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].

35 Soil viral abundance measurements vary substantially, ranging from nearly zero in dry deserts to
36 over 10^9 virus-like particles per gram in wetlands [3]. In the better studied oceans, viruses kill
37 approximately 20% of microbial biomass daily, impacting nutrient and energy cycles [4], and
38 recent work suggests that viruses may be similarly important in terrestrial ecosystems [3, 5–12].
39 For example, viruses have been suggested to affect carbon cycling in thawing permafrost
40 peatlands by preying on methanogens and methanotrophs and by encoding glycoside hydrolases
41 to break down complex carbon into simple sugars [2]. Soil viral activity has been demonstrated
42 [2, 5, 13], and soil viral communities can be spatially structured [14–16]. Despite these emerging
43 ecological patterns, comparisons of soil viral diversity within and across habitats are limited.
44 Here, we compared dsDNA viromes (viral-size fraction metagenomes) [14] from four
45 distinct habitats (wetlands, grasslands, chaparral shrublands and oak woodlands) across five UC
46 Davis Natural Reserves field sites within Northern California. We compared viral species
47 (vOTU) richness, vOTU detection patterns, and viral community beta-diversity, according to
48 habitat type, soil properties, and spatial distribution to better understand the fundamental
49 relationships between soil viruses and the ecosystems that they inhabit.
50
51 **Results and Discussion**
52
53 To compare soil viral community composition within and across terrestrial habitats on a
54 regional scale, viromes were generated from 34 near-surface (top 15 cm) soil samples, with a
55 total of 30 viromes included in downstream ecological analyses (see Supplementary Methods).
56 The analyzed samples were collected from four distinct habitats (wetlands, grasslands, chaparral
57 shrublands, and woodlands, each with 14, 7, 5, and 4 viromes, respectively) across five field sites

58 (see Table S1 for sample locations and soil properties). Following quality filtering, the 30
59 viromes generated an average of 72,950,833 reads and 416 contigs ≥ 10 Kbp each (Table S2).
60 Wetland viromes yielded more contigs ≥ 10 Kbp than viromes from other habitats (Table S2).
61 We used VIBRANT to identify 3,490 viral contigs in our assemblies, which were clustered into
62 3,432 viral operational taxonomic units (vOTUs), defined as ≥ 10 Kbp viral contigs sharing ≥
63 95% average nucleotide identity.

