Substantial differences in soil viral community composition within and among four Northern California habitats

Devyn M. Durham[1,*], Ella T. Sieradzki[1,*], Anneliek M. ter Horst[1], Christian Santos-Medellín[1], C. Winston A. Bess[1], Sara E. Geonczy[1], Joanne B. Emerson[#,1]

[1] Department of Plant Pathology, University of California Davis, CA, USA

* Equal contribution

Corresponding author email: jbemerson@ucdavis.edu

Running title: Viral community heterogeneity across soil habitats

Abstract

Viruses contribute to food web dynamics and nutrient cycles in diverse ecosystems, yet the biogeographical patterns that underlie these viral dynamics are poorly understood, particularly in soil. Here, we identified trends in soil viral community composition in relation to habitat, moisture content, and physical distance. We generated 30 soil viromes from four distinct habitats (wetlands, grasslands, woodlands, and chaparral) by selectively capturing virus-sized particles prior to DNA extraction, and we recovered 3,432 unique viral 'species' (vOTUs). Viral communities differed significantly by soil moisture content, with viral richness generally higher in wet compared to dry soil habitats. However, vOTUs were rarely shared between samples, including replicates <10 m apart, suggesting that soil viruses may not disperse well and that future soil viral community sampling strategies may need to account for extreme community differences over small spatial scales. Of the 19% of vOTUs detected in more than one sample, 93% were from the same habitat and site, suggesting greater viral community similarity in closer proximity and under similar environmental conditions. Within-habitat differences indicate that extensive sampling would be required for rigorous cross-habitat comparisons, and results belie emerging paradigms of higher viral activity in wet soils and soil viral community spatial heterogeneity.

Introduction

Viruses are abundant across Earth's ecosystems, contributing to microbial dynamics and biogeochemical cycles, yet they remain understudied, particularly in terrestrial habitats [1, 2].

Soil viral abundance measurements vary substantially, ranging from nearly zero in dry deserts to over 10^9 virus-like particles per gram in wetlands [3]. In the better studied oceans, viruses kill approximately 20% of microbial biomass daily, impacting nutrient and energy cycles [4], and recent work suggests that viruses may be similarly important in terrestrial ecosystems [3, 5–12]. For example, viruses have been suggested to affect carbon cycling in thawing permafrost peatlands by preying on methanogens and methanotrophs and by encoding glycoside hydrolases to break down complex carbon into simple sugars [2]. Soil viral activity has been demonstrated [2, 5, 13], and soil viral communities can be spatially structured [14–16]. Despite these emerging ecological patterns, comparisons of soil viral diversity within and across habitats are limited.

Here, we compared dsDNA viromes (viral-size fraction metagenomes) [14] from four distinct habitats (wetlands, grasslands, chaparral shrublands and oak woodlands) across five UC Davis Natural Reserves field sites within Northern California. We compared viral species (vOTU) richness, vOTU detection patterns, and viral community beta-diversity, according to habitat type, soil properties, and spatial distribution to better understand the fundamental relationships between soil viruses and the ecosystems that they inhabit.

Results and Discussion

To compare soil viral community composition within and across terrestrial habitats on a regional scale, viromes were generated from 34 near-surface (top 15 cm) soil samples, with a total of 30 viromes included in downstream ecological analyses (see Supplementary Methods). The analyzed samples were collected from four distinct habitats (wetlands, grasslands, chaparral shrublands, and woodlands, each with 14, 7, 5, and 4 viromes, respectively) across five field sites

(see Table S1 for sample locations and soil properties). Following quality filtering, the 30 viromes generated an average of 72,950,833 reads and 416 contigs ≥ 10 Kbp each (Table S2). Wetland viromes yielded more contigs ≥ 10 Kbp than viromes from other habitats (Table S2). We used VIBRANT to identify 3,490 viral contigs in our assemblies, which were clustered into 3,432 viral operational taxonomic units (vOTUs), defined as ≥ 10 Kbp viral contigs sharing ≥ 95% average nucleotide identity.

