fig. S1A), and recovering soils are where the fire has already passed (Lee et al., 2017; Kearns and Shade, 2018; Barnett and Shade, 2024a). The disturbance gradient provides a window into the dynamics of microbial communities during an intense press disturbance and through their post-disturbance recovery. Our previous studies have demonstrated the bacterial community's resilience to long-term heating in Centralia soils. The bacterial communities in the recovered soils were more similar to communities in reference soils than fire-affected soils (Lee et al., 2017). As fire-affected soils cooled over seven consecutive years, we further observed directional changes in the bacterial community structure, with the community

becoming more similar to the unaffected reference communities over time (Barnett and Shade, 2024a).

No studies have examined soil viral dynamics under long-term soil heating conditions like those in Centralia. Furthermore, it is not clear how much soil viral communities change over multiple years. Given the high spatial variation in soil viral communities compared across soils that were meters to tens of meters apart (Santos-Medellín et al., 2022), it may be naively expected that this pattern extends to temporal scales, with viral communities expected to be highly variable over time. We hypothesized that as Centralia soils cool and the residing bacterial communities recover (Barnett and Shade, 2024a), the viral populations also stabilize. In other words, we expect the viral population dynamics to reflect parallel resilience to the bacterial populations they infect and thus display similar trends in beta diversity over the seven years.

2 Materials and Methods

2.1 Study site and soil sampling

The Centralia study site has been described previously (Elick, 2011; Lee et al., 2017; Kearns and Shade, 2018; Barnett and Shade, 2024a). The sites in Columbia County, Pennsylvania, USA, overlay an anthracite seam of the Buck Mountain coal bed that has been burning constantly since 1962. The fire lies about 46 m below the surface, and heat and combustion gasses from the fire affect the surface, increasing surface soil temperatures and producing steam vents (Elick 2011). Based on our previous studies (Lee et al., 2017; Kearns and Shade, 2018; Barnett and Shade, 2024a), we selected ten sites (*i.e.,* SiteID) for untargeted metagenome sequencing, including three reference sites and seven fire-affected sites (Fig. S1A, Table S1).

Soils were collected from each site from 2015 to 2021 in the first or second week of October. Twenty cm depth soil cores were collected as described (Barnett and Shade, 2024a), sieved at 4 mm, hand-homogenized, and flash-frozen in liquid nitrogen. *In situ,* soil temperatures were collected with an electronic thermometer. As previously reported (Barnett and Shade, 2024a), soil temperature across the fire-affected sites decreased over the seven sampling years as the fire gradually advanced and those soils began to cool (Fig S1B, Table S1). Over the seven years of sampling, reference sites ranged from 11.1-18.0˚C and fire-affected sites ranged from 14.4-46.3˚C. In total, 70 samples were collected and processed from three reference and seven fire-affected sites that were each repeatedly sampled over seven years.

2.2 DNA extraction and sequencing

DNA was extracted from 0.25 g of soil as previously described following the Griffiths et al. phenol-chloroform protocol (Griffiths et al., 2000), with the alternate use of 0.1 mm zirconium bead containing BeadBug homogenizer tubes (#Z763764; Benchmark Scientific, Sayreville USA) for bead beating. DNA was stored at -80˚C until sequencing. Libraries were prepared and sequenced by the Research Technology Support Facility (RTSF) at Michigan State University (East Lansing, MI, USA) as previously described (Barnett and Shade, 2024b). At this stage, one sample (Site 08 from 2019) was removed due to failed library preparation. The remaining 69 barcoded samples were then pooled into one sequencing library that was sequenced in two lanes of a NovaSeq S4 flow cell in 2x150 bp paired-end format on a NovaSeq 6000 (Illumina, San Diego, USA). Demultiplexing the 69 samples from the pooled data was performed by the RTSF using standard methods. Raw sequencing data is available on the NCBI SRA (BioProject PRJNA974462).

2.3 Sequence processing

Metagenomic sequencing reads were processed as described (Barnett and Shade, 2024b) using a custom workflow based on the Joint Genome Institute (Clum et al., 2021). Quality-controlled reads were assembled within individual samples using SPAdes version v3.15.5 in metaSPAdes mode with assembler only (Nurk et al., 2017). Assemblies can be accessed through NCBI (BioProject PRJNA974462), and metagenome processing code is available via GitHub (https://github.com/ShadeLab/Centralia_7year_metagenome_processing_Barnett_2024). CRISPR spacers were identified from all contigs with minCED version 0.3.2 (https://github.com/ctSkennerton/minced). Ribosomal proteins were annotated from open reading frames identified by Prodigal (Hyatt et al., 2010) by comparing to the single copy gene profile database from anvi'o (Eren et al., 2021) using hmmscan (http://hmmer.org). We quantified the numbers of CRISPR arrays per-genome by dividing CRISPR counts by the median number of copies of ribosomal proteins in each sample.

All viral sequence processing code is available via GitHub (https://github.com/ShadeLab/Centralia_phages_Barnett). We used a custom workflow based on the VirSorter2 SOP (dx.doi.org/10.17504/protocols.io.bwm5pc86) to identify unique viral operational taxonomic units (vOTU; Fig. 1). We first ran VirSorter2 version 2.2.4 (Guo et al., 2021), removing contigs less than 5 Kbp, keeping the original sequences, and using a minimum score of 0.5. The identified contigs were assessed with CheckV version 1.0.1 (Nayfach et al., 2021) to identify viral and host genes and trim sequences to viral regions. These CheckV-trimmed sequences, including both viral and proviral sequences, were then again run through VirSorter2 but this time without filtering (--viral-gene-enrich-off and --provirus-off) to generate

files utilized by DRAM-v (Shaffer et al., 2020). Finally, these sequences were annotated with DRAM-v. The sequences were then categorized into two "keep" categories (cat 1: number of viral genes > 0, and cat 2: number of viral genes = 0 and number of host genes = 0 or score >= 0.95 or number of hallmark genes >2), or a manual curation category (number of viral genes = 0 and number of host genes = 1 and length >= 10 Kbp). All other sequences were removed. Sequences in the keep category 2 that contained genes common in both viruses and hosts and can lead to false positive assignment were moved into the manual curation category. These suspect annotations were "carbohydrate kinase," "carbohydrate-kinase," "glycosyltransferase," "glycosyl transferase," "glycosyl transferaseendonuclease," "nucleotide sugar epimerase," "nucleotide sugar-epimerase," "nucleotide-sugar epimerase," "nucleotide-sugar-epimerase," "nucleotidyltransferase," "nucleotidyl transferase," "nucleotidyl-transferase," "plasmid stability," "endonuclease". The manual curation category was then further filtered to those in one of two further categories. The first category included sequences having structural genes, hallmark genes, depletion in annotations, or enrichment for hypotheticals. The second category included sequences lacking hallmarks, but over 50% of the annotated genes hit a virus, and at least half of those have viral bitscore >100, and the contig is less than 50 Kbp in length. Final viral sequences from all samples that passed filtering were then binned into vOTU using the MMseqs2 version 14.7e284 (Steinegger and Söding, 2017) easy-cluster tool with a minimum sequence identity of 0.95 and target coverage of 0.85.

Reads across all samples were mapped to vOTUs using BBmap (Bushnell, 2019). vOTU were clustered with the INPHARED reference database (downloaded November 9th, 2023) (Cook et al., 2021) using vConTACT2 (Bin Jang et al., 2019) and taxonomy assigned using graphanalyzer (Mattia et al., 2022). vOTU quality was determined using CheckV (Nayfach et al.,

171 2021). In an attempt to match vOTU to potential hosts, CRISPR spacers were aligned to vOTU
172 using BLASTn (Altschul et al., 1990). Matches were determined by 100% identical sequence
173 mapping between the full CRISPR spacer and the vOTU.
174
175 *2.4 Statistical analysis*
176 All statistical analyses were performed in R version 4.2.2 (R Core Team, 2018) with code
177 available via GitHub (https://github.com/ShadeLab/Centralia_phages_Barnett). Since our
178 experimental design utilized repeated sampling of sites over seven years, we used linear mixed
179 effects (LME) models from package nlme (Pinheiro et al., 2020) to compare count, diversity, and
180 abundance measures across time and soil temperature. For all LME models, the site was the
181 random effect. Differences in measures between fire classification (fire-affected versus reference
182 soils) were determined using the nonparametric Wilcoxon rank sum test. We used the Benjamini-
183 Hochburg procedure for all post-hoc tests to adjust p-values for multiple comparisons.
184 vOTU were determined to be present in a sample if the average fold coverage was ≥ 1
185 and the percent of the vOTU sequence covered was ≥ 75%. vOTU abundance was measured as
186 reads per thousand base pairs per million mapped reads (RPKM). RPKM for each vOTU was
187 calculated for each sample as the number of forward and reverse reads mapped to the vOTU in
188 that sample divided by the length of the vOTU in Kbp and then further divided by the total
189 number of reads mapped to all vOTU in that sample divided by 1,000,000. Between sample
190 diversity (beta diversity) was calculated with the Bray-Curtis dissimilarity index through
191 package vegan (Oksanen et al., 2018). Variation in beta diversity was explained by fire
192 classification and year, and their interaction was determined using PERMANOVA through
193 package vegan with samples blocked by site ID to account for repeated sampling. To see if viral