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

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

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

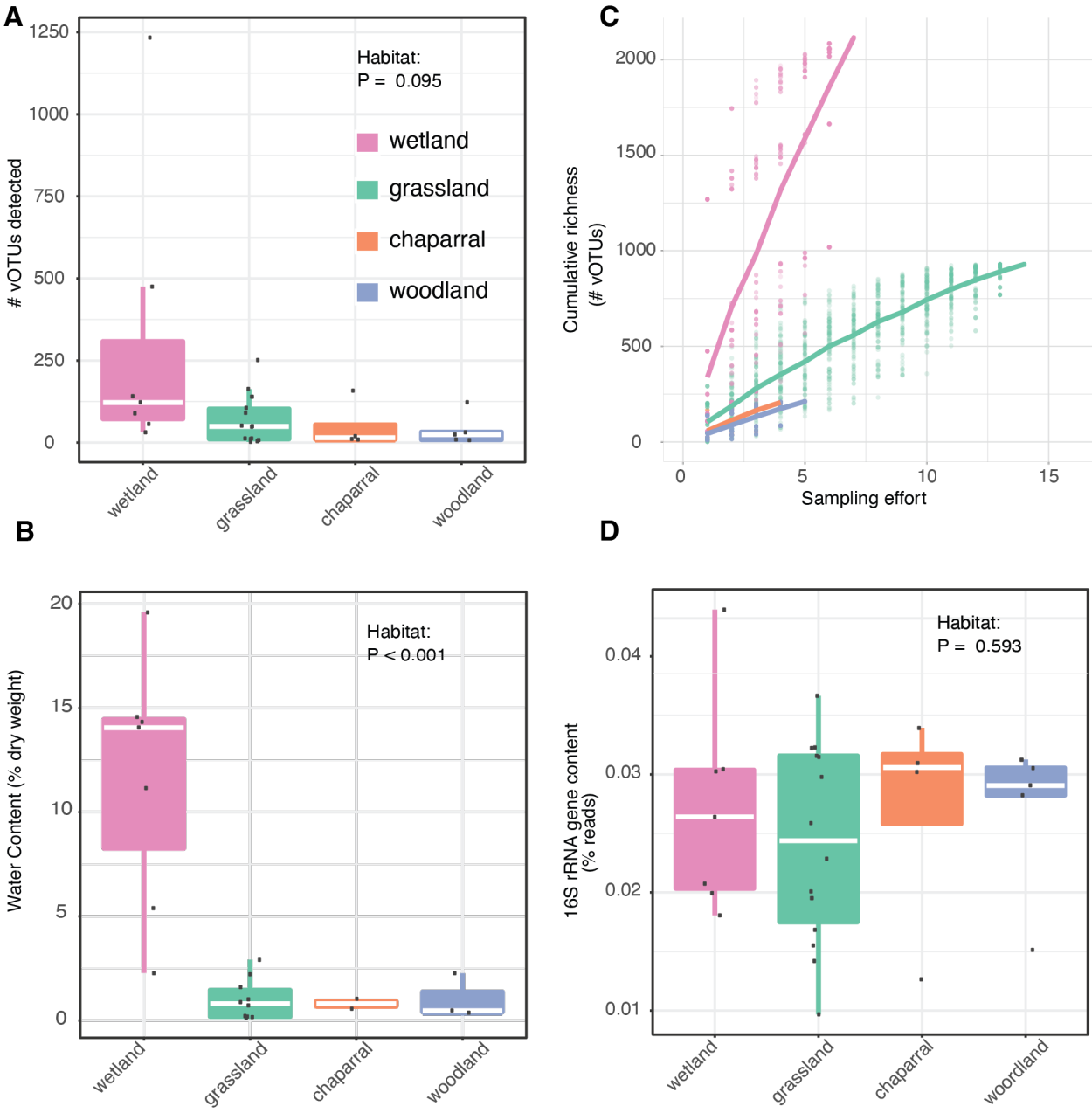
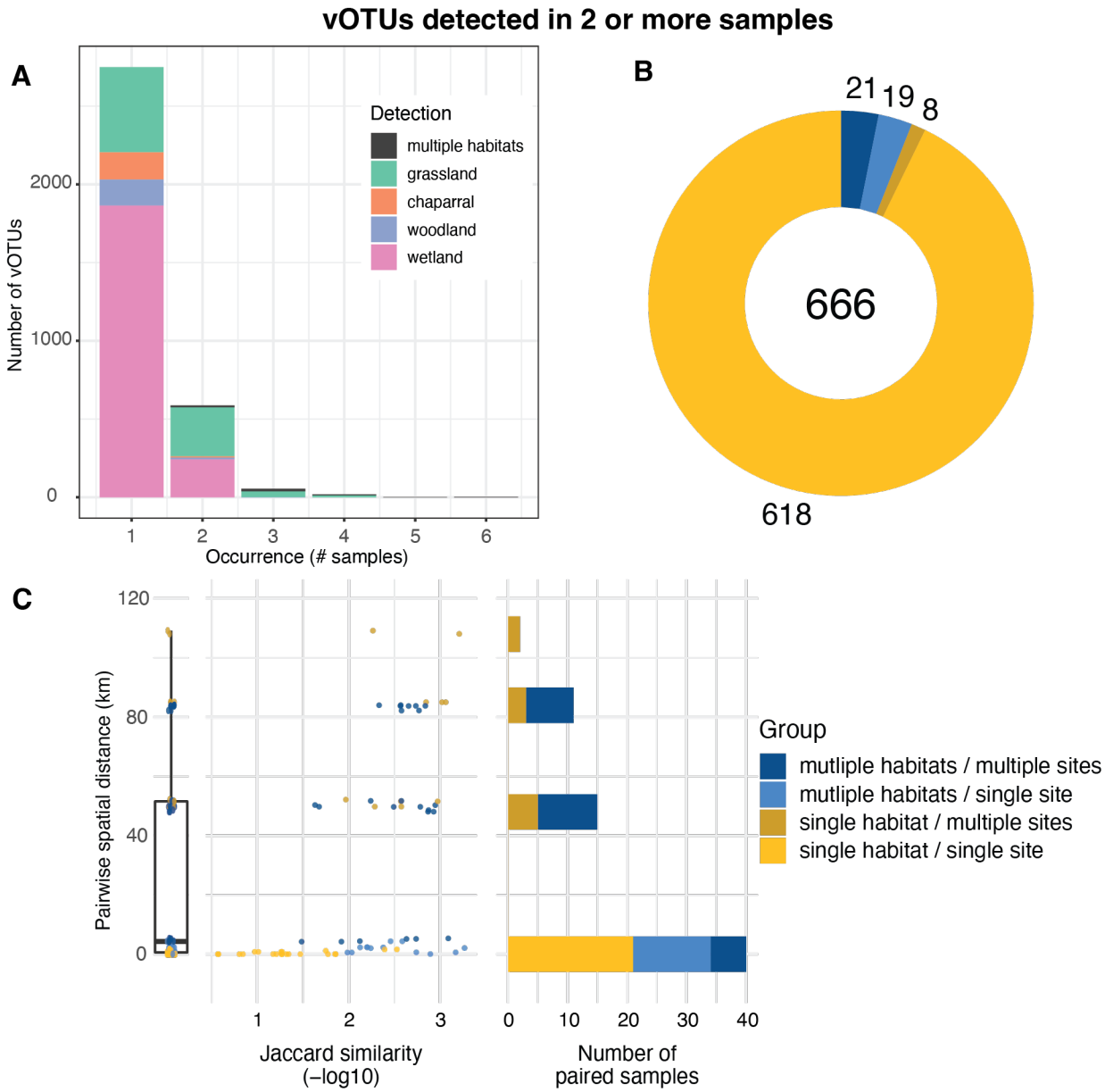


Figure 1. Evidence for higher viral richness in wetter soils. Comparisons between habitats of **(A)** Viral richness (number of identified vOTUs with coverage along at least 75% of the contig in a given virome), dots represent richness in individual samples, **(B)** Accumulation curves of cumulative 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)