Multiple lines of evidence suggest that wetter soils harbored greater viral diversity than drier soils. We recovered the most vOTUs from wetlands, both in total (56% of all vOTUs were from wetlands) and per sample (on average, 307 vOTUs were recovered per wetland sample, compared to 116 from other habitats) (Fig. 1A). Unsurprisingly, wetlands had significantly greater moisture content than other habitats (Fig. 1B; ANOVA followed by Tukey multiple comparisons of means, $p < 0.001$), especially considering that samples were collected towards the end of the Mediterranean dry season. Although viral richness was qualitatively highest in wetlands, this was not statistically significant (ANOVA model richness ~ habitat, $p = 0.095$). Nonparametric tests, which account for nonlinear correlations, revealed a significantly positive correlation between viral richness and soil moisture content (Spearman rho = 0.65, $p < 0.001$; Kendall tau = 0.43, $p = 0.002$). Viral community beta-diversity was also related to soil chemical properties overall (Mantel test, R^2=0.43, p=0.009; Table S1), while distance between sites only accounted for 5% of the variation (Partial Mantel test, R^2=0.38, p=0.009). Taken together, viral diversity was generally highest in wet soils.

We next wondered whether differences in sampling effort or bacterial content in the viromes could have produced the observed diversity patterns. For example, if viral richness had been captured more comprehensively in wetlands, and/or if bacterial content had been higher in

viromes from other habitats, viral diversity could have appeared artificially higher in wetlands. A comparison of accumulation curves across habitats revealed the opposite: wetlands were likely to be the most under-sampled, in terms of true viral diversity (Fig. 1C). Given that relic DNA and small bacteria can pass through 0.22 µm filters, bacterial sequences are known to be present in viromes [14,17]. To compare bacterial content in viromes across habitats, 16S rRNA gene fragments were recovered from raw reads (Fig. 1D). The percentage of 16S rRNA gene sequences in each virome ranged from 0.01-0.044% (consistent with prior reports of 0.028% bacterial 16S rRNA gene content in similarly prepared viromes from agricultural soils [18]) and did not differ significantly by habitat (ANOVA, P = 0.595). Thus, viral diversity estimates did not seem to be disproportionately skewed by sampling effort or the presence of non-viral sequences in viromes.

Perhaps the most striking result from this study was the relative uniqueness of each soil viral community. The majority of vOTUs (81%) in this regional study were only detected in a single sample (Fig. 2A). Of the 666 vOTUs detected in multiple samples, 93% were detected in replicate samples from the same habitat and site (Fig. 2B). Samples the shortest distance apart (less than 1 Km) and from the same habitat and site (*i.e.*, replicates) harbored the most similar viral communities (Fig. 2C). Within the same field site, viral communities were less similar between habitats, and paired samples from the same habitat at different, nearby sites (within 6 Km) did not share any vOTUs, suggesting substantial differences in viral communities over local distances both within and between habitats. At greater distances, community similarity generally decreased, even between samples from the same habitat (Fig. 2C). Still, 21 vOTUs were detected in multiple habitats across multiple sites (Fig. 2B), and some vOTUs were shared between the two farthest sites (109 Km apart, Fig. 2C), suggesting some degree of regional conservation of

viral populations. Overall, results suggest substantial differences in soil viral community composition in the same habitat on the scale of meters, greater similarity of viral communities in close proximity and under similar environmental conditions, and a small number of vOTUs shared over regional distances.

Although soil viral communities are woefully undersampled, making sweeping generalizations premature, results from this study and others converge on a picture of high local viral diversity, with communities differing substantially over space and by habitat, yet some regional and global conservation of viral 'species' (vOTUs). Viral community distance-decay relationships have been observed across an 18-m long agricultural field [14] and in a 200 m^2 grassland [20], consistent with the meters-scale differences in viral community composition between replicates from the same habitat observed in this study. However, the proportion of vOTUs shared over meters varied across these studies, with many vOTUs shared across the agricultural field but most vOTUs not shared between samples ~10 m apart in this study and in the grassland field, suggesting the potential for different spatial patterns in natural and managed soils, perhaps due to different amounts of mixing. Similarly substantial differences among viral communities on a regional scale were also identified in a study of grassland and peatland RNA viromes, which shared few viruses between samples [16]. However, ter Horst et al. showed that 4% of the vOTUs from a Minnesota, USA peatland were shared in other peatlands, often on different continents [8], consistent with the recovery of a small number of vOTUs shared over >100 Km distances here. Taken together, we propose that soil viral communities often display high heterogeneity within and among habitats, presumably due to a combination of host adaptations and microdiversity, dispersal limitation, and fluctuating environmental conditions over space and time.

