Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil

Evan P. Starr, Erin E. Nuccio, Jennifer Pett-Ridge, Jillian F. Banfield, and Mary K. Firestone

Viruses impact nearly all organisms on Earth, with ripples of influence in agriculture, health, and biogeochemical processes. However, very little is known about RNA viruses in an environmental context, and even less is known about their diversity and ecology in soil, 1 of the most complex microbial systems. Here, we assembled 48 individual metatranscriptomes from 4 habitats within a planted soil sampled over a 22-d time series: Rhizosphere alone, detritosphere alone, rhizosphere with added root detritus, and unplanted soil (4 time points and 3 biological replicates). We resolved the RNA viral community, uncovering a high diversity of viral sequences. We also investigated possible host organisms by analyzing metatranscriptome marker genes. Based on viral phylogeny, much of the diversity was Narnaviridae that may parasitize fungi or Levirviridae, which may infect Proteobacteria. Both host and viral communities appear to be highly dynamic, and rapidly diverged depending on experimental conditions. The viral and host communities were structured based on the presence of root litter. Clear temporal dynamics by Levirviridae and their hosts indicated that viruses were replicating. With this time-resolved analysis, we show that RNA viruses are diverse, abundant, and active in soil. When viral infection causes host cell death, it may mobilize cell carbon in a process that may represent an overlooked component of soil carbon cycling.

Author contributions: E.P.S., E.E.N., J.P.R., and M.K.F. designed research; E.P.S. and E.E.N. performed research; E.P.S., E.E.N., and J.F.B. analyzed data; and E.P.S. and J.F.B. wrote the paper.

Reviewers: S.W.W., The University of Tennessee; and K.E.W., The College of William and Mary.

The authors declare no competing interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. MN032676–MN036331 and MK945893–MK946421). Read files can be accessed through the Joint Genome Institute genome portal using GOLD study ID: Gs0110148. Supplementary datasets 1–6 are available at Figshare: https://figshare.com/projects/Metatranscriptomic_reconstruction_reveals_RNA_viruses_with_the_potential_to_shape_carbon_cycling_in_soil/63722.

Significance

The diversity and ecology of RNA viruses is severely understudied in complex environments. Here we investigate the diversity and community patterns of soil RNA viruses by analyzing assembled metatranscriptomes. We tracked RNA viral and host communities for 22 d in 2 soil environments central to carbon cycling, the rhizosphere and detritosphere. This work is an important step toward understanding the factors that drive RNA viral communities. The main hosts in our system may be Fungi and Proteobacteria; this is in contrast to the ocean, where diatoms and other single cellular eukaryotes are primary hosts for RNA viruses. These results greatly expand the known diversity of viruses and contribute to understanding microbial interactions in soils with major implications for carbon cycling.
truth we have much to learn about the processes influencing soil carbon stabilization.

Understanding the diversity and ecology of soil viruses may contribute to advances in biotechnology. Viral genomes have been mined for biopesticides and self-assembling nanomaterials (25, 26). Viruses have been proposed, and used, as biocontrol agents for controlling invasive organisms, including fire ants and moths (27, 28). Viruses are also being investigated as biocontrol agents for devastating plant pathogens, such as Fusarium sp., Botrytis cinerea, and Rosellinia necatrix (10, 29–34). Novel, environmentally derived viruses may be a source for new biotechnology tools and biocontrol agents.

Here, we used assembled metatranscriptomic data from a California annual grassland soil to reconstruct RNA viral genome sequences and constrain their possible hosts. We searched the assembled sequences for the RNA-dependent RNA polymerase (RdRp), a conserved gene found in all RNA viruses that lack a DNA stage (generally referred to as Baltimore type III–V RNA viruses). Although recent work tracing the deep evolutionary history of RNA viruses has merged viral families into supergroups (35–37), our focus was not on resolving deep phylogenetic viral branches. Rather, our objectives were to investigate fine-scale diversity of RNA viruses, their putative hosts, and their possible functional importance in soil. Thus, we relied heavily on Shi et al. (36), the most comprehensive metatranscriptomic study of RNA virus diversity so far, for taxonomic classification and to delineate viruses from previously undocumented clades.

The diversity of soil eukaryotes that may serve as hosts for RNA viruses remains largely understudied. Generally, RNA viruses are known to infect fungi, plants, animals and the many clades of single-celled eukaryotes, in addition to some Proteobacteria. Soil eukaryotic studies have relied heavily on primer-based sequencing or visual classification, methods that can impart biases and miss novel organisms. Using the genomic information contained in our assembled metatranscriptomes, we identified many clades of eukaryotes without reliance on primers or microscopy. In contrast to many other environmental viral studies, which sequence extracted viral particles, we sequenced whole samples, which included viral genomes and possible host transcripts. This method allowed us to explore both extracellular viruses that would be sequenced in the standard method and viruses within host cells, such as during the infection process or latent infections. We tracked both viral and eukaryotic communities in key soil habitats: The rhizosphere (soil influenced by the root), detritosphere (soil influenced by decaying particulate organic matter; bulk+litter), a combination of the 2 (rhizosphere+litter), and unamended soil (bulk) over time. Using a relatively short sampling timescale (3, 6, 12, and 22 d) allowed us to investigate viral and host community dynamics. Our research places direct constraints on the timescales for virus dynamics and provides a genomic view of RNA viruses in soil.