99 divided by dry weight, and **(D)** Bacterial 16S rRNA gene content in the viromes, based on
100 percent of viromic reads mapping to 16S rRNA reference genes. VIBRANT [19] was used to
101 identify 3,490 viral contigs in our assemblies, and these viral contigs were clustered at 95%
102 average nucleotide identity (ANI) into 3,432 viral operational taxonomic units (vOTUs). For A,
103 B, D, raw data are plotted on top of the box plots, with white lines showing the median, boxes
104 indicating 75% of the data, whiskers extending to 90%, and points beyond the whiskers
105 indicating outliers.
106
107 Perhaps the most striking result from this study was the relative uniqueness of each soil
108 viral community. The majority of vOTUs (81%) in this regional study were only detected in a
109 single sample (Fig. 2A). Of the 666 vOTUs detected in multiple samples, 93% were detected in
110 replicate samples from the same habitat and site (Fig. 2B). Samples the shortest distance apart
111 (less than 1 Km) and from the same habitat and site (*i.e.*, replicates) harbored the most similar
112 viral communities (Fig. 2C). Within the same field site, viral communities were less similar
113 between habitats, and paired samples from the same habitat at different, nearby sites (within 6
114 Km) did not share any vOTUs, suggesting substantial differences in viral communities over local
115 distances both within and between habitats. At greater distances, community similarity generally
116 decreased, even between samples from the same habitat (Fig. 2C). Still, 21 vOTUs were detected
117 in multiple habitats across multiple sites (Fig. 2B), and some vOTUs were shared between the
118 two farthest sites (109 Km apart, Fig. 2C), suggesting some degree of regional conservation of
119 viral populations. Overall, results suggest substantial differences in soil viral community
120 composition in the same habitat on the scale of meters, greater similarity of viral communities in

121 close proximity and under similar environmental conditions, and a small number of vOTUs

122 shared over regional distances.



123

124 **Figure 2.** Substantial differences in vOTU detection and viral community composition between

125 soil samples, both within and across habitats. **(A)** vOTU detection patterns, in terms of the

126 number of samples in which each vOTU was detected (x-axis) and the number of vOTUs

127 exhibiting a given detection pattern (y-axis), with stacked bar colors indicating the habitat(s) in

128 which each vOTU was detected, **(B)** vOTUs detected in more than one sample, colored by their
129 detection patterns across sites and habitats (legend to the right of panel C), numbers indicate total
130 vOTUs (center) and vOTUs per detection pattern, and **(C)** Pairwise community compositional
131 similarity (x-axis) by geographic distance (y-axis) between samples, colored according to
132 detection patterns across habitats and sites (legend on the far right). Left graph: box plot of
133 pairwise distances for all samples (condensed x-axis). Middle graph: Each point represents the -
134 log10 Jaccard similarity between two samples along the x-axis, with lower values (left side of the
135 x-axis) indicating greater viral community similarity. Right graph: Total number of sample pairs
136 included at each geographic distance. Paired samples were only included if they shared at least
137 one vOTU.
138
139 Although soil viral communities are woefully undersampled, making sweeping
140 generalizations premature, results from this study and others converge on a picture of high local
141 viral diversity, with communities differing substantially over space and by habitat, yet some
142 regional and global conservation of viral 'species' (vOTUs). Viral community distance-decay
143 relationships have been observed across an 18-m long agricultural field [14] and in a 200 m^2
144 grassland [20], consistent with the meters-scale differences in viral community composition
145 between replicates from the same habitat observed in this study. However, the proportion of
146 vOTUs shared over meters varied across these studies, with many vOTUs shared across the
147 agricultural field but most vOTUs not shared between samples ~10 m apart in this study and in
148 the grassland field, suggesting the potential for different spatial patterns in natural and managed
149 soils, perhaps due to different amounts of mixing. Similarly substantial differences among viral
150 communities on a regional scale were also identified in a study of grassland and peatland RNA

151 viromes, which shared few viruses between samples [16]. However, ter Horst et al. showed that
152 4% of the vOTUs from a Minnesota, USA peatland were shared in other peatlands, often on
153 different continents [8], consistent with the recovery of a small number of vOTUs shared over
154 >100 Km distances here. Taken together, we propose that soil viral communities often display
155 high heterogeneity within and among habitats, presumably due to a combination of host
156 adaptations and microdiversity, dispersal limitation, and fluctuating environmental conditions
157 over space and time.
158

159 **References**

160

161 1. Suttle CA. Viruses in the sea. *Nature* 2005; **437**: 356–361.
162 2. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, et al. Host-linked soil
163 viral ecology along a permafrost thaw gradient. *Nat Microbiol* 2018; **3**: 870–880.
164 3. Williamson KE, Fuhrmann JJ, Wommack KE, Radosevich M. Viruses in Soil
165 Ecosystems: An Unknown Quantity Within an Unexplored Territory. *Annu Rev Virol* 2017; **4**:
166 201–219.
167 4. Suttle CA. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol*
168 2007; **5**: 801–812.
169 5. Trubl G, Kimbrel JA, Liquet-Gonzalez J, Nuccio EE, Weber PK, Pett-Ridge J, et al.
170 Active virus-host interactions at sub-freezing temperatures in Arctic peat soil. *Microbiome* 2021;
171 **9**: 208.
172 6. Trubl G, Jang HB, Roux S, Emerson JB, Solonenko N, Vik DR, et al. Soil Viruses Are
173 Underexplored Players in Ecosystem Carbon Processing. *mSystems* 2018; **3**.