References

1. Suttle CA. Viruses in the sea. *Nature* 2005; **437**: 356–361.
2. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, et al. Host-linked soil viral ecology along a permafrost thaw gradient. *Nat Microbiol* 2018; **3**: 870–880.
3. Williamson KE, Fuhrmann JJ, Wommack KE, Radosevich M. Viruses in Soil Ecosystems: An Unknown Quantity Within an Unexplored Territory. *Annu Rev Virol* 2017; **4**: 201–219.
4. Suttle CA. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* 2007; **5**: 801–812.
5. Trubl G, Kimbrel JA, Liquet-Gonzalez J, Nuccio EE, Weber PK, Pett-Ridge J, et al. Active virus-host interactions at sub-freezing temperatures in Arctic peat soil. *Microbiome* 2021; **9**: 208.
6. Trubl G, Jang HB, Roux S, Emerson JB, Solonenko N, Vik DR, et al. Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing. *mSystems* 2018; **3**.
7. Roy K, Ghosh D, DeBruyn JM, Dasgupta T, Wommack KE, Liang X, et al. Temporal Dynamics of Soil Virus and Bacterial Populations in Agricultural and Early Plant Successional Soils. *Front Microbiol* 2020; **11**.
8. Daly RA, Roux S, Borton MA, Morgan DM, Johnston MD, Booker AE, et al. Viruses control dominant bacteria colonizing the terrestrial deep biosphere after hydraulic fracturing. *Nat Microbiol* 2019; **4**: 352–361.

9. Srinivasiah S, Lovett J, Ghosh D, Roy K, Fuhrmann JJ, Radosevich M, et al. Dynamics of autochthonous soil viral communities parallels dynamics of host communities under nutrient stimulation. *FEMS Microbiol Ecol* 2015; **91**.

10. Zablocki O, Adriaenssens EM, Cowan D. Diversity and Ecology of Viruses in Hyperarid Desert Soils. *Appl Environ Microbiol* 2016; **82**: 770–777.

11. Kuzyakov Y, Mason-Jones K. Viruses in soil: Nano-scale undead drivers of microbial life, biogeochemical turnover and ecosystem functions. *Soil Biol Biochem* 2018; **127**: 305–317.

12. Wu R, Davison MR, Gao Y, Nicora CD, Mcdermott JE, Burnum-Johnson KE, et al. Moisture modulates soil reservoirs of active DNA and RNA viruses. *Communications biology* 2021; **4**: 1–11.

13. Lee S, Sieradzki ET, Nicolas AM. Methane-derived carbon flows into host–virus networks at different trophic levels in soil. *Proceedings of the* 2021.

14. Santos-Medellin C, Zinke LA, Ter Horst AM, Gelardi DL, Parikh SJ, Emerson JB. Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. *ISME J* 2021; **15**: 1956–1970.

15. Ter Horst AM, Santos-Medellín C, Sorensen JW, Zinke LA, Wilson RM, Johnston ER, et al. Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global viral populations. *Microbiome* 2021; **9**: 233.

16. Hillary LS, Adriaenssens EM, Jones DL, McDonald JE. RNA-viromics reveals diverse communities of soil RNA viruses with the potential to affect grassland ecosystems across multiple trophic levels. *ISME Communications* 2022; **2**: 1–10.

17. Nicolas AM, Jaffe AL, Nuccio EE, Taga ME, Firestone MK, Banfield JF. Soil Candidate Phyla Radiation Bacteria Encode Components of Aerobic Metabolism and Co-occur with

Nanoarchaea in the Rare Biosphere of Rhizosphere Grassland Communities. *mSystems* 2021; **6**: e0120520.

