Spatiotemporal dynamics of RNA virus diversity in a phyllosphere microbial community

Lisa M. Bono^{1*}, Richard J. Orton^{2*}, Elena L. Peredo³, Hilary G. Morrison⁴, Mark Sistrom⁵, Sheri Simmons⁶ and Paul E. Turner^{7,8}

¹Department of Ecology, Evolution and Natural Resources, School of Environmental and Biological Sciences, Rutgers, the State University of New Jersey, 14 College Farm Rd, New Brunswick, NJ 08901

²MRC-University of Glasgow Centre for Virus Research, College of Medical, Veterinary, and Life Sciences, University of Glasgow, 464 Bearsden Road, Glasgow, G61 1QH, UK

³Ecosystems Center, Marine Biological Laboratory, Woods Hole, MA 02543 USA

⁴Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA, USA

⁵School of Natural Sciences, University of California Merced, Merced, CA, 95343 USA

⁶Seres Therapeutics, Inc., Cambridge, MA, 02139 USA

⁷Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect St., New Haven, CT 06511

⁸Program in Microbiology, Yale School of Medicine, New Haven, CT 06520 USA

*These individuals contributed equally to the work.

Corresponding author: Paul E. Turner: Tel: 203-432-5918: E-mail: paul.turner@yale.edu

Keywords: legume, white clover mosaic virus, virome, RNA viral community, *Trifolium repens*

Abstract

Although metagenomics reveals that natural virus communities harbor vast genetic diversity, the spatiotemporal dynamics of viral diversity in the wild are seldom tested, especially across small geographic scales. This problem is usefully examined in the above-ground phyllosphere, because terrestrial plants are frequently infected by taxonomically-diverse RNA viruses, whose elevated mutation rates generate abundant allele diversity. Here, we studied the problem by comparative analysis of RNA virus samples over time from three spatially-separated patches of a common perennial legume, white clover (*Trifolium repens* L.), growing in a grassy lawn in Woods Hole Village (Falmouth, MA, USA). We predicted that clover samples would show similarly high levels of virus species (alpha) diversity across space, but differing among-patch diversity of non-dominant virus taxa over time (4 samples spanning 6 weeks). Results showed that recognizable alpha diversity in clover patches was consistently dominated by RNA virus family *Alphaflexiviridae* across space, but that all patches showed inconsistent spatiotemporal presence of a diversity of minority virus families. Also, we observed that white clover mosaic virus (WCIMV) dominated all patches across space and time. The high coverage of WCIMV fostered an haplotype analysis, which revealed that two strains of the virus consistently infected clover plants during the 6-week period.

Introduction

Viruses numerically dominate biological entities on Earth and may represent the majority of its genetic diversity as well [1, 2]. Viruses can alter microbial community structure [3] and colonization [4] and impact food webs. Their effects also scale up, driving global geochemical cycles [5, 6]. However, virus sampling across different natural environments is uneven. Marine systems are perhaps the best studied to date [7], making it crucial to examine non-marine and terrestrial communities (but see [4, 8]). Understanding the viral communities of the phyllosphere (above-ground terrestrial plant biomass) has important implications for food safety, *e.g.* outbreaks of food-borne human pathogens, such as hepatitis A virus and norovirus, have been associated with fresh produce [9]. Furthermore, the broadscale ecological and evolutionary implications of plant viruses is largely unknown, even though viral infections have been detected in a variety of wild plants [10]. Even a seemingly basic question about the nature of viral communities associated with plants remains unknown: will the same viruses be observed today as they will tomorrow, the next week, *et cetera*?

Plant viruses face unique challenges when infecting their hosts, which can potentially drive the ecological and evolutionary dynamics of their microbial communities. The plant cell wall is particularly difficult to breach; viruses can avoid this issue either through vertical transmission, or by relying on passive, mechanical damage to plant cell walls, most notably via feeding by phytophagous insects but also via animal grazing and human activities such as lawn mowing. This differs from phage infection of bacterial cells and viral infection of animal cells, where viruses typically can enter host cells without relying on external forces. Plant viruses can be transmitted vertically, from parent to seed. Plants are immobile, so their viruses rely on

transmission for dispersal, with many plant viruses closely reliant on insect vectors [11]. As a consequence, plant viruses tend to be host generalists but vector specialists [12]. Plants lack highly specific immune systems, relying instead on generalized defenses; consequently, plant viruses can be selected to resist common mechanisms of host defense [11]. Plants can be persistently and asymptotically infected by viruses, particularly those in the family *Partitiviridae* or the genus *Endornavirus* [13]. In comparison with crops that often display disease phenotypes, wild plants can be infected by viruses without such phenotypic indicators, leading to their prevalence being generally underestimated and underappreciated [10]. These types of infections can lead to complex, long-term dynamics over the course of a plant's lifespan, which may vary dramatically from the population dynamics of viruses that cause acute infections.