194 community beta-diversity differed between spatial and temporal scales, we compared the Bray-
195 Curtis dissimilarities between sites within each year to dissimilarities within each site over time
196 using the Wilcoxon rank sum test. We further performed beta diversity partitioning on these
197 diversity measures to distinguish the influence of balanced variation in species abundances from
198 abundance gradients in vOTU community dissimilarity (Baselga, 2013). Beta diversity
199 portioning was performed using package betapart (Baselga and Orme, 2012) and compared
200 across spatial and temporal scales using the Wilcoxon rank sum test. Community trajectory
201 directionality, defined as the degree to which a community structure shifts in a consistent
202 direction in terms of community dissimilarity (*i.e.,* Bray-Curtis dissimilarity) over time, was
203 measured using the package ecotraj (De Cáceres et al., 2019; Sturbois et al., 2021). We tested
204 whether the dissimilarity in community structure between sites within a year was different from
205 dissimilarity over time within a site using the Wilcoxon rank sum test.
206 Finally, we wanted to compare the beta diversity of the viral community (vOTUs) to that
207 of the bacterial communities (OTU). Bacterial OTU were processed as previously described
208 (Barnett and Shade, 2024a). As the previous study had more sites than included here, we
209 removed samples not included in metagenome sequencing and rarefied the resulting OTU table
210 to the lowest sequencing depth (161,171 reads). Bray-Curtis dissimilarity in bacterial
211 communities was calculated as before. To compare viral and bacterial community beta diversity
212 structures, we performed a Mantel test using the two Bray-Curtis dissimilarity matrices and a
213 Procrustes analysis using the principal coordinate analysis ordinations (PCoA) of the two Bray-
214 Curtis dissimilarity matrices. After scaling and centering, we also performed a partial Mantel
215 test, controlling for the Euclidian distance in both pH and soil temperature between samples to
216 remove any potential correlation due to these notable edaphic factors (Barnett and Shade,

2024a). Mantel tests, PCoA ordination, and Procrustes analysis were all performed using the vegan package.

3 Results

3.1 Number of CRISPR arrays per-genome increased with disturbance intensity

The number of CRISPR arrays found per-genome is an indication of the average per-bacterial host investment in immune memory of past viral infections and thus can be a signature of previous virus-host interactions within the community. CRISPR arrays per-genome was measured as the number of CRISPR arrays identified relative to the median number of copies of ribosomal proteins found in a sample. Overall, we found that the number of CRISPR arrays per-genome was higher in the fire-affected sites than in the reference sites (Fig 2A; Wilcoxon test: $W = 49$, p-value < 0.001). Regardless of fire classification, CRISPR arrays per-genome increased with soil temperature (Fig 2B; LME: slope = 0.146, p-value < 0.001) and decreased over the seven years in the fire-affected sites but not reference sites (Fig fire-affected LME: slope = -0.386, p-value < 0.001; reference LME: p-value = 0.165).

3.2 vOTU diversity varied with soil temperature and time

We found a total of 6929 vOTU across all Centralia soils. To assess within-sample vOTU diversity, we used two main measures: the percentage of sequenced reads mapped to vOTU (Fig. S2), and richness adjusted by total number of reads mapped to vOTU (Fig. S3). Richness is an ecological metric that indicates the total number of distinct taxa detected within a community. More reads mapped to vOTU in the fire-affected soils than reference soils (Wilcoxon test: $W = 42$, p-value < 0.001), and the percentage of reads mapping to vOTU decreased over time in fire-

affected soils (LME: -7.31E-2, p-value < 0.001) but not reference soils (LME: p-value = 0.818). Correspondingly, the percentage of reads mapping to vOTU decreased with decreasing temperature (LME: slope = 3.63E-2, p-value < 0.001). However, relatively fewer vOTU were found in the fire-affected soils than in reference soils (Wilcoxon test: W = 898, p-value < 0.001), and vOTU richness increased over time in fire-affected soils (LME: slope = 3.032, p-value < 0.001) but not in reference soils (LME: p-value = 0.358). As fire-affected soils were cooling off over time (Fig. S1), we correspondingly saw, vOTU richness increased with decreasing temperature (LME: slope = -1.627, p-value < 0.001).

We next used Pielou's evenness to examine the patterns in relative abundances of vOTU within the soils (Fig. 3A-C). Abundance refers to the amount or number of individuals that collectively contribute to a taxonomically coherent population within a community. Here, we used RPKM as a measure of vOTU relativized abundance. Evenness is an ecological metric to summarize the equivalence among the taxon abundances within a community (e.g., if all taxa in a community have the same abundance, the community is perfectly "even"). vOTU abundances were less even in the fire-affected soils than in the reference soils (Wilcoxon test: W = 874, p-value < 0.001), and vOTU evenness increased over time in fire-affected soils (LME: slope = 1.46E-2, p-value < 0.001) but not reference soils (LME: p-value = 0.875). Correspondingly, vOTU evenness increased with decreasing temperature regardless of fire classification (LME: slope = -5.02E-3, p-value < 0.001). While new unique vOTUs were detected in each consecutive year, particularly in fire-affected sites (Fig. 3D), within-site vOTU accumulation approached saturation within each site over time, suggesting low cross-year variability within a site. However, when considering each additional site sampled within any single year, there was

continued and sharp increase in vOTU accumulation, suggesting high site-to-site variability within a year (Fig. 3E).

To examine across-sample vOTU diversity, we used Bray-Curtis dissimilarity with RPKM to measure abundance. Variation in vOTU community structure across samples (Fig. 4A) was explained by fire classification ($R^2 = 0.086$; p-value = 0.002), time ($R^2 = 0.064$; p-value = 0.001), and their interaction ($R^2 = 0.060$; p-value = 0.005). Unlike the bacterial OTU diversity (Barnett and Shade, 2024a), there was no variation in beta dispersion across fire classification for vOTU (p-value = 0.516). We found that time only explains variation across fire-affected samples (post-hoc PERMANOVA $R^2 = 0.102$, adjusted p-value = 0.006) but not across reference samples (adjusted p-value = 0.404). In agreement with the vOTU accumulation curves, vOTU dissimilarity varied more across sites within a year (*i.e.,* spatial scale) than within a site across years (*i.e.,* temporal scale) (Fig. 4B; Wilcoxon-test; W = 57064, p-value < 0.001). Upon partitioning these dissimilarities (Baselga, 2013), we observed greater abundance gradients (overlap and gradual changes in taxa along a gradient) within a site across years than across sites (Fig. 4C; Wilcoxon-test; W = 7557, p-value < 0.001) but greater balanced variation in species abundances (replacement of taxa along a gradient) across sites than within a site (Fig. 4D; Wilcoxon-test; W = 57210, p-value < 0.001).

Trajectory analysis, which examines shifts in community composition over time, found no statistically significant variation in viral community directionality between fire-affected and reference communities on the whole (Fig. 4E; Wilcoxon-test; W = 17, p-value = 0.183), but directionality was positively associated with maximum soil temperature at each site (*i.e.,* disturbance intensity; Fig. 4F; linear regression; slope = 2.41E-3, $R^2 = 0.379$, p-value = 0.034). However, there was no change in Bray-Curtis dissimilarity between fire-affected and reference

sites as the soils cooled (Fig. 4G; linear mixed effects model; p-value = 0.287), likely due to exceedingly high dissimilarity in viral community composition (Bray-Curtis dissimilarity > 0.9) across all sites at all time points.

3.3 Dissimilarity in viral community structure is correlated to bacterial community structure

vOTU community structure (Bray-Curtis dissimilarity) correlated strongly with bacterial structure (Mantel test: r = 0.5113, p-value < 0.001). There was also concordance between vOTU and bacterial OTU community structures across samples (Procrustes: m12 squared = 0.1502, p-value = 0.001). After accounting for correlation explained by soil temperature and pH, vOTU and bacterial community structures were more weakly correlated (partial Mantel test: r = 0.3863, p-value < 0.001). Notably, bacterial community structure was highly correlated to differences in soil temperature and pH (Mantel test: r = 0.6717, p-value < 0.001), as previously shown (Barnett and Shade, 2024a),.

3.4 vOTU taxonomy was dominated by unknown viruses

Over 90% of the vOTUs were unclassified, with another 8.6% classified in the *Caudoviricetes* class but unclassified at the family level. Of the other classified taxa, in the *Caudoviricetes* class there were nine *Autographiviridae*, five *Steigviridae*, three *Mesyanzhinovviridae* and *Winoviridae*, two *Herelleviridae* and *Peduoviridae*, and one each from *Anaerodiviridae, Kyanoviridae, Podoviridae, Pungoviridae,* and *Zobellviridae*. We also found 27 *Tectiviridae* from class *Tectiliviricetes*, two *Inoviridae* from class *Faserviricetes*, and one *Microviridae* from class *Malgrandaviricetes*.

3.5 Few vOTUs had matching CRISPRs

There were 184 vOTU (2.66% of vOTU) that matched one or more CRISPR spacers across samples. These CRISPR spacers (2672 spacers) represented 2.46% of all identified CRISPR spacers across samples but 10.48% of all CRISPR arrays (674 arrays). Over half of these CRISPRs (1726 spacers and 425 arrays) were detected in the same samples as their mapped vOTUs, indicating the co-occurrence of many of these host-virus pairings (Fig. S4). Notably, at least one vOTU-CRISPR spacer co-occurrence was observed in all seven fire-affected sites, but there were no such co-occurrences detected in any reference site. While over half of CRISPR arrays that matched at least one vOTU were present on contigs unclassified at the phylum level (366), 120 were classified in the phylum *Pseudomonadota*, 86 were classified as *Actinomycetota*, 44 were classified as *Planctomycetota*, and less than ten each were classified in *Cyanobacteriota, Verrucomicrobiota, Acidobacteriota, Bacillota, Bacteroidota, Myxococcota, Deinococcota, Euryarchaeota, Thermomicrobiota, Chloroflexota,* and *Thermotogota*.