18. Sorensen JW, Zinke LA, Ter Horst AM, Santos-Medellín C, Schroeder A, Emerson JB. DNase Treatment Improves Viral Enrichment in Agricultural Soil Viromes. *mSystems* 2021; **6**: e0061421.

19. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 2020; **8**: 90.

20. Santos-Medellín C, Estera-Molina K, Yuan M, Pett-Ridge J, Firestone MK, Emerson JB. Spatial turnover of soil viral populations and genotypes overlain by cohesive responses to moisture in grasslands. *bioRxiv*. 2022. 2022.03.24.485562

Acknowledgements

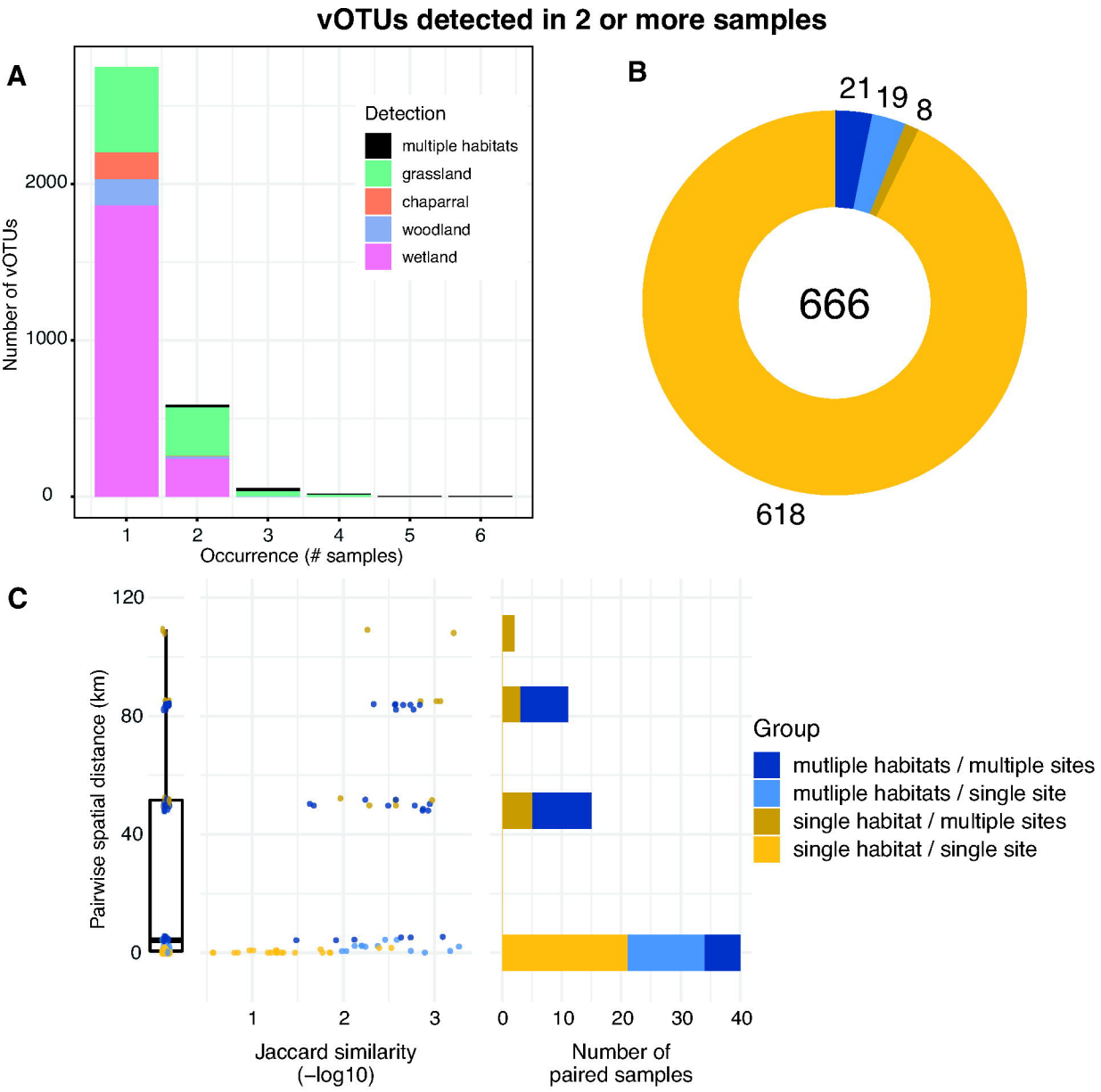
We thank the University of California, Davis Natural Reserves site directors and staff, particularly Shane Waddell, Catherine Koehler, Jeffrey Clary, Sarah Oktay, Suzanne Olyarnik, and Jacqueline Sones, for facilitating site access, showing us around the field sites, and providing logistical support for this work. Funding for this work was provided by new lab start-up to JBE from the University of California, Davis College of Agricultural and Environmental Sciences and Department of Plant Pathology, as well as by the U.S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research (BER), Genomic Science Program, award number DE-SC0021198 (grant to JBE).