Nearly all plant viruses have genomes composed of RNA, with most being +ssRNA [14]. This is in contrast to other well-studied viral communities—marine viromes—which are dominated by dsDNA viruses [15–20], although some data suggest that perhaps half of marine viruses have RNA genomes [21]. Due to their typically high mutation rates, RNA virus populations are expected to harbor extensive genetic variation and to be capable of rapid evolution. These properties may foster RNA virus ability to jump (emerge) easily onto new host species and to cause pandemics, relative to DNA viruses [22]; however, [11] found no difference between host range of RNA and DNA plant viruses.

We chose to focus our sampling efforts on white clover (*Trifolium repens* L.), an herbaceous perennial plant that is geographically widespread in temperate regions, ranging from the subarctic to the subtropical [23, 24]. Originally cultivated in the 16th century in the Netherlands [25], this legume serves as a vital forage crop but also thrives as a weedy species in

disturbed, moist areas, particularly mowed lawns. Agriculturally, *T. repens* often composes only a small proportion of a grass-legume mixed sward but provides high nutritional quality as forage. As a cover crop and companion to grass species, it is often used to enrich the soil, because it is capable of symbiotically fixing atmospheric nitrogen [26]. As an obligate outbreeding allotetraploid, cultivars and natural populations of *T. repens* are extremely diverse [24, 27, 28]. Additionally, since *T. repens* has extensive natural populations and often grows—intentionally planted or not—in and around farms, it may act as a reservoir for viruses that infect agricultural plants.

Previous studies suggest that *T. repens* is most commonly infected by the aptly-named White clover mosaic virus (WCIMV). This (+)ssRNA monopartite virus (genus *Potexvirus*, family *Alphaflexiviridae*) has a 5.84 kb genome packaged in a non-enveloped, flexuous filamentous capsid. WCIMV infection causes mosaic-like and mottle-like symptoms but cannot be distinguished by these discolorations alone, as its symptoms are very similar to other clover viruses. WCIMV has a global distribution and a narrow host range [29, 30] able to infect other clover species as well as broad bean (*Vicia faba*) and sweet pea (*Lathyrus odoratus*). Plants inoculated with WCIMV produced a third less dry material [31]. In addition to WCIMV, *T. repens* is often coinfecting with Clover yellow mosaic virus (CIYMV), a sister species to WCIMV [32]. Both species are transmitted vertically via seed or mechanically via contact with material from infected plants [29] with grazing and mowing being particularly important for transmission between pastures [33].

Here, we compared the dynamics of RNA viral communities associated with a single host species, *T. repens*, over the small scale with the goal of testing how much a community of viruses varies in a single patch over a short time course. We collected samples of clover from 3

plots spaced every 5 m along a transect every 2 weeks for 6 weeks in total. Since we only collected tissue above ground, the viral communities are associated with the phyllosphere. Specifically, we test if the viral communities would show similar levels of species (alpha) diversity but differ in the diversity of non-dominant virus taxa. We also characterized the complexity of the quasispecies of the dominant viral species, including performing a haplotype reconstruction and tested for a signature of space or time among haplotypes. Finally, we examined the diversity of the rare species within the viral community.

Materials and Methods

Sampling, Extraction and Library Preparation

Clover samples were collected along a transect consisting of three plots spaced 5 m apart in the Woods Hole Ball Field and Taft's Playground in Falmouth, MA (41°31'42.85"N, 70°40'12.59"W, Figure S1). Each sample contained at least 4 g of aboveground plant material collected in an area of one square yard and was stored in a clean, unused zip top plastic bag. Sampling equipment was surface sterilized between samples. Samples were collected every 2 weeks, allowing *T. repens* biomass to recover in between sampling events. Hereafter, we refer to the three sampled patches as A, B and C, and the time points of sampling as $t = 0, 2, 4$ and 6 weeks. Thus, the sample from patch A at time 0 is designated 'A0' and so on.