3.6 Few vOTU had annotated auxiliary metabolic genes

Out of all vOTU, we found 790 putative auxiliary metabolic genes (AMGs) in 533 vOTU (7.69% of vOTU). Considering all AMGs (Fig. S5), the cumulative abundance of vOTU-containing an annotated AMG did not vary across fire-affected and reference sites (Wilcoxon test: $W = 491$, p-value = 0.995), soil temperature (LME: p-value = 0.581), or time (LME: fire-affected p-value = 0.713, reference p-value = 0.951). When examining only AMGs annotated as carbohydrate-active enzyme genes (CAZymes), the cumulative abundance of vOTU-containing CAZymes was higher in fire-affected sites than the reference sites (Fig. 5A; Wilcoxon test: $W = 192$, p-value < 0.001) and decreased with decreasing temperature (Fig. 5B; LME: slope = 44.9,

p-value = 0.040). However, while the cumulative abundance of vOTU-containing peptidase annotated AMGs was also high in the fire-affected sites (Fig. 5C; Wilcoxon test: W = 192, p-value < 0.001), there was no relationship with temperature (Fig. 5D; LME: p-value = 0.791).

4 Discussion

We examined seven years of viral diversity dynamics in soils associated with the Centralia, Pennsylvania, coal mine fire. This multi-year time series provided the opportunity to consider not only long-term "natural" trends in soil viral community structure, but also the long-term responses to an ongoing but lessening heat disturbance. We surprisingly observed that there was greater variation in soil viral communities across space than over time. In other words, while we found high site-to-site variation in viral community structure, as previously well reported (Noah et al., 2007; Srinivasiah et al., 2015; Emerson et al., 2018; Trubl et al., 2018; Chevallereau et al., 2022; Durham et al., 2022; Santos-Medellín et al., 2022; Barnett and Buckley, 2023; Graham et al., 2023), we found relatively consistent viral communities within a site across the seven years, with the reference soils having a relatively more stable community than the fire-affected soils. Furthermore, while the bacterial and viral communities were correlated to each other and to the major environmental drivers of temperature and pH, the viral communities did not exhibit the same degree of resilience and recovery to the fire as the bacterial communities in terms of community structure.

It has been suggested that the high spatial variability in soil viral communities, even at small scales (*i.e.,* meters), is due to strong dispersal limitation (Santos-Medellín et al., 2022). Our finding of greater abundance variation and lower balanced variation contributions to vOTU dissimilarity within sites over time relative to across sites within a timepoint supports this

hypothesis (Figs. 4C and D). There were fewer unique vOTU discovered with successive annual sampling within a site as compared to successive site sampling within a year (Figs. 3D and E). In other words, the composition of the vOTUs within a site did not change as much over time, despite dynamic environmental conditions within and across years. This lack of viral community turnover may be due to limited immigration of new viruses into a site even over the years, with viral community changes primarily attributed to variations in the abundances of persistent members. As an expected limitation in using metagenome methods to understand viral communities (e.g., due to imperfect assembly and annotation methods particularly for viruses), some turnover observed in Centralia may be due to the recovery of previously unassembled viral sequences. Regardless, these results were robust across the sites included in this study, which represented a broad range of environmental contexts. Notably, our annual sampling approach (soils collected in autumn in early October) cannot provide insights into this ecosystem's seasonal and shorter-term environmental changes (Barnett and Shade, 2024a) that are also expected to influence the viral communities (*e.g.,* snowfall and seasonal precipitation changes). More resolved seasonal sampling is required to observe such seasonal drivers and understand the viral community responses to them. For example, a study of high altitude watershed soils observed some seasonal trends in viral communities linked to summer snowmelt and host population dynamics (Coclet et al., 2023). Viral particle abundances in soil have also been observed to change throughout the seasons, though this is may be associated with agricultural practices and other short term disturbances and is not consistent across soil ecosystem types (Roy et al., 2020; Cornell et al., 2021). Nonetheless, this study suggests that there are either temporally stable or annually recurring viral communities that were relatively consistent within a soil sampling location.

Of note, the vast majority of our vOTUs (> 90%) were unclassified using current viral reference databases (INPHARED downloaded November 9th, 2023). This is unsurprising considering the vast richness of viruses across diverse ecosystems (Paez-Espino et al., 2016) and the reported under sampling of viruses from the soil environment in particular (Graham et al., 2019; Jansson, 2023). The observation of high spatial variability in viral communities further demonstrates the daunting task of surveying and classifying the viral "dark matter" in soils, even at relatively local scales. However, environmental viral reference databases are constantly and rapidly growing (Roux et al., 2019, 2021; Trubl et al., 2020; Santiago-Rodriguez and Hollister, 2022), and as more studies are conducted and bioinformatic methods for recovering viral genomes are improved, the percentage of unclassified vOTUs are expected to decrease. While the current lack of sequenced representatives limits our ability to classify most of the vOTU discovered here, since we are using vOTU as our finest taxonomic unit, it does not limit our ability to compare the viral community structures within our system.

While we previously observed strong resilience in the bacterial communities as the temperatures cooled with fire advancement in Centralia (Barnett and Shade, 2024a), there was evidence for relatively lower community resilience for the viruses than for the bacteria. Over time, and even as soils cooled, highly dissimilar viral communities remained across fire-affected and reference communities (Fig. 4G), likely due to the inherently high spatial variability in viral structure across sites. While viral community directionality was highest in the soils undergoing the hottest recorded temperature over the seven sampling years, there was no overall difference in directionality between the viral communities across fire-affected and reference soils (Figs. 4E and F). These results suggest that the viral community composition remains distinct across disturbed soils for a longer duration than that of the bacterial communities. Viral diversity has

even been shown to be uncoupled from microbial diversity in some soil systems (Ma et al., 2024). Thus, changes in bacterial communities during recovery may not lead to comparable changes in their viral communities. While both viral and bacterial community dissimilarity across samples were correlated, this correlation was reduced when controlling for correlation with soil temperature and pH. This uncoupling of viral and host communities may be partly due to the previously suggested limited dispersal of viruses across local spatial scales.

Our results suggest that there could have been a change in the interactions between viruses and bacterial hosts across the disturbance gradient and during soil recovery. The number of CRISPR arrays per genome and viral community evenness shifted as soils cooled, with both measures becoming more like those of undisturbed reference sites (Figs. 2 and 3A-C). The observed highly uneven viral communities suggest that there were a few highly abundant viruses, typical of predator-prey dynamics such as “kill-the-winner” (Rodriguez-Brito et al., 2010). It is possible that Centralia fire-affected soils may be under more intense predator-prey relationships than undisturbed soils, possibly due to an observed decrease in bacterial community richness (e.g., fewer host taxa that have relatively more viral infections) and greater selection for heat tolerant bacterial species in the hotter sites. Increased viral infection and bacterial biomass turnover has previously been observed in experimental soils incubated at 23˚C relative to 18˚C incubations leading to increased soil carbon mineralization (Wang et al., 2022). High temperatures may also induce temperate viruses to become lytic, increasing host population turnover (Jansson and Wu, 2022).

We observed more CRISPR arrays in the fire-affected soils than in the reference soils (Fig. 2A). CRISPR mechanisms are used by bacteria and archaea to combat phage infection through targeted and adaptive destruction of viral DNA. In this study we cannot directly

423 determine the mechanism behind this selection pressure, whether it be increased viral infection in
424 heated soils leading to increased CRISPR counts in genomes of resident microbes or the heating
425 directly selecting for bacterial taxa that tend to have naturally higher CRIPSR investment.
426 However, overall CRISPR abundance is correlated with viral abundance across many
427 ecosystems, linking CRISPR system prevalence with environmental viral diversity (Meaden et
428 al., 2022). Previous work that examined metagenomes from eight years of experimental warming
429 of soils to about 4˚C above ambient demonstrated an increased prevalence of the CRISPR
430 associated marker gene, *cas1*, among populations of diverse phyla. In that study, increased
431 prevalence was defined as the abundance of *cas1* genes normalized to the abundance of the
432 phylum they were classified to. Such increased prevalence suggested selection for phage defense
433 in warmed soil (Wu et al., 2020). Our study used a comparable measure of CRISPR prevalence
434 by normalizing CRISPR array counts with estimated genome counts determined by median
435 ribosomal protein gene counts, regardless of taxonomy. Our findings thus agree with this
436 previous work despite differences in warming intensity (generally over 10˚C above ambient
437 temperature in Centralia) and determination of genomic prevalence.
438 While we did not find many auxiliary metabolic genes (AMG) in the vOTUs detected and
439 annotated, the AMGs we found suggest that these viruses may influence bacterial metabolisms
440 associated with ecosystem functioning (Breitbart et al., 2007). AMGs expressed in infected
441 bacterial cells can reprogram the cellular metabolism and cellular function in ecosystem
442 processes such as carbon and nitrogen cycling (Breitbart, 2011; Hurwitz et al., 2013). In soils, in
443 particular, some AMGs are believed to be important in carbon processing by providing hosts
444 with new or alternative genes for polysaccharide binding, breakdown of complex carbohydrates,
445 and central carbon metabolism (Trubl et al., 2018). Across the Centralia soils, while there was

not any variation in the abundance of vOTUs containing one or more annotated AMGs across fire classifications and temperature (Fig. S5), there was a higher abundance of vOTU with CAZyme and peptidase AMGs in fire-affected soils relative to reference soils (Figs 5A and C). There was also a positive linear relationship between the abundance of vOTU-encoding CAZyme AMGs and soil temperature (Fig. 5B). Therefore, given their relatively higher abundance in warmer Centralia soils, vOTU may indirectly influence carbohydrate and protein metabolisms within hotter, more disturbed soils than undisturbed soils. Despite the limitations of work utilizing untargeted metagenomes to assay viral communities (*i.e.,* under-sampling of the viral community, incomplete annotations of AMG, and unquantified expression of these AMGs), our results demonstrate that disturbances influence the representation of AMGs in soil viruses which may influence viral roles in ecosystem processes beyond the direct response to the changes in their host populations.