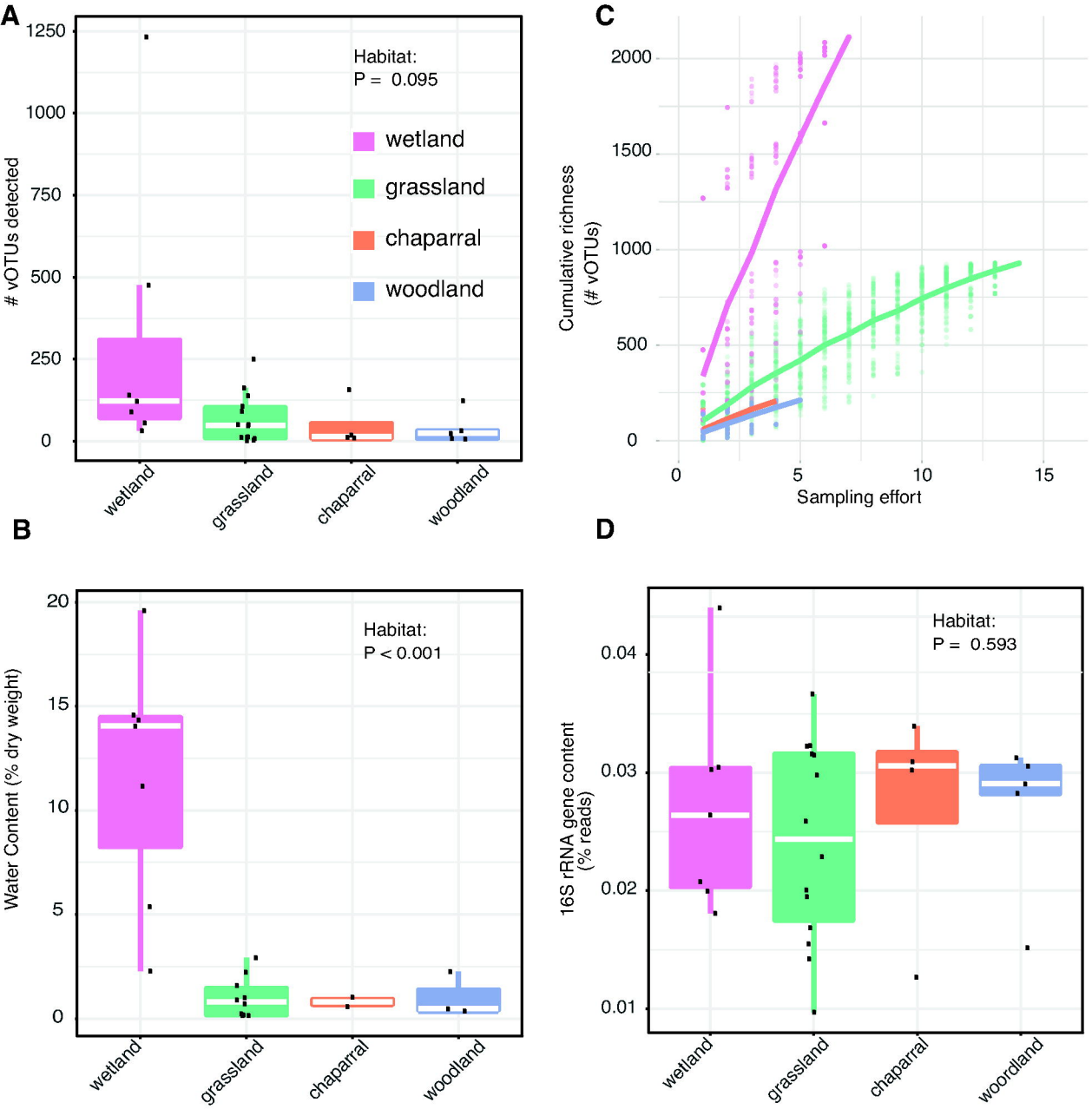
Figure titles

Figure 1. Evidence for higher viral richness in wetter soils. Comparisons between habitats of **(A)** Viral richness (number of identified vOTUs with coverage along at least 75% of the contig in a given virome), dots represent richness in individual samples, **(B)** Accumulation curves of cumulative vOTU richness as sampling effort increased, dots represent cumulative richness at each sampling effort across 100 permutations of sample order; the overlaid lines display the mean cumulative richness per habitat, **(C)** Water content, calculated as (wet weight – dry weight) divided by dry weight, and **(D)** Bacterial 16S rRNA gene content in the viromes, based on percent of viromic reads mapping to 16S rRNA reference genes. VIBRANT [19] was used to identify 3,490 viral contigs in our assemblies, and these viral contigs were clustered at 95% average nucleotide identity (ANI) into 3,432 viral operational taxonomic units (vOTUs). For A, B, D, raw data are plotted on top of the box plots, with white lines showing the median, boxes indicating 75% of the data, whiskers extending to 90%, and points beyond the whiskers indicating outliers.

Figure 2. Substantial differences in vOTU detection and viral community composition between soil samples, both within and across habitats. **(A)** vOTU detection patterns, in terms of the number of samples in which each vOTU was detected (x-axis) and the number of vOTUs exhibiting a given detection pattern (y-axis), with stacked bar colors indicating the habitat(s) in which each vOTU was detected, **(B)** vOTUs detected in more than one sample, colored