Plant material was pulverized using sterile ceramic mortars and pestles in 30 mL of sterile 0.85% NaCl, incubated overnight in a 50 mL Falcon tube, and then pulverized again in 10 additional ml of buffer. To eliminate plant debris, samples were centrifuged 1h at 4000 rpm, and the supernatant was transferred to a new Falcon tube and centrifuged for 20 additional minutes. The supernatant was poured through a 0.2 micron filter. The filtrate was concentrated using a

Centricon Plus-70 (Millipore, Ireland) by 20-30 min centrifugation at 4000 rpm. The concentrate was re-suspended in sterile 0.1x TE buffer (pH 8.0, Fisher Waltham, MA USA) and stored at -80° C. The plant material were transferred to 1.5-mL centrifuge with 0.85% NaCl.

RNA was extracted using QIAamp Viral Extraction kit (QIAgen Hilden, Germany), treated with Turbo DNase (Ambion Foster City, CA, USA), cleaned and concentrated using RNA Clean & ConcentratorTM-5 (Zymo Research Irvine, CA, USA), and converted to cDNA using Ovation RNA-Seq System V2 (NuGEN San Carlos, CA, USA). Libraries were prepared using Ovation Rapid DR Multiplex System 9-16 (NuGEN San Carlos, CA, USA), and size selection performed using 1.5% Pippin Prep cassettes (Sage Science, Beverly, MA, USA). RNA, cDNA, and libraries were quantified using Quant-iT RiboGreen RNA, PicoGreen dsDNA Assay Kits (Life Technologies, Carlsbad, CA, USA), Eukaryote total RNA pico Bioanalyzer chips, and DNA 7500 Bioanalyzer chips (Agilent Technologies, Santa Clara, CA), as appropriate. Equimolar amounts of each library were mixed and run in 1.5% Pippin Prep cassettes (Sage Science). Samples were size-selected in the range of 450-650. Recovered size-selected libraries were concentrated using Ampure beads and adjusted for optimal sequencing using Quant-iTTM PicoGreen® dsDNA Assay Kit (Life Technologies, Carlsbad, CA, USA) and DNA 7500 Bioanalyzer chips (Agilent Technologies, St Clara, CA, USA). Sequencing was performed on Illumina HiSeq 2500 in the Josephine Bay Paul Center at Marine Biological Laboratory.

Bioinformatic analyses

The paired end reads were first adapter trimmed and quality filtered with trim_galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) using a quality threshold of 25 and length of 75; this was followed by additional filtering with prinseq-lite

(<http://prinseq.sourceforge.net>) to remove low complexity sequences (dust method, threshold 7) and to remove duplicate sequences. Reads were then mapped to the genome of the barrel clover *Medicago truncatula* (GenBank BioProject PRJNA10791) to remove host reads, and also to the genome of the bacteriophage ϕ X which is common Illumina control, using bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>). Reads were subsequently filtered to remove ribosomal sequences using ribopicker (<http://ribopicker.sourceforge.net>).

Initial read level metagenomics analysis was performed by applying BLASTx to the filtered reads against a database of all RefSeq genomes using DIAMOND (<https://github.com/bbuchfink/diamond>), with MEGAN (<http://www-ab.informatik.uni-tuebingen.de/software/megan6/>) used to map BLASTx hits to the GenBank taxonomy tree.

Metagenomics

Only filtered reads that did not map to the WCIMV virus genome were used for *de novo* assembly using SPAdes (<http://bioinf.spbau.ru/spades>), metagenomics of the assembled contigs was performed by applying BLASTx to the contigs against a non-redundant database of all protein sequences using DIAMOND. Metagenomic pie charts were generated using KronaTools (<https://github.com/marbl/Krona/wiki/KronaTools>).

Diversity measures within dominant viral species

Ultra-high coverage of the dominant viral species WCIMV allowed examination of within-species genetic diversity. WCIMV was subset by filtering reads that aligned to the RefSeq genome of WCIMV (GenBank accession number NC_003820.1) using bwa (<http://bio-bwa.sourceforge.net>); an iterative mapping approach was used to call the consensus sequence of

the WCIMV in each sample (involving re-alignment to the initial consensus sequence, and re-calling of the consensus). The consensus sequences were then combined with all available WCIMV complete genomes on GenBank, aligned using mafft (<https://mafft.cbrc.jp/alignment/software/>) and a maximum likelihood tree was constructed using MEGA7 with bootstrap of 100 and the GTR (G+I) substitution model. Intra-host diversity of each sample was investigated using DiversiTools (<http://josephhughes.github.io/DiversiTools/>) to quantify the amount of variation in each sample and determine whether sub-consensus mutations were synonymous/non-synonymous.