Results

Experimental Design. This analysis utilized data generated by a previous study on microbial niche differentiation in soil habitats (38). Briefly, wild oat (Avena fatua), an annual grass common in Mediterranean climates, was grown in microcosms with a sidecar, allowing us to track root age and sample directly from the rhizosphere (39). The experimental setup used microcosms, half of which included soil mixed with dried ground A. fatua root litter and the other half contained soil without litter amendment. All microcosms contained bulk soil bags, which included roots, providing control soil uninfluenced by the root. Once A. fatua was mature, roots were allowed to enter the sidecar and the growth of individual roots was tracked. We destructively harvested rhizosphere soil (and paired bulk soil) that had been in contact with the root for 3, 6, 12, and 22 d. In total, we sampled paired rhizosphere and bulk samples from 4 time points with 2 treatments (with and without litter), with 3 biological replicates, for a total of 48 samples for metatranscriptome sequencing. Using rRNA-depleted RNA, we sequenced a total of 408 Gbp (average of 8.7 Gbp per sample).

Eukaryotic RNA Viruses. We used profile hidden Markov models (HMMs) to search our assembled metatranscriptomes and found a total of 3,884 unique viral RdRp sequences (dereplicated at 99% amino acid sequence identity: AAI). This includes 1,350 predicted RNA bacteriophage (phage) viruses that infect bacteria, and 2,534 predicted viruses that infect eukaryotes. Our eukaryotic viruses group into 15 major clades that span the majority of known viral diversity (Fig. 1). Many were included into the supergroup-like clades defined by Shi et al. (36). For the remainder, we constructed phylogenetic trees to define additional viral families. Overall, in trees that include both existing and newly generated sequences, we noted a strong grouping of our RNA viral sequences into “fans of diversity,” much like the seed head of a dandelion. Many of these fans included a single reference sequence previously used to propose a new viral family. For example, in the Hepe-Virga, sequences grouped with the Agaricus bisporus virus 16 [proposed family Ambsetviridae (40)], in the Astro-Pody clad sequences group with Bifivirus UC1 from wastewater (37) and in the Partiti-Picobirna, sequences cluster with Purpureocillium lilacinum nonsegmented virus 1 (41). We substantially expanded the Barnaviridae (Luteo-Sobemo superfamily) and Mynonaviridae (Tombus-Noda superfamily) families (42, 43), which replicate in fungi, and the newly proposed Zhaovirus and Qivovirus families, which are found in invertebrates (36).

We predict that fungi are the most common hosts for many of our reported RNA viruses (mycoviruses) (SI Appendix, Fig. S1). We are most confident when they fall into the Barnaviridae, Megabirnaviridae, Quadirviridae, and mitoviruses, groups currently thought to only infect fungi (44, 45). The most frequently encountered virus in our dataset, accounting for over 50% of the eukaryotic viral strains identified, came from the mitovirus genus in the Narnaviridae family. Mitoviruses are linear single-stranded RNA viruses that replicate within fungal mitochondria and spread vertically through spores and horizontally through hyphal fusion (33, 46). We suspect that most of the mitoviruses we detect are infecting fungi because, although some mitoviral sequences have been found integrated into the genomes of plants, they are frequently truncated and not transcribed (46, 47). Mitoviruses were also the most abundant viral clade in every sample (SI Appendix, Fig. S2).

Until recently, the Narnaviridae group, which includes mitoviruses, were thought to only encode an RdRp, but a recent discovery suggests some narnaviruses (only distantly related to mitoviruses) encode additional proteins, including capsids and helicases (36). We identified some mitovirus genomes with possible additional genes, which could play a part in infection efficiency or infection of new hosts. However, the majority of mitoviruses we identified contained only a single RdRp gene. We predicted several additional proteins on some near-complete mitoviral genomes (SI Appendix, Fig. S3). These putative genes are small (average 79 amino acids), and functions could not be predicted for them. Sometimes the additional genes are predicted to be transcribed in the same direction as RdRp and in other cases they are predicted to be transcribed in the opposite direction (SI Appendix, Fig. S3).

We reconstructed many sequences for viruses that likely infect eukaryotes other than fungi. For the Picorna–Calici, hosts are likely vertebrates, insects, algae, and plants, based on viral phylogenetic placements (48). In the Tombus-Noda tree, many of the RdRps group with umbraviruses, well-recognized plant viruses. Other RdRp sequences group with sequences from complex environmental samples where the hosts are unknown.

Starr et al. | PNAS | December 17, 2019 | vol. 116 | no. 51 | 25901
RNA viruses from some previously defined superfamilies were conspicuously absent. We did not identify any soil RNA viruses belonging in the Nidovirales-like, Reoviridae-like or Orthomixoviridae-like superfamilies. The absence of Reoviridae is interesting, as these RNA viruses can infect fungi (49). We also did not confidently identify any Ortervirales, retroviruses that replicate with a DNA intermediate. We identified many reverse transcripases, but none of the corresponding scaffolds encoded capsid proteins, so the sequences may be retrotransposons or fragments of retroviruses.

We classified 2 clades within the Bunya-Arena viral superfamily as novel, given that they exhibit sequence divergence comparable to that which separates known RNA viral families (SI Appendix). In both cases, the scaffolds only encode the polymerase with a Bunyavirus RNA-dependent RNA polymerase domain. This genome structure is shared by Ixodes scapularis-associated virus 6 (36) and Ixodes scapularis-associated virus 3 (50).