174 7. Roy K, Ghosh D, DeBruyn JM, Dasgupta T, Wommack KE, Liang X, et al. Temporal
175 Dynamics of Soil Virus and Bacterial Populations in Agricultural and Early Plant Successional
176 Soils. *Front Microbiol* 2020; **11**.
177 8. Daly RA, Roux S, Borton MA, Morgan DM, Johnston MD, Booker AE, et al. Viruses
178 control dominant bacteria colonizing the terrestrial deep biosphere after hydraulic fracturing. *Nat*
179 *Microbiol* 2019; **4**: 352–361.
180 9. Srinivasiah S, Lovett J, Ghosh D, Roy K, Fuhrmann JJ, Radosevich M, et al. Dynamics
181 of autochthonous soil viral communities parallels dynamics of host communities under nutrient
182 stimulation. *FEMS Microbiol Ecol* 2015; **91**.
183 10. Zablocki O, Adriaenssens EM, Cowan D. Diversity and Ecology of Viruses in Hyperarid
184 Desert Soils. *Appl Environ Microbiol* 2016; **82**: 770–777.
185 11. Kuzyakov Y, Mason-Jones K. Viruses in soil: Nano-scale undead drivers of microbial
186 life, biogeochemical turnover and ecosystem functions. *Soil Biol Biochem* 2018; **127**: 305–317.
187 12. Wu R, Davison MR, Gao Y, Nicora CD, Mcdermott JE, Burnum-Johnson KE, et al.
188 Moisture modulates soil reservoirs of active DNA and RNA viruses. *Communications biology*
189 2021; **4**: 1–11.
190 13. Lee S, Sieradzki ET, Nicolas AM. Methane-derived carbon flows into host–virus
191 networks at different trophic levels in soil. *Proceedings of the* 2021.
192 14. Santos-Medellin C, Zinke LA, Ter Horst AM, Gelardi DL, Parikh SJ, Emerson JB.
193 Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural
194 soil viral communities. *ISME J* 2021; **15**: 1956–1970.

195 15. Ter Horst AM, Santos-Medellín C, Sorensen JW, Zinke LA, Wilson RM, Johnston ER, et
196 al. Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global
197 viral populations. *Microbiome* 2021; **9**: 233.
198 16. Hillary LS, Adriaenssens EM, Jones DL, McDonald JE. RNA-viromics reveals diverse
199 communities of soil RNA viruses with the potential to affect grassland ecosystems across
200 multiple trophic levels. *ISME Communications* 2022; **2**: 1–10.
201 17. Nicolas AM, Jaffe AL, Nuccio EE, Taga ME, Firestone MK, Banfield JF. Soil Candidate
202 Phyla Radiation Bacteria Encode Components of Aerobic Metabolism and Co-occur with
203 Nanoarchaea in the Rare Biosphere of Rhizosphere Grassland Communities. *mSystems* 2021; **6**:
204 e0120520.
205 18. Sorensen JW, Zinke LA, Ter Horst AM, Santos-Medellín C, Schroeder A, Emerson JB.
206 DNase Treatment Improves Viral Enrichment in Agricultural Soil Viromes. *mSystems* 2021; **6**:
207 e0061421.
208 19. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and
209 curation of microbial viruses, and evaluation of viral community function from genomic
210 sequences. *Microbiome* 2020; **8**: 90.
211 20. Santos-Medellín C, Estera-Molina K, Yuan M, Pett-Ridge J, Firestone MK, Emerson JB.
212 Spatial turnover of soil viral populations and genotypes overlain by cohesive responses to
213 moisture in grasslands. *bioRxiv*. 2022. 2022.03.24.485562
214

215 **Acknowledgements**

216 We thank the University of California, Davis Natural Reserves site directors and staff,
217 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).

Supplemental Methods

Experimental design adjustments on account of COVID-19 and wildfire

The sample set analyzed in this study, collected in November 2019, was meant to be the first of several time points, with plans for periodic sampling of the same sites in 2020-2021. However, COVID-19 prevented sample collection throughout much of 2020, and then the LNU Complex Fires completely burned 2 and partially burned 1 of our 5 field sites in August 2020. Thus, the decision was made to analyze this November 2019 dataset on its own.