After WCIMV was subset, diversity within the populations was examined. Mean entropy for each sample was calculated to examine the viral complexity of the populations [34]. Shannon entropy was calculated over each nucleotide site and then averaged over the entire genome to measure the amount of “disorder” in a population. Next, the degree of polymorphism was quantified at every nucleotide site across the genome, creating a mutation spectrum [34], to assess the heterogeneity of WCIMV populations. Mismatch frequency of the aligned reads to the consensus sequence was determined for each sample, and each nucleotide site in the WCIMV genome was binned by the proportion of reads by the mismatch frequency at that nucleotide site. Note: although conceptually similar, this mutation spectrum differs from the “spectrum of spontaneous mutations” as described in [35], which refers to characterization of the types of mutations, *e.g.* the number of transitions and transversions, *etc.*

To further examine the polymorphic diversity within the viral populations, haplotypes were constructed using QuRe (<https://sourceforge.net/projects/quire/>) and PredictHaplo (<http://bmda.cs.unibas.ch/software.html>). Due to the ultra-high depth of the WCIMV samples and in order for haplotype reconstruction to complete in a reasonable time, reads aligning to the

white clover mosaic virus genome were subsampled to give an even coverage of 10,000 across the WCIMV genome.

Results

Metavirome production and contig assembly

We characterized the diversity of viral communities associated with *T. repens* from 3 plots separated by 5 m at 4 time points separated by 2 weeks with 12 samples collected in total. Sample C4 was lost in transport, so it was excluded in our analyses. For our sample preparation, we modified a protocol originally designed to isolate viruses in human stool samples for plant tissue. Our approach produced high quality results. The number of raw reads ranged between 2-24 million reads per sample. 79-99.6% reads per sample passed our quality filters, ranging from 1.6-2.1 million reads per sample. After subsequently removing reads that aligned to *M. trunculata*, the closest relative to *T. repens* on GenBank, 76-85% aligned to known viruses with 11 unique viral 11 families represented.

Assembly of metagenomic reads produced 214-34,279 contigs per sample ($\mu = 13927$, $\sigma = 11613$). The average length of the contigs was XX. Consistent with the read alignments, the contigs are dominated the *Alphaflexiviridae* family. The remaining 161 contigs were assigned to the families *Podoviridae*, *Siphoviridae*, *Myoviridae*, *Betaflexiviridae*, *Luteoviridae*, *Potyviridae*, *Virgaviridae*, *Closteroviridae*, *Cystoviridae*, and *Rhabdoviridae*.

Taxonomic composition of the viral community associated with T. repens

We predicted that the viral communities associated with *T. repens* would be dominated by a few viral species with many more rare species. Specifically, we predicted that we would

detect WCIMV and CIYMV based on previous studies of viruses typically detected infecting *T. repens*. Our results confirmed that the communities were consistently dominated by these two viral species, with WCIMV as the most frequent. Notably, these two species were extremely common relative to the other virus species detected in the viral community. After filtering, WCIMV accounted for 50-83% of the reads in each sample, and that for CIYMV was 2.2-24%. These two species consistently dominated the viral communities, accounting for 76-85% of the filtered reads (Figure 1). Furthermore, when CIYMV accounted for only a small proportion of the reads, WCIMV compensated by comprising a larger proportion of the remaining viral community rather than another viral species. We examined the distribution of the viral species within the communities associated with *T. repens* after removing WCIMV and CIYMV. We found that removal of WCIMV and CIYMV showed that other members of the *Alphaflexiviridae* family still dominated the remaining identifiable species (Figure 2, Table S2). 76-85% of the filtered reads aligned to known sequences in RefSeq included other plant viruses, e.g. *Clover yellow vein virus*, *Potato yellow dwarf nucleorhabdovirus*, and *Beet yellow stunt virus*, as well as bacteriophages (e.g. *Pseudomonas* phage ϕ 2954, *Klebsiella* phage SopranoGao, and *Shigella* phage Sf11), and viruses associated with invertebrates (e.g. Hubei picorna-like virus 51 [36]).