**Eukaryotic Hosts.** As viruses replicate within their host, changes in virus abundance levels implicate introduction of new vectors or hosts, shifts in infected and uninfected host abundance levels, changes in infected host physiology, or changes in host susceptibility to infection. To better understand the diversity and ecology of RNA viral hosts and carriers in soil, we used the mitochondrially encoded cytochrome C oxidase subunit 1 (Cox1) gene as a marker to define the eukaryotic populations present (51–56). However, mitochondrial transcription is determined by metabolic activity of a cell (57) and can be impacted by the switch to a symbiotic state [e.g., in arbuscular mycorrhizal fungi (58)] or nutrient sensing and hormonal signals [e.g., in animals (59)]. Thus, the number of reads mapped to the Cox1 genes is determined by the number of organisms present and the organism’s activity.

We identified 726 eukaryotic Cox1 sequences and clustered them at 98% AA1 to approximate a species-level view. The eukaryotes we discovered span the diversity of known soil organisms. Not surprisingly, sequences from *A. fatua* were the most abundant Cox1 transcripts in many samples. In some samples, the most abundant Cox1 were from an *Enchytraeidae*-related worm or an *Amoeboboa* sp. Other samples were dominated by a mixture of fungi, *Amoeboboa*, Viridiplantae, and unknown eukaryotes.

*Amoeboboa* were the most diverse eukaryotic clade in this dataset, >25% of the identified eukaryotes, followed by species of fungi and unknown eukaryotes. The many unknown eukaryotes reflect the lack of environmental Cox1 sequences from micro- and mesoeukaryotes in public databases. For this reason, we reconstructed 18S rRNA gene sequences for classification, as they are better represented than Cox1 in public databases. However, the transcripts we analyzed had been depleted in bacterial and plant ribosomal RNA (via Rib-Zero rRNA Removal kits), which undoubtedly influenced the composition and relative amount of other rRNA sequences. As such, the identified 18S rRNA sequences could not be used for abundance measurement or as a quantitative measure of diversity, only for identification of eukaryotes not identified using COX1. In total we identify 521 distinct species of fungi and unknown eukaryotes. The many unknown eukaryotes span the diversity of known soil organisms.

**Eukaryotic Virus and Host Ecology.** As viruses replicate within their host, changes in virus abundance levels implicate introduction of new vectors or hosts, shifts in infected and uninfected host abundance levels, changes in infected host physiology, or changes in host susceptibility to infection. To better understand the diversity and ecology of RNA viral hosts and carriers in soil, we used the mitochondrially encoded cytochrome C oxidase subunit 1 (Cox1) gene as a marker to define the eukaryotic populations present (51–56).

We compared the abundances of eukaryotic Cox1 transcripts in rhizosphere-litter, bulk-litter, and rhizosphere samples to the bulk samples in order to identify species statistically enriched in each case (SI Appendix, Fig. S5). The results showed that the presence of root litter and rhizosphere enriched for many *Amoeboboa* and fungi. However, enrichment patterns indicate that litter had a greater selective force on more individual species than the presence of a growing root.

The eukaryotic RNA viral community was influenced by some of the experimental treatments. Root litter had a significant effect on the eukaryotic RNA viral community (PERMANOVA...
used a marker gene approach to find RdRps for these RNA phage viruses. All previously documented Leviviridae infect Proteobacteria, which in-cluding those infecting Pseudomonas sp. (60, 61), and the Leviviridae, which infect Gamma-proteobacteria and Alphaproteobacteria (62–64). We used a marker gene approach to find RdRps for these RNA phage in the assembled transcripts. After dereplication, we identified 12 Cystoviridae RdRp sequences and 1,338 unique Leviviridae RdRp sequences. This is a significant increase in the known diversity of this group, as there are currently just over 200 Leviviridae RdRp sequences in public databases. Some of our sequences group with Allolevivirus and Levivirus, well-studied genera, but importantly, other novel Leviviridae sequences resolved new clades (Fig. 1).

The specific Leviviridae taxa enriched in bulk soil are relatively phylogenetically novel (SI Appendix, Fig. S8) and have unconventional genetic architectures based on metagenomic reconstruction. The most abundant Leviviridae sequence in 46 of 48 samples groups into a predominantly novel branch. The near-complete 4,668-bp genome of this abundant RNA phage encodes 4 nonoverlapping genes, as opposed to the frequently overlapping sequences. This is a significant increase in the known diversity of this group, as there are currently just over 200 Leviviridae RdRp sequences in public databases. Some of our sequences group with Allolevivirus and Levivirus, well-studied genera, but importantly, other novel Leviviridae sequences resolved new clades (Fig. 1).

The specific Leviviridae taxa enriched in bulk soil are relatively phylogenetically novel (SI Appendix, Fig. S8) and have unconventional genetic architectures based on metagenomic reconstruction. The most abundant Leviviridae sequence in 46 of 48 samples groups into a predominantly novel branch. The near-complete 4,668-bp genome of this abundant RNA phage encodes 4 nonoverlapping genes, as opposed to the frequently overlapping sequences. This is a significant increase in the known diversity of this group, as there are currently just over 200 Leviviridae RdRp sequences in public databases. Some of our sequences group with Allolevivirus and Levivirus, well-studied genera, but importantly, other novel Leviviridae sequences resolved new clades (Fig. 1).

We used ribosomal protein S3 (rpS3) phylogeny to identify possible hosts for Leviviridae viruses. All previously documented Leviviridae infect Proteobacteria, so we narrowed our analysis to these bacteria, which represent a possible pool of hosts for the Leviviridae, although other clades cannot be ruled out. In our samples, we identified 355 species of Proteobacteria, which differ in abundance between rhizosphere, bulk, rhizosphere+ litter, and bulk-litter treatments (PERMANOVA P < 9.999e-05) (SI Appendix, Fig. S10) after only 3 d of root growth (SI Appendix, Fig. S10B). The Leviviridae communities also diverged within 3 d and changed over time (PERMANOVA P < 0.02) (Fig. 3A). The communities separated based on the presence or absence of root litter (PERMANOVA P < 9.999e-05), whereas the influence of growing roots was undetectable (PERMANOVA P < 0.09) (Fig. 3B).