by their detection patterns across sites and habitats (legend to the right of panel C), numbers indicate total vOTUs (center) and vOTUs per detection pattern, and **(C)** Pairwise community compositional similarity (x-axis) by geographic distance (y-axis) between samples, colored according to detection patterns across habitats and sites (legend on the far right). Left graph: box plot of

pairwise distances for all samples (condensed x-axis). Middle graph: Each point represents the -log10 Jaccard similarity between two samples along the x-axis, with lower values (left side of the x-axis) indicating greater viral community similarity. Right graph: Total number of sample pairs included at each geographic distance. Paired samples were only included if they shared at least one vOTU.


vOTUs detected in 2 or more samples
A
Detection
multiple habitats
grassland
chaparral
woodland
wetland
Number of vOTUs
0
1000
2000
1
2
3
4
5
6
Occurrence (# samples)
B
21
19
8
666
618
C
Pairwise spatial distance (km)
0
40
80
120
1
2
3
Jaccard similarity
(−log10)
0
10
20
30
40
Number of
paired samples
Group
mutliple habitats / multiple sites
mutliple habitats / single site
single habitat / multiple sites
single habitat / single site


A
Habitat:
P = 0.095
wetland
grassland
chaparral
woodland
vOTUs detected
1250
1000
750
500
250
0
wetland
grassland
chaparral
woodland
C
Cumulative richness
(# vOTUs)
2000
1500
1000
500
0
0
5
10
15
Sampling effort
B
Habitat:
P < 0.001
Water Content (% dry weight)
20
15
10
5
0
wetland
grassland
chaparral
woodland
D
Habitat:
P = 0.593
16S rRNA gene content
(% reads)
0.04
0.03
0.02
0.01
wetland
grassland
chaparral
woordland