Analysis of the dominant viral species: WCIMV

WCIMV consistently dominated the viral communities across all samples, composing approximately two thirds of all RefSeq hits ($\mu = 0.667$, $\sigma = 0.117$). Given the high coverage ($\mu = 8.32 \times 10^6$, $\sigma = 4.36 \times 10^6$ read pairs per sample), we were able to reconstruct the full viral genome from each sample and also analyze the dynamics of WCIMV. We compared the consensus sequence of WCIMV from each sample to all other complete WCIMV genome

sequences on GenBank on a phylogenetic tree (Figure 3). Our samples clustered with the only other sample of WCIMV from North America and the original sample of WCIMV sequenced from New Zealand. The other sample from New Zealand clustered with samples from South Korea and Japan. As for our samples, all time points from plot B clustered together as did the majority of samples from plot A; sample A2 is an outlier that does not cluster with the other samples. Although plot C samples do not form a distinct clade, they cluster with the other WCIMV consensus sequences from this study.

In order to examine the dynamics of WCIMV, we characterized the complexity of the viral quasispecies within each sample using Shannon entropy. This metric assesses the diversity of the genomes within a sample. The mean Shannon entropy per nucleotide site within a sample was relatively high with values between 0.105-0.109 for all time points at plots A and B. However, the entropy varied more for plot C ranging 0.093-0.106 (Figure 4A). Of the 5843 nucleotide sites, the number of nucleotide sites that is invariant (Shannon entropy = 0) is 1-103 per sample ($\mu = 25.5$, $\sigma = 31.7$), although this may be due in part to the high coverage. Next, we generated a mutation spectrum [37] to examine the frequency of mismatches [non-reference bases] across all genome positions (Figure 4B). The spectrum groups genome positions by their total mismatch frequency, to give an overview of the average mismatch frequency throughout the genome. We compared the mutation spectrum of the samples against the spectrum of another RNA virus, namely foot-and-mouth disease virus (FMDV) from a single foot lesion [38] for reference. The samples of WCIMV clearly have a substantially higher heterogeneity than the FMDV sample, with a large proportion of genome positions displaying mismatch frequencies greater than 10%. Further investigation of these mutations showed that the vast majority were non-synonymous suggesting that these were functional mutations and not mutation errors

introduced during sequencing or sample preparation. Due to the number of high frequency mutations observed within the viral population, we suspected that the sample contained multiple strains of WCIMV and therefore proceeded to haplotype reconstruction. The phylogeny of the reconstructed haplotypes (Figure 4C) suggested that the haplotypes could be divided into two different groups with high confidence, supporting the hypothesis that two different strains of WCIMV coexisted within the samples.

Discussion

We sampled the RNA viral communities associated with a single host species over a scale of meters and weeks. We found a striking pattern of viral diversity in the communities associated with *T. repens*: all communities were dominated by a single viral species, WCIMV. The second most abundant species was its sister species, CIYMV. Most of the remaining rare species were also members of the *Alphaflexiviridae* family, but there were no discernible patterns of diversity among the very rare viruses outside this family. Because we were able to recover a high level of reads per sample, we were able to identify diversity within a species rather than higher taxonomic ranks. Although we saw low diversity at the community level, we observed very high levels of diversity within the dominant species. This enabled us to examine the population dynamics of WCIMV further. We found more similarity among WCIMV reads by sampling location rather than time point and the consensus sequence for each of our samples tended to cluster together when compared with all other whole genome sequences on GenBank. Also noteworthy is that such a high proportion of the filtered reads (76-85%) aligned with known viral genomes on RefSeq. For instance, only 10% of the virus-derived sequences from plants

belong to known virus species [39]. This may be explained by the fact that our samples are dominated by known viruses.

Since we were able to recover such a high number of reads per sample, we want to highlight the steps in our sample preparation that we believe contributed to our success. We modified the sample preparation as described by Minot and colleagues [40], which isolated DNA viruses from human stool samples, for plant samples. Similarly, we used a large plant sample, ground it, and mixed it with a large volume of buffer. The large volume allowed us to mechanically lyse cells and separate cellular debris from the plant while the filtration steps allowed us recover viruses in a concentrated form. Unlike Minot et al [40], we opted to omit their chloroform step to avoid disrupting membrane-enclosed viruses. Additionally, they used a different set of kits for genome extraction and library preparation, because their study focused on DNA viruses. Regardless of those differences, both studies recovered a high number of reads (ours: 2-24+ million reads, theirs: 15-39 million) per sample. To our knowledge, this is the first publication of this methodology as applied plant viral metagenomes. This approach ultimately proved successful at separating viral and host genomes: we found that >10% of the samples aligned with the host reference genome, *Medicago trunculata*, a relative of *T. repens*. In comparison with other methods for isolating viral communities associated with plants, this method is relatively unbiased (see [39] for a review). The only amplification step occurred during the library preparation; the kit amplified samples with random primers so no bias is anticipated. Overall, this approach is a promising method for examining viral communities associated with plants.