We used ribosomal protein S3 (rpS3) phylogeny to identify possible hosts for Leviviridae viruses. All previously documented Leviviridae infect Proteobacteria, so we narrowed our analysis to these bacteria, which represent a possible pool of hosts for the Leviviridae, although other clades cannot be ruled out. In our samples, we identified 355 species of Proteobacteria, which differ in abundance between rhizosphere, bulk, rhizosphere+ litter, and bulk-litter treatments (PERMANOVA P < 9.999e-05) (SI Appendix, Fig. S10) after only 3 d of root growth (SI Appendix, Fig. S10B). The Leviviridae communities also diverged within 3 d and changed over time (PERMANOVA P < 0.02) (Fig. 3A). The communities separated based on the presence or absence of root litter (PERMANOVA P < 9.999e-05), whereas the influence of growing roots was undetectable (PERMANOVA P < 0.09) (Fig. 3B).

**Discussion**

To our knowledge, no prior study has genomically investigated the ecology and diversity of RNA viruses in soils. Using assembled metatranscriptomes, we uncovered a vast diversity of RNA viruses compared the number of enriched eukaryotes (SI Appendix, Fig. S7). This may be due to the substantial heterogeneity in viral abundance patterns, even within replicates. The RNA viruses displayed a higher level of microheterogeneity than the eukaryotes. In contrast, transcripts for individual eukaryotes were more ubiquitous across samples. The viral strains most frequently enriched relative to bulk soil were from the mitoviruses. This is unsurprising, as the abundance of the obligately intermitochon-drial mitoviruses is tied to the abundance and activity of the host and fungi responded strongly to the specific treatments.

To explore possible correlations between the hosts and RNA viruses we created a cooccurrence network. The network was constructed from positive cooccurrence edges between a eukaryote and virus derived from the network analysis (SI Appendix, Fig. S11). The number of statistically significant correlations is the number of positive cooccurrence edges between a eukaryote and virus derived from the network analysis (SI Appendix, Fig. S11).

**Table 1. Number of eukaryotes identified based on marker genes and significant correlations between eukaryotes and RNA viruses**

| Clade            | Cox1 18S rRNA* | Narnaviridae Bunya-Arena Tombus-Noda Luteo-Sobemo Toti-Chryso Picorna-Calici Leviviridae Parti-Picobirna |
|------------------|----------------|----------------------------------------------------------------------------------------------------------------|
| Amoebozoa        | 193            | 295 14 16 3 2                                                                                                     |
| Fungi            | 174            | 33 15 11 2 1                                                                                                      |
| Unknown eukaryotes | 108         | 14 1                                                                                                               |
| Metazoa          | 85             | 28                                                                                                                 |
| Stramenopiles    | 55             | 17 11 6 3                                                                                                          |
| Alveolata        | 54             | 14                                                                                                                 |
| Heterolobosea    | 11             | 41                                                                                                                 |
| Nucleariida      | 11             | 0 11 2                                                                                                             |
| Cryptophyta      | 9              | 1                                                                                                                   |
| Euglenozoa       | 7              | 28                                                                                                                  |
| Viridiplantae    | 6              | 2                                                                                                                   |
| Choanoflagellida | 5              | 8 1 3 1                                                                                                             |
| Rhizaria         | 3              | 21                                                                                                                  |
| Apusozoa         | 2              | 1                                                                                                                   |
| Alveida          | 2              | 0                                                                                                                   |
| Ichthysoporea    | 1              | 0                                                                                                                   |
| Centroheliozoa   | 0              | 11                                                                                                                  |
| Malawimonadida   | 0              | 6                                                                                                                   |
| Jakobida         | 0              | 1                                                                                                                   |

*Note that the absolute values may be skewed by rRNA depletation.*
viruses (>3,000 sequences) from a single grassland soil and discovered several possible unique families. Some viruses grouped phylogenetically with viral families previously proposed based on only a single member, indicating soils may hold much of the diversity of these understudied groups. The viruses we reconstructed came from nearly all known groups of RNA viruses. We identified hundreds of eukaryotes and bacteria that may represent RNA viral hosts or vectors that can pass infections among plants and other organisms. We know that the transfer of viral agents can have agriculturally relevant effects (66–68) but there are likely many unexplored impacts on the soil community and host-mediated process, such as nutrient cycling.

Many of the RNA viruses we identified group phylogenetically with fungal viruses (mycoviruses), and fungi appear to be one of the main RNA viral hosts in this soil. The apparent dominance of mycoviruses in our dataset provides a striking contrast to aquatic systems, where Picornavirales dominate (14). The most diverse mycoviruses in our samples are mitoviruses (Narnaviridae) which, despite their simplicity, can have significant impacts on fungal fitness and physiology (33, 69, 70).

Previous studies of environmental RNA virus ecology have primarily addressed marine systems. The RNA viruses of oceans appear to be dominated by Picornavirales, thought to infect diatoms and other single-cell eukaryotes (7, 14, 71). While our samples did contain Picornavirales, they did not appear to be dominant in the soil we sampled. In addition to Picornavirales, some aquatic systems contain plant, human, and livestock RNA viruses that may have

Fig. 2. Nonmetric multidimensional scaling ordination of soil eukaryotic communities in response to 4 soil resource treatments, based on coverage of the Cox1 gene (A. fatua sequences removed) (A), and ordination of eukaryotic RNA viral communities with coverage calculated at the scaffold level (B).