5 Conclusion

Altogether, this work provides insights into two major unknowns in our understanding of soil virus communities. It first showed that viral communities can be remain relatively predictable in their community structure given a consistent sampling time across years, and that, within a particular soil site, it is possible to exhaust the needed observation effort for the local virus community. Given more resolved temporal sampling, we expect that it also will be possible to understand seasonal influences and intra-annual dynamics within a soil site, which have not yet been well-documented. Second, this study showed how viral communities can respond to the long-term ecosystem disturbance of warming. The comparison of the viral community responses with those of their host communities has implications for bacterial biogeochemical cycling that

can be influenced by phage infection. Thus, understanding when, how, and to what extent soil virus and host dynamics are coupled, and the consequences of disturbances like warming on that coupling, provides critical information about how microbial functions will be altered given the climate crisis.

Acknowledgments

We thank Tammy Tobin for Centralia research consultation and local sampling support and Keara L Grady for coordinating fieldwork logistics and organization of multi-year soil procurement and processing. We additionally thank SH Lee, JW Sorensen, TK Dunivin, JL Chodkowski, N Stopnisek, and M Mechan Llontop for collecting samples. This work was supported by the U.S. National Science Foundation CAREER award #1749544 to AS. AS acknowledges the French Centre National de la Recherche Scientifique (CNRS) support. We declare no conflicts of interest.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410. doi:https://doi.org/10.1016/S0022-2836(05)80360-2

Barnett, S.E., Buckley, D.H., 2023. Metagenomic stable isotope probing reveals bacteriophage participation in soil carbon cycling. Environmental Microbiology 25, 1785–1795. doi:https://doi.org/10.1111/1462-2920.16395

Barnett, S.E., Shade, A., 2024a. Arrive and wait: Inactive bacterial taxa contribute to perceived soil microbiome resilience after a multidecadal press disturbance. Ecology Letters 27,

492 e14393. doi:10.1111/ele.14393
493 Barnett, S.E., Shade, A., 2024b. Seven years of microbial community metagenomes from
494 temperate soils affected by an ongoing coal seam fire. Microbiology Resource
495 Announcements 0, e00198-24. doi:10.1128/mra.00198-24
496 Baselga, A., 2013. Separating the two components of abundance-based dissimilarity: balanced
497 changes in abundance vs. abundance gradients. Methods in Ecology and Evolution 4, 552–
498 557. doi:https://doi.org/10.1111/2041-210X.12029
499 Baselga, A., Orme, C.D.L., 2012. betapart: an R package for the study of beta diversity. Methods
500 in Ecology and Evolution 3, 808–812. doi:https://doi.org/10.1111/j.2041-
501 210X.2012.00224.x
502 Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J.H., Roux, S., Adriaenssens, E.M., Brister, J.R.,
503 Kropinski, A.M., Krupovic, M., Lavigne, R., Turner, D., Sullivan, M.B., 2019. Taxonomic
504 assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks.
505 Nature Biotechnology 37, 632–639. doi:10.1038/s41587-019-0100-8
506 Braga, L.P.P., Spor, A., Kot, W., Breuil, M.-C., Hansen, L.H., Setubal, J.C., Philippot, L., 2020.
507 Impact of phages on soil bacterial communities and nitrogen availability under different
508 assembly scenarios. Microbiome 8, 52. doi:10.1186/s40168-020-00822-z
509 Breitbart, M., 2011. Marine Viruses: Truth or Dare. Annual Review of Marine Science 4, 425–
510 448. doi:10.1146/annurev-marine-120709-142805
511 Breitbart, M., Thompson, L.R., Suttle, C.A., Sullivan, M.B., 2007. Exploring the Vast Diversity
512 of Marine Viruses. Oceanography 20, 135–139.
513 Bushnell, B., 2019. BBMap short-read aligner, and other bioinformatics tools [WWW
514 Document].

Chevallereau, A., Pons, B.J., van Houte, S., Westra, E.R., 2022. Interactions between bacterial and phage communities in natural environments. Nature Reviews Microbiology 20, 49–62. doi:10.1038/s41579-021-00602-y

Clum, A., Huntemann, M., Bushnell, B., Foster, Brian, Foster, Bryce, Roux, S., Hajek, P.P., Varghese, N., Mukherjee, S., Reddy, T.B.K., Daum, C., Yoshinaga, Y., O'Malley, R., Seshadri, R., Kyrpides, N.C., Eloe-Fadrosh, E.A., Chen, I.-M.A., Copeland, A., Ivanova, N.N., 2021. DOE JGI Metagenome Workflow. MSystems 6, e00804-20. doi:10.1128/mSystems.00804-20

Coclet, C., Sorensen, P.O., Karaoz, U., Wang, S., Brodie, E.L., Eloe-Fadrosh, E.A., Roux, S., 2023. Virus diversity and activity is driven by snowmelt and host dynamics in a high-altitude watershed soil ecosystem. Microbiome 11, 237. doi:10.1186/s40168-023-01666-z

Cook, R., Brown, N., Redgwell, T., Rihtman, B., Barnes, M., Clokie, M., Stekel, D.J., Hobman, J., Jones, M.A., Millard, A., 2021. INfrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. PHAGE 2, 214–223. doi:10.1089/phage.2021.0007

Cornell, C.R., Zhang, Y., Van Nostrand, J.D., Wagle, P., Xiao, X., Zhou, J., 2021. Temporal Changes of Virus-Like Particle Abundance and Metagenomic Comparison of Viral Communities in Cropland and Prairie Soils. MSphere 6, 10.1128/msphere.01160-20. doi:10.1128/msphere.01160-20

De Cáceres, M., Coll, L., Legendre, P., Allen, R.B., Wiser, S.K., Fortin, M.-J., Condit, R., Hubbell, S., 2019. Trajectory analysis in community ecology. Ecological Monographs 89, e01350. doi:https://doi.org/10.1002/ecm.1350

DiPietro, A.G., Bryant, S.A., Zanger, M.M., Williamson, K.E., 2023. Understanding Viral

538 Impacts in Soil Microbial Ecology Through the Persistence and Decay of Infectious
539 Bacteriophages. Current Microbiology 80, 276. doi:10.1007/s00284-023-03386-x
540 Durham, D.M., Sieradzki, E.T., ter Horst, A.M., Santos-Medellín, C., Bess, C.W.A., Geonczy,
541 S.E., Emerson, J.B., 2022. Substantial differences in soil viral community composition
542 within and among four Northern California habitats. ISME Communications 2, 100.
543 doi:10.1038/s43705-022-00171-y
544 Elick, J.M., 2011. Mapping the coal fire at Centralia, Pa using thermal infrared imagery.
545 International Journal of Coal Geology 87, 197–203.
546 doi:https://doi.org/10.1016/j.coal.2011.06.018
547 Emerson, J.B., Roux, S., Brum, J.R., Bolduc, B., Woodcroft, B.J., Jang, H. Bin, Singleton, C.M.,
548 Solden, L.M., Naas, A.E., Boyd, J.A., Hodgkins, S.B., Wilson, R.M., Trubl, G., Li, C.,
549 Frolking, S., Pope, P.B., Wrighton, K.C., Crill, P.M., Chanton, J.P., Saleska, S.R., Tyson,
550 G.W., Rich, V.I., Sullivan, M.B., 2018. Host-linked soil viral ecology along a permafrost
551 thaw gradient. Nature Microbiology 3, 870–880. doi:10.1038/s41564-018-0190-y
552 Eren, A.M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S.E., Schechter, M.S., Fink, I., Pan, J.N.,
553 Yousef, M., Fogarty, E.C., Trigodet, F., Watson, A.R., Esen, Ö.C., Moore, R.M., Clayssen,
554 Q., Lee, M.D., Kivenson, V., Graham, E.D., Merrill, B.D., Karkman, A., Blankenberg, D.,
555 Eppley, J.M., Sjödin, A., Scott, J.J., Vázquez-Campos, X., McKay, L.J., McDaniel, E.A.,
556 Stevens, S.L.R., Anderson, R.E., Fuessel, J., Fernandez-Guerra, A., Maignien, L., Delmont,
557 T.O., Willis, A.D., 2021. Community-led, integrated, reproducible multi-omics with anvi'o.
558 Nature Microbiology 6, 3–6. doi:10.1038/s41564-020-00834-3
559 Graham, E.B., Camargo, A.P., Wu, R., Neches, R.Y., Nolan, M., Paez-Espino, D., Copeland, A.,
560 Kyrpides, N.C., Jansson, J.K., McDermott, J.E., Hofmockel, K.S., Consortium, the S.V.,

2023. Global Biogeography of the Soil Virosphere. BioRxiv 2023.11.02.565391. doi:10.1101/2023.11.02.565391

Graham, E.B., Paez-Espino, D., Brislawn, C., Hofmockel, K.S., Wu, R., Kyrpides, N.C., Jansson, J.K., McDermott, J.E., 2019. Untapped viral diversity in global soil metagenomes. BioRxiv 583997. doi:10.1101/583997

Griffiths, R.I., Whiteley, A.S., O'Donnell, A.G., Bailey, M.J., 2000. Rapid Method for Coextraction of DNA and RNA from Natural Environments for Analysis of Ribosomal DNA- and rRNA-Based Microbial Community Composition. Applied and Environmental Microbiology 66, 5488–5491. doi:10.1128/AEM.66.12.5488-5491.2000