The high diversity that our populations exhibited suggest that these plants may have long-term WCLMV infections in *T. repens*. WCLMV is a long-term infection, rather than an acute

infection, allowing viral diversity to arise and accumulate via mutation. Since *T. repens* is a perennial plant and can survive for many years, these viral infections may also be several years old. Additionally, these samples were collected from a municipal ball field that is regularly used, meaning that normal use and mowing can allow mixing of viruses from throughout the field. Furthermore, the mowing may introduce virus from other recently mown fields. These factors may help account for the high Shannon entropy values. In comparison, the Shannon entropy values for individuals experimentally inoculated with an acute infection of Foot and Mouth Disease Virus (FMDV) maxed out at 0.0008 [34], while the values for our populations peaked at 0.11.

The second most abundant viral species in our communities was CIYMV. This virus has also been associated with white clover for decades. Interestingly, CIYMV was historically limited to the Western US [30]. However, a more recent disease note documented CIYMV east of the Mississippi River for the first time, in *Verbena canadensis*, which is in a different family from *T. repens*, in Florida [41]. Furthermore, CIYMV appears to be expanding its geographical ranges recent reports for the first time in the UK [42] and in *T. repens* in Australia [43]. Taken together, this suggests that CIYMV may be expanding both its host and geographical range, first jumping into ornamental *Verbena* spp. and then moving globally through the sale of live plant material [44].

Beyond the two dominant viral species, most of the viruses we detected were also from the family *Alphaflexiviridae*, but we were not able to identify many of these reads beyond the XX taxonomic rank. Although we were not targeting DNA viruses with our methods, we did detect several DNA viruses present among the rare species, which we likely detected during transcription. Specifically, we detected three tailed phage families—*Myoviridae*, *Siphoviridae*,

and *Podoviridae*—all of which have a linear dsDNA genomes. This is consistent with other studies, *e.g.* the findings from a study of viral communities associated with *Capsicum annuum*: despite extracting RNA from the plant tissue, ~3/4 of the viral reads aligned DNA viruses [45]. If these results are real, this suggests that this method was able to detect DNA viruses during transcription and that these phages were actively replicating during infection of bacteria associated with the plant.

Acknowledgements

We thank researchers in the Josephine Bay Paul Center at Marine Biological Laboratory for helpful advice and useful comments on early versions of the manuscript.

References

1. Suttle CA. Viruses in the sea. *Nature* 2005; **437**: 356–361.
2. Cesar Ignacio-Espinoza J, Solonenko SA, Sullivan MB. The global virome: Not as big as we thought? *Curr Opin Virol* 2013; **3**: 566–571.
3. Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. *Nature* 1999; **399**: 541–548.
4. Morella NM, Gomez AL, Wang G, Leung MS, Koskella B. The impact of bacteriophages on phyllosphere bacterial abundance and composition. *Mol Ecol* 2018; **27**: 2025–2038.
5. Evans C, Pearce I, Brussaard CPD. Viral-mediated lysis of microbes and carbon release in the sub-Antarctic and polar frontal zones of the Australian Southern Ocean. *Environ Microbiol* 2009; **11**: 2924–2934.
6. Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A, et al. A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME J* 2015; **9**: 1352–1364.
7. Breitbart M. Marine Viruses: Truth or Dare. *Ann Rev Mar Sci* 2012; **4**: 425–448.
8. de Cárcer DA, López-Bueno A, Alonso-Lobo JM, Quesada A, Alcamí A. Metagenomic analysis of lacustrine viral diversity along a latitudinal transect of the Antarctic Peninsula. *FEMS Microbiol Ecol* 2016; **92**: 1–10.
9. Whipps JM, Hand P, Pink DAC, Bending GD. Human Pathogens and the Phyllosphere. *Adv Appl Microbiol* 2008; **64**: 183–221.
10. Malmstrom CM, Melcher U, Bosque-Pérez NA. The expanding field of plant virus ecology: Historical foundations, knowledge gaps, and research directions. *Virus Res* 2011; **159**: 84–94.