Fig. 3. Nonmetric multidimensional scaling ordination of the Leviviridae community identified in an annual grassland soil exposed to 4 resource treatments, colored by time (A) and treatment (B).
been introduced through runoff; we identified no viruses known to infect humans (14, 72). Some aquatic studies have noted viruses related to RNA fungal viruses, however not at the abundances or diversity we have in our soil (7). Again, in contrast to the diversity of RNA phage that we found in soil, there is only a single report of an RNA phage from an aquatic ecosystem (71). Compared to marine systems, soil virology, especially investigations of RNA viruses, is in its infancy.

In our soil, both eukaryotic RNA viral and eukaryotic communities were impacted by the presence of decaying root litter. By the first sampling time point, the eukaryotic and eukaryote-associated RNA viral communities were different in the presence and absence of litter. Saprophytic fungi appeared to respond favorably to the dead litter biomass. In contrast, dissolved organic compounds exuded by roots may be less accessible to nonfungal soil eukaryotes. Accordingly, roots had little impact on the nonplant eukaryotic community composition. Addition of decaying plant biomass promoted the activity of detritivores, which likely immigrated into the litter from surrounding soil, increased transcription levels, grew, and/or germinated from spores. However, the shifts in the eukaryotic RNA viral community we observed are likely due to proliferation following increased activity of their hosts, although they may have also been transported by vectors.

We conducted a network analysis between the RNA viruses and possible hosts to provide new hypotheses about interacting partners. The correlation between virus and host provides possible associations and fodder for future studies. However, overinterpretation of network analysis results, especially in complex systems, should be avoided (73). Within the RNA virus–host cooccurrence network the fungi had the greatest number of cooccurrence links with the Narnaviridae, which provides another piece of evidence that fungi may be hosts for RNA viruses in this soil. Other eukaryotes had cooccurrence links with the Narnaviridae as well, but it is not possible to tell if these viruses are infecting or simply cooccurring with the eukaryote. The Amoebozoa had the largest number of cooccurrence linkages, including 3 with Leviviridae, which are unlikely to infect the Amoebozoa themselves but may be associated with a cooccurring bacteria. We know of only 1 RNA virus that infects Amoebozoa, a Mononegavirales (Mono-Chu) (74). These results suggest a possible link between Amoebozoa and the viral Bunyaviridae clade. The largest module (cluster of linages) contained a phylogenetically diverse array of eukaryotes and viruses including single virus types connected to multiple clades of hosts; this may represent a group of cooccurring entities that respond to similar environmental or biotic conditions (SI Appendix, Fig. S11).

RNA phage, which replicate in bacteria, have received relatively little study of their ecological context. Here, we provide a glimpse into their ecology and diversity in soil. Given that over half of all of the identified RdRp sequences in our dataset were Leviviridae, we encourage additional studies (in other soil types, land uses, tillage practices, and cover vegetations) to assess whether soils generally host a large diversity of Leviviridae. As Leviviridae do not have a known lysogenic life stage, most of their population changes likely reflect replication in their hosts. Since Leviviridae communities and their hosts became distinctly different within the first 3 d of our experiment, we conclude that these RNA viruses infected and replicated within days. Infections over this timescale likely have dramatic effects on host communities, and thus soil ecology.

RNA viruses and RNA phage were heterogeneously distributed across samples, including replicates. In contrast, eukaryotes and bacteria appeared to have more even transcript abundances across samples. In combination, these observations suggest that virus and phage abundances are not solely determined by the presence of hosts, but also factors—such as sporadic blooms or variation in viral resistance levels in the host population—that lead to patchy viral distribution patterns in soil. These patterns may be analogous to those documented in marine virus blooms, although the exact causes of these events are still unknown (13, 75). In addition, viruses have different release rates and persistence times in soil, which may be impacted by a variety of factors, such as soil type, viral strain, host physiology, microsite of release, clay content, and extracellular enzymes. This effect would be most pronounced for viruses with an extracellular life stage and less impactful on obligate intracellular viruses like the Narnaviridae.

In general, the magnitude of viral impacts on the soil carbon cycle is underexplored. Little research has been done on phage-induced bacterial lysis in soil and even less on viral-induced death of fungi and other eukaryotes, which can contain an equal or greater biomass compared to bacteria in some soils (76). In addition to lysis, viruses of fungi can have subtle yet profound effects on fungal biology, pathogenesis, mating success, toxin production, symbiotic relationships, and other physiological effects (77–79). These effects of viruses on fungal biology could have impacts for fungal functioning and thus soil performance as a whole. In our samples, the high diversity and abundance of identified RNA viruses, combined with their dynamic population changes, indicates that there was substantial viral replication. We hypothesize that proliferation of lytic RNA phage and RNA viruses will have substantial impacts on the form, abundance, and distribution of carbon compounds in soil and on the biology and fitness of soil fungi. For example, lysis of host bacterial cells and viral-induced cell death of eukaryotes will cause release of dissolved low molecular weight carbon compounds. These will likely be quickly consumed by nearby bacteria and much of the carbon respired back to the atmosphere, mostly as CO2. However, as happens when marine snow settles into the deep ocean, a portion of the cellular and viral debris and soluble carbon released may be stabilized in soil. For example, bacterial and eukaryotic lipids and polysaccharides could adhere to mineral surfaces or become occluded within soil aggregates.