Site description and field sampling

Soil samples were collected from woodland, chaparral, grassland, coastal bluff, and wetland habits at five UC Davis Natural Reserves sites in Northern California, USA: Bodega Bay (BB), Jepson Prairie (JP), McLaughlin (ML), Quail Ridge (QR), and Stebbins Cold Canyon (SCC). From November 14 through November 16, 2019, 34 samples were collected from 17 plots within the five sites. We sampled at least two distinct habitats from each site, and two samples (replicates) were retrieved from each plot, with each plot consisting of one contiguous

241 habitat. Inter-sample distance between replicates in each plot ranged from 6.5-74.2 m and was
242 meant to capture a large area within one instance of the habitat (*i.e.*, no distinct habitat types
243 were permitted between replicate samples), considering available habitat space and accessibility.
244 The number of samples per site and habitat, along with GPS coordinates (Table S1) and distance
245 between sites, appear in Tables S3 and S4, and soil chemistry appears in Table S1.
246 For each sample, three 7.6 cm diameter soil cores were collected from the 0-15 cm depth
247 range (excluding aboveground biomass) with a slide-hammer corer, approximately 15 cm apart
248 from each other. The three fresh cores for each sample were homogenized and transferred into
249 sterile bags on site, kept on ice for 1-2 days in the field, and then passed through an 8 mm mesh
250 sieve within 3 days of sample collection and then stored at -80 °C prior to processing for DNA
251 extraction and soil properties.
252
253 *Soil viral purification and DNA extraction*
254 Viral particle purification and DNA extraction proceeded similarly to the Santos-
255 Medellin et al. [1] protocol (which was largely derived from Goller et al. [2]), with slight
256 modifications. Briefly, for each sample, a suspended soil solution was prepared by adding 9 mL
257 of PPBS Buffer (2% bovine serum albumin, 10% phosphate-buffered saline, 1% potassium
258 citrate, and 150 mM MgSO4) to 10 g of soil [2]. To elute virions, soil suspensions were shaken
259 by inversion and placed in an orbital shaker for 20 min at 300 RPM and 4 °C, then centrifuged
260 (Sorvall RC5C, Thermo-Fisher Scientific, Waltham, MA, USA) for 10 min at 10,000 x g and 4
261 °C. Supernatant was briefly stored at 4 °C while the soil pellet was resuspended with 9 mL of
262 PPBS Buffer, and the entire process of mixing, shaking, and centrifugation was repeated two
263 more times, resulting in a total of ~27 mL of supernatant per soil sample. To pellet and remove

264 remaining soil, supernatant was centrifuged for 8 min at 5,000 x g and 4 °C, with supernatant
265 transferred, excluding the pellet. This step was repeated twice. The pooled supernatants were
266 filtered through a 0.22 µm polyether sulfone filter and centrifuged (Optima LE-80K
267 ultracentrifuge, Beckman Coulter, Brea, CA, USA; 50.2 Ti rotor) for 2 hours 25 minutes at
268 32,000 g and 4 °C. The supernatant was removed and the pellet containing viral particles was
269 resuspended in 100 µL of ultrapure water. DNase was not used here, as samples were frozen,
270 which may compromise the integrity of viral capsids [3, 4]. However, ecological trends from
271 viromes can still be detected without DNase treatment [3, 4].
272 For DNA extraction, 250 µL of phenol:chloroform:isoamyl alcohol (pH 8) was added to
273 each 100 µL sample. The mixture was vortexed and then incubated on ice for 3 cycles of 1 min
274 on and 5 min off intervals. Samples were then centrifuged for 5 min at 14 000 x g and the
275 supernatant was transferred to a Phase Lock Gel™ tube, and then centrifuged for 5 min at 14,000
276 x g. The viral DNA supernatant (100 µL) was mixed with 25 µL of 3 M sodium acetate, 1.5 µL
277 of Glycoblue™ Coprecipitant, and 250 µL isopropanol. The mixture was vortexed and then
278 incubated at -80 °C for 20 min. After incubation, samples were centrifuged for 20 min at 14,000
279 x g. Supernatant was removed and the viral pellet was washed with 500 µL of 70 % ethanol and
280 centrifuged for 5 min at 14,000 x g. The ethanol was removed and the samples were centrifuged
281 for another minute, then left to dry for 5 min. The viral DNA pellet was resuspended with 100
282 µL of non-DEPC-treated ultrapure water. DNA yields were quantified using an Invitrogen Qubit
283 4 Fluorometer with 1X High Sensitivity DNA assay (Thermo Fisher Scientific, Inc., Waltham,
284 MA, USA). Of the 34 samples from which viral DNA was extracted, one replicate from a BB
285 wetland site (sample ID: BB_3_2) and one replicate from a QR woodland site (sample ID:

286 QR_2_1) yielded insufficient DNA and were thus excluded, resulting in 32 samples remaining
287 for library construction and sequencing.
288
289 *Library construction and sequencing*
290 Libraries from the 32 viromes were constructed with the KAPA Roche kit, as described
291 in Santos-Medellin et al. [1], pooled in equimolar concentrations, and sequenced to a target depth
292 of 10 Gbp per virome on the Illumina NovaSeq platform at the UC Davis DNA Technologies
293 Core.
294
295 *Bioinformatics*
296 Raw sequencing reads were quality-filtered and trimmed using Trimmomatic-0.39, as in
297 Roux et al. [5, 6]. Removal of PhiX147 sequences was performed with BBDUK from the
298 BBMap package (version: BBMap-38.87) [7], followed by assembly with MEGAHIT-1.2.9 in
299 meta-large mode [8, 9]. Predicted viral contigs were identified using VIBRANT-1.2.1 with
300 default settings [10]. dRep version 2.0.0 [11] was used for clustering sequences at 95% average
301 nucleotide identity across 85% of the length of the shorter contig, resulting in a reference set of
302 4,712 viral population sequences (vOTUs). Reads from the 32 viromes were mapped to these
303 reference vOTU sequences with Bowtie 2-2.4.2 [12], using sensitive mode, and the vOTU read-
304 mapped coverage table was created with BamM-1.7.3 [13]. Where vOTU breadth (contig length
305 fraction with at least 1x coverage depth) was below 75%, coverage was set to zero [8]. Two
306 viromes (sample IDs: BB_6_1 and BB_6_2) were excluded from downstream comparative
307 analyses at this point because they were the only samples from a coastal bluff habitat across the