Guo, J., Bolduc, B., Zayed, A.A., Varsani, A., Dominguez-Huerta, G., Delmont, T.O., Pratama, A.A., Gazitúa, M.C., Vik, D., Sullivan, M.B., Roux, S., 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome 9, 37. doi:10.1186/s40168-020-00990-y

Hurwitz, B.L., Hallam, S.J., Sullivan, M.B., 2013. Metabolic reprogramming by viruses in the sunlit and dark ocean. Genome Biology 14, R123. doi:10.1186/gb-2013-14-11-r123

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119. doi:10.1186/1471-2105-11-119

Jansson, J.K., 2023. Soil viruses: Understudied agents of soil ecology. Environmental Microbiology 25, 143–146. doi:https://doi.org/10.1111/1462-2920.16258

Jansson, J.K., Wu, R., 2022. Soil viral diversity, ecology and climate change. Nature Reviews Microbiology. doi:10.1038/s41579-022-00811-z

Kearns, P.J., Shade, A., 2018. Trait-based patterns of microbial dynamics in dormancy potential

584 and heterotrophic strategy: case studies of resource-based and post-press succession. The
585 ISME Journal 12, 2575–2581. doi:10.1038/s41396-018-0194-x
586 Kuzyakov, Y., Mason-Jones, K., 2018. Viruses in soil: Nano-scale undead drivers of microbial
587 life, biogeochemical turnover and ecosystem functions. Soil Biology and Biochemistry 127,
588 305–317. doi:https://doi.org/10.1016/j.soilbio.2018.09.032
589 Lee, S.-H., Sorensen, J.W., Grady, K.L., Tobin, T.C., Shade, A., 2017. Divergent extremes but
590 convergent recovery of bacterial and archaeal soil communities to an ongoing subterranean
591 coal mine fire. The ISME Journal 11, 1447–1459. doi:10.1038/ismej.2017.1
592 Liang, X., Zhang, Y., Wommack, K.E., Wilhelm, S.W., DeBruyn, J.M., Sherfy, A.C., Zhuang, J.,
593 Radosevich, M., 2020. Lysogenic reproductive strategies of viral communities vary with
594 soil depth and are correlated with bacterial diversity. Soil Biology and Biochemistry 144,
595 107767. doi:https://doi.org/10.1016/j.soilbio.2020.107767
596 Liao, H., Li, H., Duan, C.-S., Zhou, X.-Y., Luo, Q.-P., An, X.-L., Zhu, Y.-G., Su, J.-Q., 2022.
597 Response of soil viral communities to land use changes. Nature Communications 13, 6027.
598 doi:10.1038/s41467-022-33771-2
599 Ma, B., Wang, Y., Zhao, K., Stirling, E., Lv, X., Yu, Y., Hu, L., Tang, C., Wu, C., Dong, B.,
600 Xue, R., Dahlgren, R.A., Tan, X., Dai, H., Zhu, Y.-G., Chu, H., Xu, J., 2024. Biogeographic
601 patterns and drivers of soil viromes. Nature Ecology & Evolution. doi:10.1038/s41559-024-
602 02347-2
603 Mattia, P., Andrea, T., Gioele, L., M., A.E., Nicola, V., 2022. MetaPhage: an Automated
604 Pipeline for Analyzing, Annotating, and Classifying Bacteriophages in Metagenomics
605 Sequencing Data. MSystems 7, e00741-22. doi:10.1128/msystems.00741-22
606 Meaden, S., Biswas, A., Arkhipova, K., Morales, S.E., Dutilh, B.E., Westra, E.R., Fineran, P.C.,

2022. High viral abundance and low diversity are associated with increased CRISPR-Cas prevalence across microbial ecosystems. Current Biology 32, 220-227.e5. doi:10.1016/j.cub.2021.10.038

Muscatt, G., Cook, R., Millard, A., Bending, G.D., Jameson, E., 2023. Viral metagenomics reveals diverse virus-host interactions throughout the soil depth profile. MBio 14, e02246-23. doi:10.1128/mbio.02246-23

Nayfach, S., Camargo, A.P., Schulz, F., Eloe-Fadrosh, E., Roux, S., Kyrpides, N.C., 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nature Biotechnology 39, 578–585. doi:10.1038/s41587-020-00774-7

Noah, F., Mya, B., James, N., Peter, S., Catherine, L., Ryan, J., Michael, R., A., E.R., Ben, F., Steve, R., Rob, K., Forest, R., B., J.R., 2007. Metagenomic and Small-Subunit rRNA Analyses Reveal the Genetic Diversity of Bacteria, Archaea, Fungi, and Viruses in Soil. Applied and Environmental Microbiology 73, 7059–7066. doi:10.1128/AEM.00358-07

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Research 27, 824–834. doi:10.1101/gr.213959.116

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2018. vegan: Community Ecology Package.

Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N.N., Kyrpides, N.C., 2016. Uncovering Earth's virome. Nature 536, 425–430. doi:10.1038/nature19094

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2020. nlme: Linear and Nonlinear Mixed Effects Models.

630 R Core Team, 2018. R: A language and environment for statistical computing.
631 Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., Buchanan, J.,
632 Desnues, C., Dinsdale, E., Edwards, R., Felts, B., Haynes, M., Liu, H., Lipson, D., Mahaffy,
633 J., Martin-Cuadrado, A.B., Mira, A., Nulton, J., Pašić, L., Rayhawk, S., Rodriguez-Mueller,
634 J., Rodriguez-Valera, F., Salamon, P., Srinagesh, S., Thingstad, T.F., Tran, T., Thurber,
635 R.V., Willner, D., Youle, M., Rohwer, F., 2010. Viral and microbial community dynamics
636 in four aquatic environments. The ISME Journal 4, 739–751. doi:10.1038/ismej.2010.1
637 Roux, S., Adriaenssens, E.M., Dutilh, B.E., Koonin, E. V, Kropinski, A.M., Krupovic, M., Kuhn,
638 J.H., Lavigne, R., Brister, J.R., Varsani, A., Amid, C., Aziz, R.K., Bordenstein, S.R., Bork,
639 P., Breitbart, M., Cochrane, G.R., Daly, R.A., Desnues, C., Duhaime, M.B., Emerson, J.B.,
640 Enault, F., Fuhrman, J.A., Hingamp, P., Hugenholtz, P., Hurwitz, B.L., Ivanova, N.N.,
641 Labonté, J.M., Lee, K.-B., Malmstrom, R.R., Martinez-Garcia, M., Mizrachi, I.K., Ogata,
642 H., Páez-Espino, D., Petit, M.-A., Putonti, C., Rattei, T., Reyes, A., Rodriguez-Valera, F.,
643 Rosario, K., Schriml, L., Schulz, F., Steward, G.F., Sullivan, M.B., Sunagawa, S., Suttle,
644 C.A., Temperton, B., Tringe, S.G., Thurber, R.V., Webster, N.S., Whiteson, K.L., Wilhelm,
645 S.W., Wommack, K.E., Woyke, T., Wrighton, K.C., Yilmaz, P., Yoshida, T., Young, M.J.,
646 Yutin, N., Allen, L.Z., Kyrpides, N.C., Eloe-Fadrosh, E.A., 2019. Minimum Information
647 about an Uncultivated Virus Genome (MIUViG). Nature Biotechnology 37, 29–37.
648 doi:10.1038/nbt.4306
649 Roux, S., Páez-Espino, D., Chen, I.-M.A., Palaniappan, K., Ratner, A., Chu, K., Reddy, T.B.K.,
650 Nayfach, S., Schulz, F., Call, L., Neches, R.Y., Woyke, T., Ivanova, N.N., Eloe-Fadrosh,
651 E.A., Kyrpides, N.C., 2021. IMG/VR v3: an integrated ecological and evolutionary
652 framework for interrogating genomes of uncultivated viruses. Nucleic Acids Research 49,

D764–D775. doi:10.1093/nar/gkaa946

Roy, K., Ghosh, D., DeBruyn, J.M., Dasgupta, T., Wommack, K.E., Liang, X., Wagner, R.E., Radosevich, M., 2020. Temporal Dynamics of Soil Virus and Bacterial Populations in Agricultural and Early Plant Successional Soils. Frontiers in Microbiology 11, 1494.

Santiago-Rodriguez, T.M., Hollister, E.B., 2022. Unraveling the viral dark matter through viral metagenomics. Frontiers in Immunology 13.

Santos-Medellín, C., Blazewicz, S.J., Pett-Ridge, J., Emerson, J.B., 2023. Viral but not bacterial community succession is characterized by extreme turnover shortly after rewetting dry soils. BioRxiv 2023.02.12.528215. doi:10.1101/2023.02.12.528215

Santos-Medellín, C., Estera-Molina, K., Yuan, M., Pett-Ridge, J., Firestone, M.K., Emerson, J.B., 2022. Spatial turnover of soil viral populations and genotypes overlain by cohesive responses to moisture in grasslands. Proceedings of the National Academy of Sciences 119, e2209132119. doi:10.1073/pnas.2209132119

Shaffer, M., Borton, M.A., McGivern, B.B., Zayed, A.A., La Rosa, S.L., Solden, L.M., Liu, P., Narrowe, A.B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M.C., Daly, R.A., Smith, G.J., Vik, D.R., Pope, P.B., Sullivan, M.B., Roux, S., Wrighton, K.C., 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Research 48, 8883–8900. doi:10.1093/nar/gkaa621

Srinivasiah, S., Lovett, J., Ghosh, D., Roy, K., Fuhrmann, J.J., Radosevich, M., Wommack, K.E., 2015. Dynamics of autochthonous soil viral communities parallels dynamics of host communities under nutrient stimulation. FEMS Microbiology Ecology 91, fiv063. doi:10.1093/femsec/fiv063

Starr, E.P., Nuccio, E.E., Pett-Ridge, J., Banfield, J.F., Firestone, M.K., 2019.

Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. Proceedings of the National Academy of Sciences 116, 25900 LP – 25908. doi:10.1073/pnas.1908291116

Starr, E.P., Shi, S., Blazewicz, S.J., Koch, B.J., Probst, A.J., Hungate, B.A., Pett-Ridge, J., Firestone, M.K., Banfield, J.F., 2021. Stable-Isotope-Informed, Genome-Resolved Metagenomics Uncovers Potential Cross-Kingdom Interactions in Rhizosphere Soil. MSphere 6, 10.1128/msphere.00085-21. doi:10.1128/msphere.00085-21

Steinegger, M., Söding, J., 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nature Biotechnology 35, 1026–1028. doi:10.1038/nbt.3988

Sturbois, A., De Cáceres, M., Sánchez-Pinillos, M., Schaal, G., Gauthier, O., Mao, P. Le, Ponsero, A., Desroy, N., 2021. Extending community trajectory analysis: New metrics and representation. Ecological Modelling 440, 109400. doi:https://doi.org/10.1016/j.ecolmodel.2020.109400

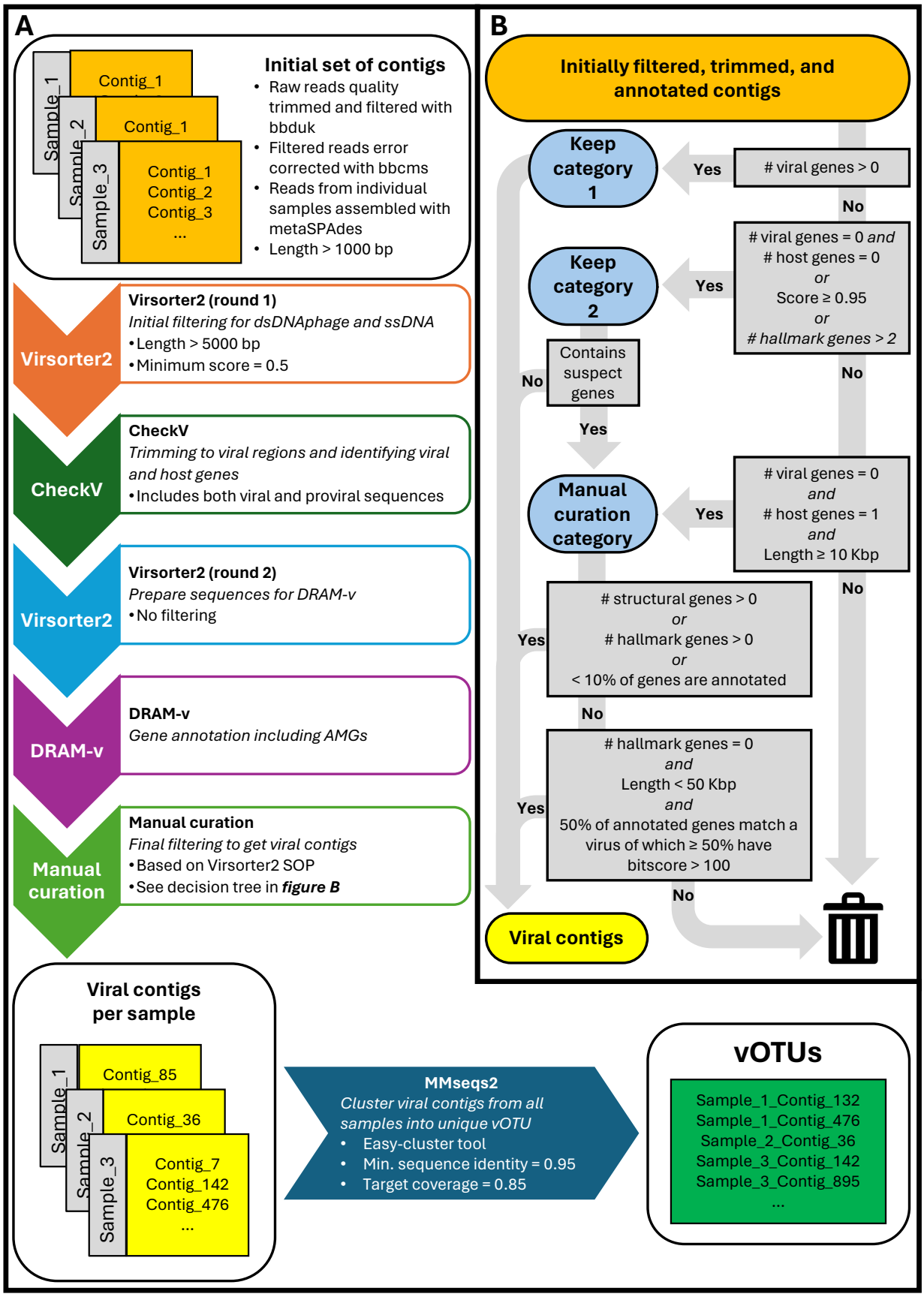
Tong, D., Wang, Y., Yu, H., Shen, H., Dahlgren, R.A., Xu, J., 2023. Viral lysing can alleviate microbial nutrient limitations and accumulate recalcitrant dissolved organic matter components in soil. The ISME Journal 17, 1247–1256. doi:10.1038/s41396-023-01438-5

Trubl, G., Hyman, P., Roux, S., Abedon, S.T., 2020. Coming-of-Age Characterization of Soil Viruses: A User's Guide to Virus Isolation, Detection within Metagenomes, and Viromics. Soil Systems. doi:10.3390/soilsystems4020023

Trubl, G., Jang, H. Bin, Roux, S., Emerson, J.B., Solonenko, N., Vik, D.R., Solden, L., Ellenbogen, J., Runyon, A.T., Bolduc, B., Woodcroft, B.J., Saleska, S.R., Tyson, G.W., Wrighton, K.C., Sullivan, M.B., Rich, V.I., 2018. Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing. MSystems 3, e00076-18. doi:10.1128/mSystems.00076-

699 18
700 Trubl, G., Kimbrel, J.A., Liquet-Gonzalez, J., Nuccio, E.E., Weber, P.K., Pett-Ridge, J., Jansson,
701 J.K., Waldrop, M.P., Blazewicz, S.J., 2021. Active virus-host interactions at sub-freezing
702 temperatures in Arctic peat soil. Microbiome 9, 208. doi:10.1186/s40168-021-01154-2
703 Trubl, G., Roux, S., Borton, M.A., Varsani, A., Li, Y.-F., Sun, C., Jang, H. Bin, Woodcroft, B.J.,
704 Tyson, G.W., Wrighton, K.C., Saleska, S.R., Eloe-Fadrosh, E.A., Sullivan, M.B., Rich, V.I.,
705 2023. Population ecology and potential biogeochemical impacts of ssDNA and dsDNA soil
706 viruses along a permafrost thaw gradient. BioRxiv 2023.06.13.544858.
707 doi:10.1101/2023.06.13.544858
708 Van Goethem, M.W., Swenson, T.L., Gareth, T., Simon, R., Northen, T.R., Martiny, J.B.H.,
709 2019. Characteristics of Wetting-Induced Bacteriophage Blooms in Biological Soil Crust.
710 MBio 10, e02287-19. doi:10.1128/mBio.02287-19
711 Voigt, E., Rall, B.C., Chatzinotas, A., Brose, U., Rosenbaum, B., 2021. Phage strategies facilitate
712 bacterial coexistence under environmental variability. PeerJ 9, e12194.
713 doi:10.7717/peerj.12194
714 Wang, S., Yu, S., Zhao, Xiaoyan, Zhao, Xiaolei, Mason-Jones, K., Zhu, Z., Redmile-Gordon,
715 M., Li, Y., Chen, J., Kuzyakov, Y., Ge, T., 2022. Experimental evidence for the impact of
716 phages on mineralization of soil-derived dissolved organic matter under different
717 temperature regimes. Science of The Total Environment 846, 157517.
718 doi:https://doi.org/10.1016/j.scitotenv.2022.157517
719 Weinbauer, M.G., Rassoulzadegan, F., 2004. Are viruses driving microbial diversification and
720 diversity? Environmental Microbiology 6, 1–11. doi:https://doi.org/10.1046/j.1462-
721 2920.2003.00539.x

722 Wigington, C.H., Sonderegger, D., Brussaard, C.P.D., Buchan, A., Finke, J.F., Fuhrman, J.A.,
723 Lennon, J.T., Middelboe, M., Suttle, C.A., Stock, C., Wilson, W.H., Wommack, K.E.,
724 Wilhelm, S.W., Weitz, J.S., 2016. Re-examination of the relationship between marine virus
725 and microbial cell abundances. Nature Microbiology 1, 15024.
726 doi:10.1038/nmicrobiol.2015.24
727 Wu, R., Chai, B., Cole, J.R., Gunturu, S.K., Guo, X., Tian, R., Gu, J.-D., Zhou, J., Tiedje, J.M.,
728 2020. Targeted assemblies of cas1 suggest CRISPR-Cas's response to soil warming. The
729 ISME Journal 14, 1651–1662. doi:10.1038/s41396-020-0635-1
730
731