Conclusions

By reconstructing soil metatranscriptomes drawn from multiple soil habitats and time points within the context of a controlled experiment we greatly increased the known diversity of RNA viruses. Phylogenetic analyses suggest that fungi are the most common hosts for RNA viruses in the studied grassland soil. While the diversity of hosts for RNA phage remain to be definitively identified, they likely also include Proteobacteria. Shifts in eukaryote, RNA phage, and RNA viral abundances over a few-day period reveal that entire soil communities can rapidly respond to altered resource availability. Our experiments indicate that the form of carbon inputs (root-derived low molecular weight C inputs versus macromolecular carbon compound in litter) may impact eukaryotic and bacterial abundance patterns, and that these in turn may be a major determinant of RNA viral and RNA phage dynamics.

Methods

Experimental Design. Wild oat (A. fatua) was grown in microcosms with a clear sidecar designed to allow access and visual tracking of the soil and rhizosphere (39, 80). Experimental soil was collected from Hopland Research and Extension Center (Hopland, CA). The soil is a fine-loamy, mixed, active, mesic Typic Haploxeralf, supporting dominant stands of Avena species (81, 82). Microcosms were filled with soil at field bulk density and seedlings were planted and grown for 6 wk in the main chamber before the start of the experiment. Six days before the start of the experiment the divider separating the main chamber and the sidecar was removed and the sidecar was packed with soil. The litter-amended microcosms received 0.4 g of dried A. fatua root litter mixed with 50 g of Hopland soil. Bulk soil was contained in 1-μm mesh bags embedded into the microcosm, which
allowed solutes to pass but not roots. Experimental design and soil edaphic characteristics are fully explained in a companion publication, which provides additional details regarding sample collection protocol, RNA extraction and processing, and sequencing (38); what follows is a brief description of the methods.

Sample Collection. The ages of individual roots were tracked to collect rhizosphere soil that had been influenced by the root for 3, 6, 12, and 22 d. Three replicate microcosms were destructively harvested for paired rhizosphere and bulk soil. Rhizosphere soil was cut out from the rest of the soil along the edge of the root hair zone (<2 mm from the main root). Root sections and adhering soil were placed immediately in ice-cold Lifegard Soil Preservation Reagent (MoBio) and soil was agitated off the root by vortexing for 2 min on medium speed and pelleted; bulk soil was treated in the same manner. Pelleted soils were frozen on dry ice and stored at −80 °C.

RNA Extraction. RNA was extracted from 0.5 g of frozen soil using a phenol-chloroform extraction protocol (83). Then Qiagen AllPrep kits were used to separate DNA from RNA. RNA was treated with TURBO DNase (Thermo Fisher Scientific) following the manufacturer’s protocol to remove residual DNA, and was concentrated using an ethanol precipitation.

Sequencing. Metatranscriptome libraries were prepared and sequenced at the Joint Genome Institute. Ribosomal RNA was depleted from 1 μg of total RNA using the Ribo-Zero rRNA Removal Kit (Epicentre) for Plants and Bacteria. RNA was sequenced on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v3, and Illumina’s cBot instrument to generate a cluster flowcell for sequencing. Sequencing of the flowcell used a TruSeq SBS sequencing kit, v3, following a 2 × 150 indexed run recipe.

Sequence Analysis. Reads were trimmed using Sickle (https://github.com/najoshi/sickle) and BBTools (https://sourceforge.net/projects/gembmap/) was used to remove Illumina adapters and trace contaminants. The reads from all 48 samples were assembled individually using IDBA-UD with default settings (84). Genes were predicted using Prodigal in the anonymous mode (85).

To find the host marker genes, Cox1 and rpS3, we used an HMM profile from Pfam (86), Cox1 (PF00115), and HMMS for Bacteria and Archaea (https://github.com/AJProbst/rpS3_trckr) and searched using hmmpress (-E 0.0001) from the HMMER suite (87). The identified proteins were classified using both National Center for Biotechnology Information (NCBI) blast and by making trees. Once classified, we predicted genes again using Prodigal in the single mode with the appropriate translation table. The Cox1 protein sequences were dereplicated and clustered at 98% AAI representing an estimated species-level designation (88, 89). Assembly errors were found and fixed in the scaffolds containing the Cox1 using ra2.py (https://github.com/christophertbrown/fixed_assembly_errors) (90). The bacterial and archaeal rpS3 genes was clustered at 99% AAI, representing species-level differences (91), using USEARCH (-cluster_fast) (92). The Cox1 and rpS3 trees were generated using UPGMA trees and confirm that the protein sequences were then aligned using MAFFT v7.402 (93) with the E-INS-i option on Cipres (94). Then, alignments were trimmed for the conserved domain manually on Geneious and automatically trimmed using trimal (95, 96). The tree was made using RAxML (97) with the JTT protein substitution model (98). The trees were analyzed and figured using iTOL (99). The 18S genes were found and the alignment was generated using susu_tree.py (https://github.com/christophertbrown/bioscripts27), which searches for rRNA genes using an HMM method then the sequences were dereplicated and clustered at 98% nucleic acid identity, again representing a possible species level designation (100, 101), and aligned using SSU-ALIGN (102). Assembly errors were found and fixed in the scaffolds containing the 18S using ra2.py (https://github.com/christophertbrown/fixed_assembly_errors) (90). The tree was generated using the approach described above.