308 dataset, thus this habitat was deemed insufficiently sampled for cross-habitat comparisons. The
309 final dataset for comparative analyses consisted of 30 viromes.
310 16S rRNA gene abundance recovery from viromes was performed by identifying reads
311 that map to this gene with SortMeRNA v4.2.0 [15] against the SILVA database [16], as in [4,5].
312
313 *Soil properties*
314 Soil was weighed upon arrival to the lab (Wet Weight, WW), allowed to dry completely
315 in a biological hood, and weighed again (Dry Weight, DW). Soil water content was calculated as
316 (WW-DW)/DW.
317 All other soil properties appearing in Table S1 were analyzed by Ward Laboratories
318 (Kearney, NE, USA). Phosphorus and sulfate were extracted with Mehlich-3 buffer. Iron,
319 manganese, copper, and zinc were extracted with DTPA buffer. Sodium, potassium, calcium, and
320 magnesium were extracted with ammonium-acetate. Soil organic matter was calculated by
321 percentage loss on ignition (LOI). Soil pH and soluble salts were measured using a 1:1 soil:water
322 suspension. Nitrate was extracted with KCl.
323
324 *Data analysis*
325 Data analysis was performed in R on the vOTU read-mapped coverage table generated
326 previously (see *Bioinformatics*). When calculating community similarity (beta-diversity), we
327 used Jaccard similarity, as opposed to Bray-Curtis similarity, as there were so few vOTUs shared
328 between samples. Similarity matrices were calculated using vegan [17]. Richness of vOTUs
329 detected per site and per habitat was calculated using tidyverse [18] and plotted with cowplot
330 [19] and ggplot2 [20]. Comparisons of vOTU richness between sites were performed with rstatix

331 [21], using the Games-Howell test to account for uneven sample size between habitats.
332 Correlations between water content and viral richness, ANOVA comparison of water content
333 between environments, and Tukey post hoc tests were calculated using the R stats package. The
334 correlation analysis was performed under the hypothesis that higher moisture leads to higher
335 richness and was therefore one-sided (alternative=greater). A Mantel test comparing viral
336 community composition to soil chemistry was performed by creating Jaccard distance matrices
337 (vegdist) and comparing them (mantel, 100 permutations; mantel.partial, 100 permutations) with
338 functions from the Vegan R package [18].
339
340 *Data availability*
341 The datasets generated and/or analyzed in this study have been submitted to the NCBI
342 sequence read archive (SRA) under BioProject number PRJNA831438 and will become
343 available upon publication. The vOTU fasta sequences are available on
344 https://github.com/ellasiera/Nat_res_vOTUs.
345
346 **Supplementary references**
347 1. Santos-Medellín C, Estera-Molina K, Yuan M, Pett-Ridge J, Firestone MK, Emerson JB.
348 Spatial turnover of soil viral populations and genotypes overlain by cohesive responses to
349 moisture in grasslands. *bioRxiv* 2022. 2022.03.24.485562
350 2. Göller PC, Haro-Moreno JM, Rodriguez-Valera F, Loessner MJ, Gómez-Sanz E.
351 Uncovering a hidden diversity: optimized protocols for the extraction of dsDNA bacteriophages
352 from soil. *Microbiome* 2020; **8.1**: 1-16

353 3. Emerson JB, Thomas BC, Andrade K, Allen EE, Heidelberg KB, Banfield JF. Dynamic
354 viral populations in hypersaline systems as revealed by metagenomic assembly. *Appl Environ*
355 *Microbiol* 2012; **78**: 6309–6320.
356 4. Sorensen JW, Zinke LA, Ter Horst AM, Santos-Medellín C, Schroeder A, Emerson JB.
357 DNase Treatment Improves Viral Enrichment in Agricultural Soil Viromes. *mSystems* 2021; **6**:
358 e0061421.
359 5. Santos-Medellin C, Zinke LA, Ter Horst AM, Gelardi DL, Parikh SJ, Emerson JB.
360 Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural
361 soil viral communities. *ISME J* 2021; **15**: 1956–1970.
362 6. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. Benchmarking viromics: an in
363 silico evaluation of metagenome-enabled estimates of viral community composition and
364 diversity. *PeerJ* 2017; **5**: e3817.
365 7. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
366 data. *Bioinformatics* 2014; **30**: 2114–2120.
367 8. Bushnell B. BBTools software package. *URL http://sourceforge net/projects/bbmap*
368 2014.
369 9. Ter Horst AM, Santos-Medellín C, Sorensen JW, Zinke LA, Wilson RM, Johnston ER, et
370 al. Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global
371 viral populations. *Microbiome* 2021; **9**: 233.
372 10. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node
373 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
374 *Bioinformatics* 2015; **31**: 1674–1676.