A
Sample_1
Contig_1
Sample_2
Contig_1
Sample_3
Contig_1
Contig_2
Contig_3
...
Initial set of contigs
• Raw reads quality trimmed and filtered with bbduk
• Filtered reads error corrected with bbcms
• Reads from individual samples assembled with metaSPAdes
• Length > 1000 bp
Virsorter2
Virsorter2 (round 1)
Initial filtering for dsDNAphage and ssDNA
• Length > 5000 bp
• Minimum score = 0.5
CheckV
CheckV
Trimming to viral regions and identifying viral and host genes
• Includes both viral and proviral sequences
Virsorter2
Virsorter2 (round 2)
Prepare sequences for DRAM-v
• No filtering
DRAM-v
DRAM-v
Gene annotation including AMGs
Manual curation
Manual curation
Final filtering to get viral contigs
• Based on Virsorter2 SOP
• See decision tree in figure B
Viral contigs per sample
Sample_1
Contig_85
Sample_2
Contig_36
Sample_3
Contig_7
Contig_142
Contig_476
...
MMseqs2
Cluster viral contigs from all samples into unique vOTU
• Easy-cluster tool
• Min. sequence identity = 0.95
• Target coverage = 0.85
vOTUs
Sample_1_Contig_132
Sample_1_Contig_476
Sample_2_Contig_36
Sample_3_Contig_142
Sample_3_Contig_895
...
B
Initially filtered, trimmed, and annotated contigs
Keep category 1
Yes
viral genes > 0
No
Keep category 2
Yes
viral genes = 0 and # host genes = 0 or Score ≥ 0.95 or # hallmark genes > 2
No
Contains suspect genes
No
Yes
Manual curation category
Yes
viral genes = 0 and # host genes = 1 and Length ≥ 10 Kbp
No
Yes
structural genes > 0 or # hallmark genes > 0 or < 10% of genes are annotated
No
Yes
hallmark genes = 0 and Length < 50 Kbp and 50% of annotated genes match a virus of which ≥ 50% have bitscore > 100
No
Viral contigs

732

Figure 1: The overview of the bioinformatic sequence processing used in this study. A) The flowchart describing the overall processing of assembled contigs into vOTUs. B) The decision tree of how the initially filtered, trimmed, and annotated contigs were filtered into viral contigs.

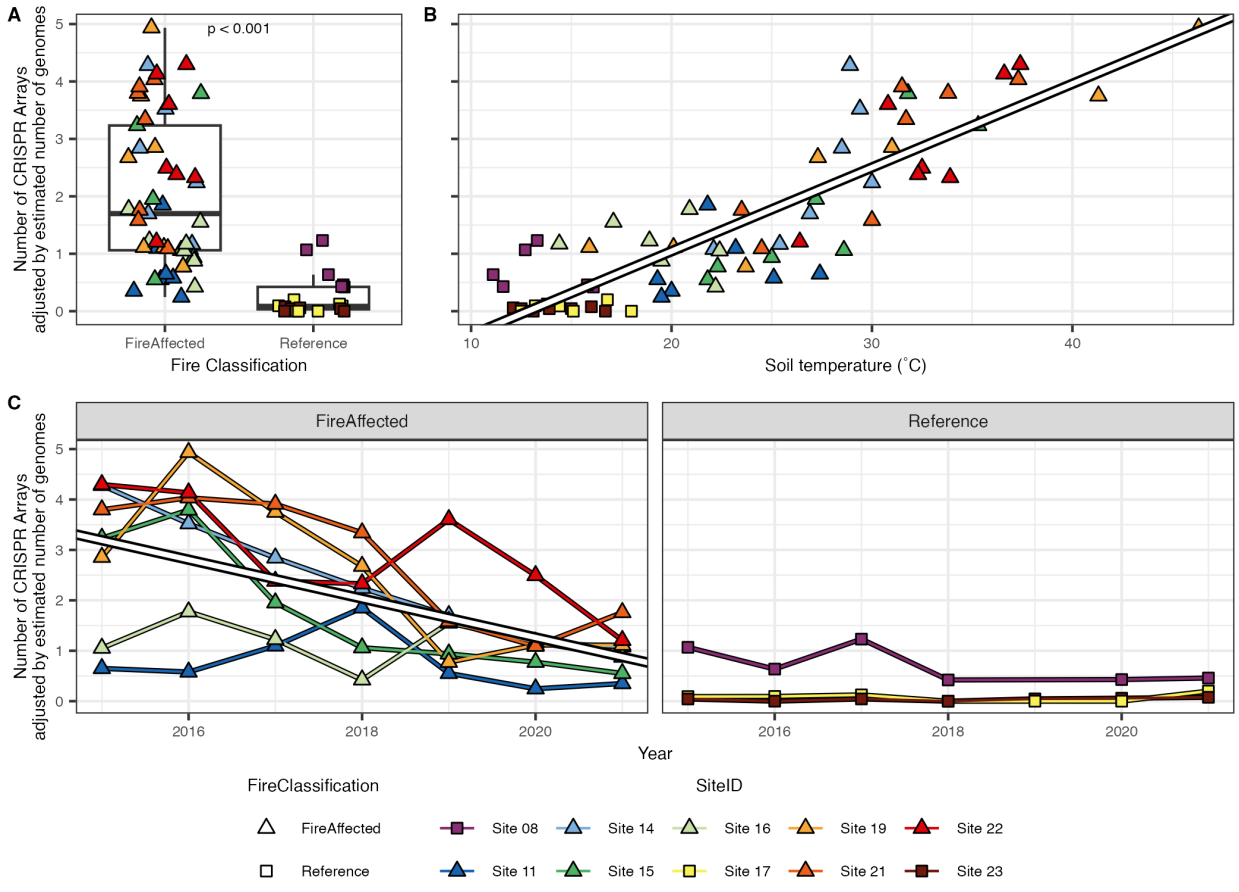


Figure 2: The number of CRISPR arrays per-genome was higher in fire-affected soils than in reference soils and decreased as the fire-affected soils cool. The number of CRISPR arrays per-genome was calculated as the number of CRISPR arrays found divided by the median count of ribosomal protein within each sample, which is a proxy for the number of individual genomes detected given the sequencing effort. A) The number of CRISPR arrays per genome was compared between fire-affected and reference soils with the p-values of the Wilcoxon rank sum

745 test indicated. B) The relationship between the number of CRISPR arrays per genome to soil
746 temperature (˚C). The line shows the LME regression (slope = 0.146, p-value < 0.001). C) The
747 number of CRISPR arrays per genome was compared across sampling years within fire-affected
748 or reference soils, with the lines indicating the statistically significant LME regression (fire-
749 affected slope = -0.386, p-value < 0.001; reference p-value = 0.165).
750
751