We identified the RNA viral scaffolds using a combined profile HMM approach, using HMMs from Pfam (86) for different types of RdRps. The Leviriviridae were identified using RNA_replicate_B (PF03431) and scaffolding errors were identified with ra2.py. There were no usable HMMs for the Cystoviridae, so we generated our own. We aligned public RdRp sequences from the Cystoviridae using MAFFT v7.402 (E-INS-i) on Cipres and generated an HMM using hmmbuild from the HMMER suite with default settings (87, 93, 94). The Cystoviridae HMM is publicly available on the figshare repository corresponding to this report. We used many publicly available HMMs to find the RdRp of eukaryotic RNA viruses (Mononeg, RNA_pol [PF00946], RdRp_5 [PF07925], Flavi_NNS [PF00972], Bunya_RdRp [PF04196], Mitovir_RNA_pol [PF00172], RdRP_4 [PF00978], RdRP_1 [PF00980], RdRP_2 [PF00984], RdRP_3 [PF00985], RdRP_4 [PF02123], RVT_1 [PF00078], RVT_2 [PF07727], Viral_RdRp_C [PF17501], and Birna_RdRp [PF04197]). The RdRp genes were initially classified using a blast and tree-building method to determine the correct translation table to use for Prodigal in the single mode (85). For mitoviruses we used translation table 4, which to our knowledge is used by all mitoviruses. However, the additional mitoviral genes we predicted may have been incorrectly called if these genomes use a modified genetic code, as occurs in some fungi (103, 104). We examined gene predictions for indications of this (e.g., interruption of the RdRp gene) but as this phenomenon was not identified, no alternative codes were used for gene prediction. RdRp amino acid sequences were dereplicated and clustered at 99% amino acid identity using USEARCH (92). We used previously published alignments for many viral families (36) and added our sequences and key references using the Mafft v7.402 with the seed (93) and E-INS-i options on Cipres (94). For viral clades without published alignments or where the published alignment was inappropriate (Fasiraviridae, Narnaviraviridae, Leviriviridae, and Cystoviridae) we generated our own alignments using reference sequences from NCBI and the same alignment and tree-building steps as described above.

To identify lysis proteins in the Leviriviridae genomes we used the Genedius ORF prediction tool to find all possible ORFs (96). The amino acid sequences were run through PSORTb v3.0.2 with the gram-negative setting to find possible lysis proteins (105). To obtain coverage values for the viruses, we mapped against the entire scaffold containing the RdRp gene. For the presumed hosts we only mapped against the RdRp using MAFFT. Raw reads were then mapped using Bowtie2 (−sensitive and−frg 200, 300, 300 options), then reads were filtered for 2 mismatches using calculate_breadth.py (https://github.com/banfieldlab/mattom-public-scripts/blob/master/calculate_breadth.py) (106). Coverage values were converted to read counts and normalized with DESeq2 (107). Ordinations were generated from DESeq2 normalized count data in R. The data were ordinated using nonmetric multidimensional scaling (R package: vegan) and significantly different clusters were determined using adonis (108). Using DESeq2 we determined the viral and eukaryotic habitat enrichments in the treatments (rhizosphere-litter, bulk-litter, and rhizosphere) relative to bulk soil at each time point. We also conducted the opposite test, identifying enrichments in bulk soil by comparing normalized abundance to each treatment (rhizosphere-litter, bulk-litter, and rhizosphere) at each time point. P values were corrected for multiple comparisons using a Benjamini–Hochberg correction (109).

Network analysis was conducted using the DESeq2 normalized coverage for the eukaryotic viruses and bacterial viruses and the eukaryotic Cox1 and bacterial rpS3. These values were filtered for coverage greater than 1,000. Spearman’s rank correlation coefficient values were computed in R using the Hmisc package; the P values were corrected for multiple testing (109, 110). The data were then filtered to contain only organisms present in 20 or greater samples, a corrected P value of 0.0001 or less, a correlation value of 0.7 or greater, and removed all host-host correlations. These data were then fed into Cytoscape and a network was constructed using default settings (111). Only clusters that contained at least 1 virus as the dependent variable and 1 host as the independent variable were analyzed and presented.