375 11. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and
376 curation of microbial viruses, and evaluation of viral community function from genomic
377 sequences. *Microbiome* 2020; **8**: 90.
378 12. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic
379 comparisons that enables improved genome recovery from metagenomes through de-replication.
380 *ISME J* 2017; **11**: 2864–2868.
381 13. Langmead B, Salzberg SL. Langmead. 2013. Bowtie2. *Nat Methods* 2013; **9**: 357–359.
382 14. BamM: Metagenomics-focused BAM file manipulation. Github.
383 15. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal
384 RNAs in metatranscriptomic data. *Bioinformatics* 2012; **28**: 3211–3217.
385 16. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA
386 ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic*
387 *Acids Res* 2013; **41**: D590–6.
388 17. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB, et al. Vegan:
389 community ecology package. R package version 2.3-0; 2015. *Sci Rep* 2020; **10**: 20354.
390 18. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to
391 the tidyverse. *J Open Source Softw* 2019; **4**: 1686.
392 19. Wilke CO. cowplot: streamlined plot theme and plot annotations for 'ggplot2'. *R package*
393 *version 0 9* 2019; **4**.
394 20. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2016. Springer.
395 21. Kassambara A. rstatix: Pipe-friendly framework for basic statistical tests. *R package*
396 *version 0 6 0* 2020.
397

398 **Supplementary Tables**

399 Table S1. Soil chemistry characteristics of all samples.

400 Table S2. Summary of assembly statistics including detection of viral genomes per habitat

	Across all habitats	**Grassland**	**Wetland**	**Woodland**	**Chaparral**
Total reads	2,160,621,059	943,863,746	577,454,144	358,356,831	280,946,338
Total contigs > 10 kbp	12,850	3,376	8,513	704	257
Average reads per virome	72,950,833	67,401,933	82,493,449	71,671,366	70,236,585
Average contigs > 10 kbp per virome	416	241	1,216	14,164	
Average viral contigs predicted per virome (VIBRANT)	116	68	307	39	50
Total vOTUs	3,490	949	2,148	195	198