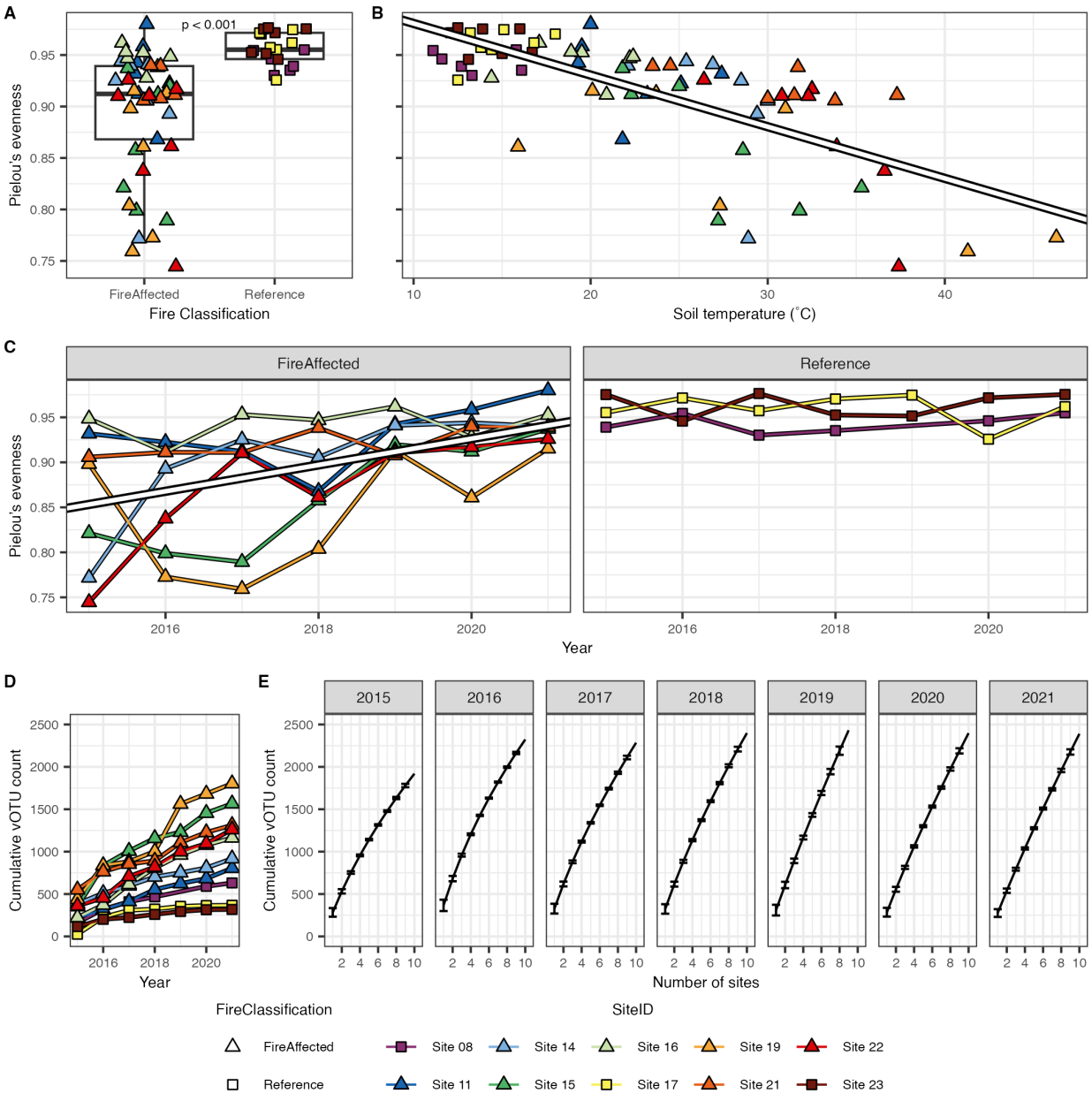
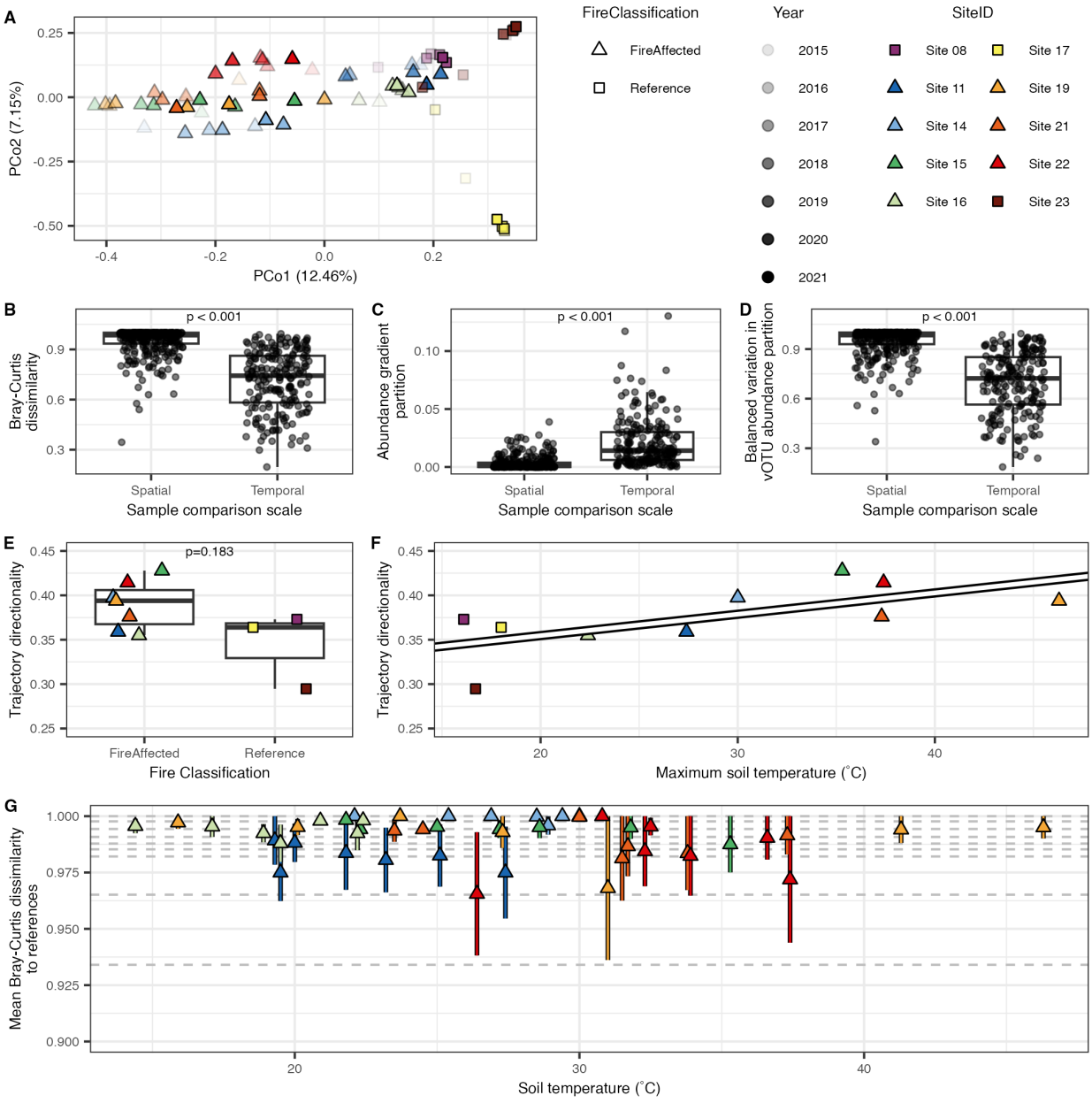


Figure 3: The vOTU community evenness was higher in fire-affected soils than reference and decreased as these fire-affected soils cooled. vOTU community evenness was measured as Pielou's evenness. A) The vOTU community evenness was compared between fire-affected and reference soils with the p-values of the Wilcoxon rank sum test indicated. B) The vOTU community evenness was compared across soil temperature (˚C) regardless of fire classification with the line indicating LME regression (slope = -5.02E-3, p-value < 0.001). C) The vOTU

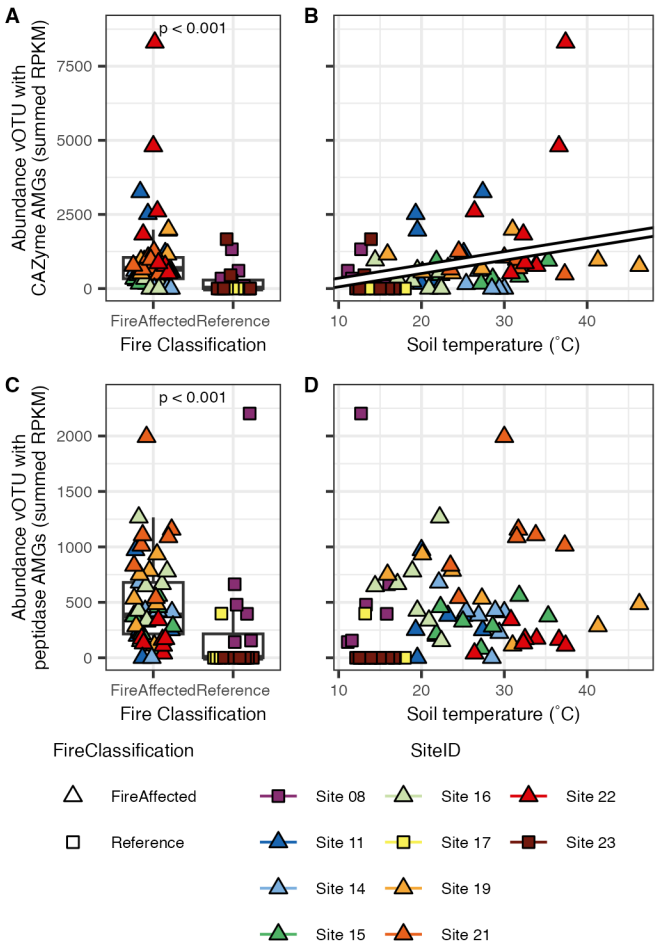
community evenness was compared across sampling years within fire-affected or reference soils with the line indicating statistically significant LME regression (fire-affected slope = 1.46E-2, p-value < 0.001; reference p-value = 0.875). D) Cumulative vOTUs were identified at each site over time, with each successive sampling year. E) Cumulative vOTUs were identified over space, with additional site sampling within each sampling year, with the line indicating mean unique vOTU count and error bars indicating standard error across all sets of sites.



767

768 **Figure 4:** Variation in vOTU diversity, as measured by Bray-Curtis dissimilarity, was explained
769 by fire classification and sampling year. The community directionality over time was higher in
770 fire-affected than in reference soils and increased with increased temperature, an indicator of
771 disturbance intensity. A) The principal coordinate analysis ordination with site ID, fire
772 classification, and sampling year indicated. The axis labels give the percent of the variation in
773 vOTU diversity that is explained by each axis. B) The Bray-Curtis dissimilarity was compared

across samples differing spatially (*i.e.,* different sites but same sampling year) or temporally (*i.e.,* same site but different sampling years). C) The partition of Bray-Curtis dissimilarity that was explained by an abundance gradient was compared across samples differing spatially or temporally, which is similar to community nestedness. D) The partition of Bray-Curtis dissimilarity that was explained by balanced variation in vOTU abundance was compared across samples differing spatially or temporally, which is similar to community turnover. E) The community trajectory directionality for each site over time compared between fire-affected and reference soils. F) The community trajectory directionality for all sites regardless of fire classification was compared against maximum soil temperature. The line represents the linear regression (slope = 2.41E-3, R^2 = 0.379, p-value = 0.034). G) The mean community dissimilarity between each fire-affected site and the three reference sites was compared across fire-affected soil temperature in the corresponding sampling year. The error bars indicate standard error (n=3) while the dashed lines indicate the community dissimilarity between the reference sites within sampling years. There was no statistically significant relationship across dissimilarity between fire-affected and reference soil and soil temperature (LME: p-value = 0.287). P-values included in the figures indicate the Wilcoxon rank sum tests.



792

793 **Figure 5:** The abundances of vOTU-encoded and annotated CAZyme and peptidase AMGs were
794 higher in the fire-affected soils than in the reference soils, and there was a positive relationship
795 between CAZyme encoding vOTU abundance and soil temperature. The abundances are
796 measured as summed RPKM. The CAZyme-encoding vOTU abundances were compared A)
797 between fire-affected and reference soils and B) across soil temperature, where solid line
798 represents the LME regression (slope = 44.9, p-value = 0.040). The peptidase-encoding vOTU
799 abundances were compared C) between fire-affected and reference soils and D) across soil
800 temperature. There was no statistically significant relationship between peptidase-encoding
801 vOTU abundances and temperature (LME: p-value = 0.791). The p-values for Wilcoxon rank
802 sum tests between fire-affected and reference soils are indicated above the points for box plots.