ACKNOWLEDGMENTS. This work was made possible by the support and expertise of the personnel at the Hopland Research and Extension Center (Hopland, CA). Plants were grown at the University of California, Berkeley Donald Tract Greenhouse facility. Sampling efforts and technical expertise of the personnel at the Hopland Research and Extension Center were provided by Jonathan Schilling and the MycoClub at the University of Minnesota. We thank the reviewers for their constructive comments and input. This research was supported by the US Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Awards DE-SC0010570 and DOE-SC0016247 to M.K.F., DOE-SC10010566 (to J.F.B.), and SCW1589 and SCW1632 (to M.K.F.). E.P.S. was supported by a National Science Foundation fellowship, (to M.K.F.), DOE-SC10010566 (to J.F.B.), and SCW1589 and SCW1632 (to J.P.-R.). Work conducted at Lawrence Livermore National Laboratory was contributed under the auspices of the US Department of Energy under Contract DE-AC52-07NA27344. Sampling was conducted as part of Community Sequencing Awards 1487 (to J.P.-R.) and 1472 (to M.K.F.) awarded to the Department of Energy Office of Science User Facility, supported under Contract DE-AC02-05CH11231.
72. A. Djikeng, R. Kuzmickas, N. G. Anderson, D. J. Spiro, Metagenomic analysis of RNA viruses in a fresh water lake. PLoS One 4, e7264 (2009).
73. A. Carr, C. Diener, N. S. Baliga, S. M. Gibbons, Use and abuse of correlation analyses in microbial ecology. ISME J. 13, 2647–2655 (2019).
74. R. G. Bird, T. F. McCaul, The rhabdoviruses of Entamoeba histolytica and Entamoeba invadens. Ann. Trop. Med. Parasitol. 70, 81–93 (1976).
75. J. R. Seymour, L. Seuront, M. Doubell, R. L. Waters, J. G. Mitchell, Microscale patchines of viroplankton. J. Mar. Biol. Assoc. U. K. 86, 551–561 (2006).
76. Y. M. Bar-On, R. Phillips, R. Milo, The biomass distribution on Earth. Proc. Natl. Acad. Sci. U.S.A. 115, 6506–6511 (2018).
77. L. M. Marquez, R. S. Redman, R. J. Rodriguez, M. J. Roossinck, A virus in a fungus in a plant: Three-way symbiosis required for thermal tolerance. Science 315, 513–515 (2007).
78. L. Zhang, R. A. Baasiri, N. K. Van Alfen, Viral repression of fungal pheromone precursor gene expression. Mol. Cell. Biol. 18, 953–959 (1998).
79. N. Rodriguez-Coqueiro et al., A new wine Saccharomyces cerevisiae killer toxin (Kius), encoded by a double-stranded ma virus, with broad antifungal activity is evolutionarily related to a chromosomal host gene. Appl. Environ. Microbiol. 77, 1822–1832 (2011).
80. K. M. DeAngelis et al., Selective progressive response of soil microbial community to wild oat roots. ISME J. 3, 168–178 (2009).
81. T. Whitman et al., Microbial community assembly differs across minerals in a rhizosphere microcosm. Environ. Microbiol. 20, 4444–4460 (2018).
82. S. A. Placella, E. L. Brodie, M. K. Firestone, Rainfall-induced carbon dioxide pulses result from sequential resuscitation of phylogenetically clustered microbial groups. Proc. Natl. Acad. Sci. U.S.A. 109, 10931–10936 (2012).
83. R. L. Barnard, C. A. Osborne, M. K. Firestone, Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. ISME J. 7, 2229–2241 (2013).
84. Y. Peng, H. C. M. Leung, S. M. Yiu, F. Y. L. Chin, IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28, 1420–1428 (2012).
85. D. Hyatt et al., Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119 (2010).
86. S. E. Gebali et al., The Pfam protein families database in 2019. Nucleic Acids Res. 47, D427–D432 (2019).
87. R. D. Finn, J. Clements, S. R. Eddy, HHMNER web server: Interactive sequence similarity searching. Nucleic Acids Res. 39 W29–W37 (2011).
88. P. E. Court, M. C. Pierrat, J. Garbaye, Temporal changes in the ectomychorial community in two soil horizons of a temperate oak forest. Appl. Environ. Microbiol. 74, 5792–5801 (2008).
89. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
90. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. Roy. Statist. Soc. 57, 289–300 (1995).
91. F. E. Harrell, Package “rms,” Version 4.2-0 (2019). https://github.com/harrelfe/rms. Accessed 26 June 2019.
92. P. Shannon et al., Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).
93. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461 (2010).
94. M. A. Miller, W. Pfeiffer, T. Schwartz, Creating the CIPRES science gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), (Institute of Electrical & Electronic Engineers, New York, 2010), pp 1–8.
95. S. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973 (2009).
96. M. Keane et al., Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649 (2012).
97. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461 (2010).
98. D. T. Jones, W. R. Taylor, J. M. Thornton, The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282 (1992).
99. I. Letunic, P. Bork, Interactive tree of life (iTOl) v3: An online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242–W245 (2016).
100. K. Hadziavdic et al., Characterization of the 18S RNA gene for designing universal eukaryote specific primers. PLoS One 9, e87624 (2014).
101. S. Wu, J. Xiang, Y. Yu, Taxonomic resolutions based on 18S rRNA genes: A case study of subclass copepoda. PLoS One 10, e0131498 (2015).
102. E. P. Nawrocki, “Structural RNA homology search and alignment using covariance models,” PhD thesis; 282, Washington University in St. Louis, MO (2009).
103. L. Forget, J. Ustinova, Z. Wang, V. A. R. Hsu, B. F. Lang, Hyaloraphidium curvatum: A linear mitochondrial genome, trNA editing, and an evolutionary link to lower fungi. Mol. Biol. Evol. 19, 310–319 (2002).
104. M. J. Laforest, I. Roewer, B. F. Lang, Mitochondrial tRNAs in the lower fungus Spizellomyces punctatus: tRNA editing and UAG ‘stop’ codons recognized as leuine. Nucleic Acids Res. 25, 626–632 (1997).
105. N. Y. Yu et al., PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26, 1608–1615 (2010).
106. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bovtirse 2. Nat. Methods 9, 357–359 (2012).
107. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
108. J. Oksanen et al., Package “vegan” Title Community Ecology Package, version 2.5-4 (2019). Available at: https://cran.r-project.org/web/packages/vegan/vegan.pdf. Accessed 2 April 2019.
109. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. Roy. Statist. Soc. 57, 289–300 (1995).
110. L. Forget, J. Ustinova, Z. Wang, V. A. R. Hsu, B. F. Lang, Hyaloraphidium curvatum: A linear mitochondrial genome, trNA editing, and an evolutionary link to lower fungi. Mol. Biol. Evol. 19, 310–319 (2002).
111. M. J. Laforest, I. Roewer, B. F. Lang, Mitochondrial tRNAs in the lower fungus Spizellomyces punctatus: tRNA editing and UAG ‘stop’ codons recognized as leuine. Nucleic Acids Res. 25, 626–632 (1997).
112. N. Y. Yu et al., PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26, 1608–1615 (2010).