401

402 Table S3. Distance (km) between UC Davis Natural Reserves sites sampled

	QR	SCC	JP	ML	BB
QR		6	36	50	85
SCC			36	50	85
JP				85	109
ML					85

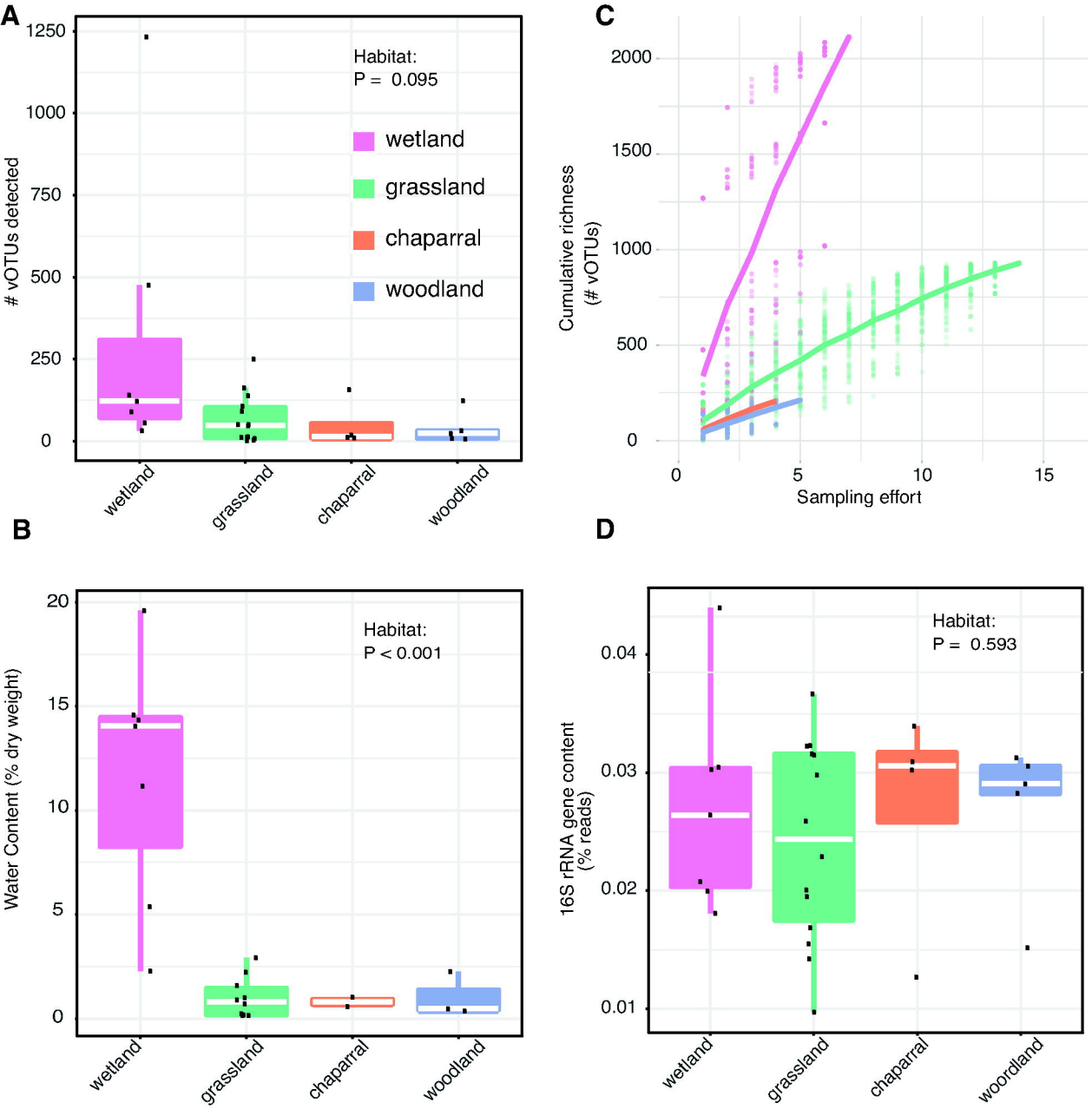
403

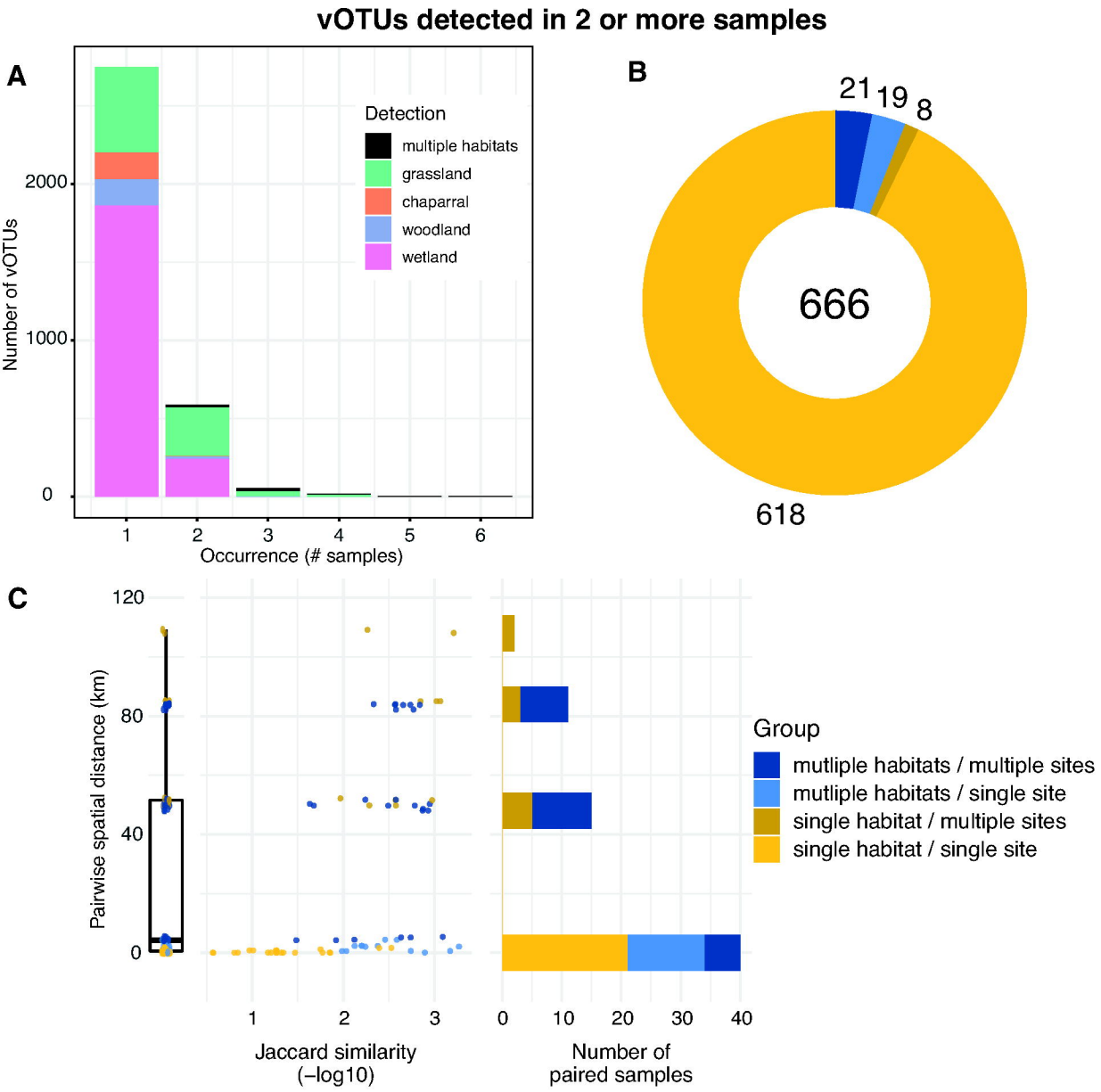
404 Table S4. Number of samples analyzed by habitat and site

# in cell represents # of samples		Habitat				
		Woodland	Chaparral	Grassland	Wetland	Total
	BB			4	5	9

Site	JP			4	2	6
	ML	2	2	4		8
	QR	1	2			3
	SCC	2		2		4
	Total	5	4	14	7	30

405


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


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