- 375 11. Power AG. Community Ecology of Plant Viruses. In: Roossinck MJ (ed). *Plant Virus*
376 *Evolution*. 2008. Springer, Berlin, pp 15–26.
- 377 12. Power AG, Flecker AS. Virus Specificity in Disease Systems: Are Species Redundant?
378 *The Importance of Species*. 2015. pp 330–346.
- 379 13. Roossinck MJ. Lifestyles of plant viruses. *Philos Trans R Soc B Biol Sci* 2010; **365**: 1899–
380 1905.
- 381 14. Gergerich R, Dolja V V. Introduction to plant viruses, the invisible foe. *Plant Heal Instr*
382 2006.
- 383 15. Brum JR, Ignacio-espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, et al. Ocean
384 Viral Communities. *Science (80-)* 2015; **348**: 1261498-1–11.
- 385 16. Steward GF, Wikner J, Cochlan WP, Smith DC, Azam F. Estimation of virus production
386 in the sea: 2. Field results. *Aquat Microb Ecol* 1992; **6**: 79–90.
- 387 17. Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, et al. Genomic
388 analysis of uncultured marine.pdf. 2002.
- 389 18. Weinbauer MG. Ecology of prokaryotic viruses. *FEMS Microbiol Rev* 2004; **28**: 127–181.
- 390 19. Comeau AM, Tremblay D, Moineau S, Rattei T, Kushkina AI, Tovkach FI, et al. Phage
391 morphology recapitulates phylogeny: The comparative genomics of a new group of
392 myoviruses. *PLoS One* 2012; **7**: 1–11.
- 393 20. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, et al.
394 Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like
395 myoviruses from diverse hosts and environments. *Environ Microbiol* 2010; **12**: 3035–
396 3056.
- 397 21. Steward GF, Culley AI, Mueller JA, Wood-Charlson EM, Belcaid M, Poisson G. Are we

missing half of the viruses in the ocean? *ISME J* 2013; **7**: 672–679.

22. Woolhouse MEJJ, Gowtage-Sequeria S. Population biology of emerging and re-emerging pathogens. *Emerg Infect Dis* 2005; **11**: 1842–7.

23. Frame J, Newbould P. Agronomy of white clover. *Advances in Agronomy* . 1986. Academic Press.

24. Williams WM, Nichols SN. Trifolium. *Wild Crop Relatives: genomic and breeding resources*. 2011. Springer, Berlin, Heidelberg, pp 249–272.

25. Zeven AC. Four hundred years of cultivation of Dutch white clover landraces. *Euphytica* 1991; **54**: 93–99.

26. Elgersma A, Hassink J. Effects of white clover (*Trifolium repens* L.) on plant and soil nitrogen and soil organic matter in mixtures with perennial ryegrass (*Lolium perenne* L.). *Plant Soil* 1997; **197**: 177–186.

27. George J, Dobrowolski MP, Van Zijll De Jong E, Cogan NOI, Smith KF, Forster JW. Assessment of genetic diversity in cultivars of white clover (*Trifolium repens* L.) detected by SSR polymorphisms. *Genome* 2006; **49**: 919–930.

28. Van Treuren R, Bas N, Goossens PJ, Jansen J, Van Soest LJM. Genetic diversity in perennial ryegrass and white clover among old Dutch grasslands as compared to cultivars and nature reserves. *Mol Ecol* 2005; **14**: 39–52.

29. Barnett OW, Diachun S. Virus diseases of clovers. *Clover Sci Technol* . 1985. , **25**: 235–268

30. Pratt MJ. Studies on clover yellow mosaic and white clover mosaic viruses. *Can J Bot* 1961; **39**: 655–65.

31. Dudas B, Woodfield DR, Tong PM, Nicholls MF, Cousins GR, Burgess R, et al.

Estimating the agronomic impact of white clover mosaic virus on white clover performance in the North Island of New Zealand. *New Zeal J Agric Res* 1998; **41**: 171–178.

32. Biddle JM, Linde C, Godfree RC. Co-infection patterns and geographic distribution of a complex pathosystem targeted by pathogen-resistant plants. *Ecol Appl* 2012; **22**: 35–52.

33. McKirdy SJ, Jones RAC, Sivasithamparam K. Determining the effectiveness of grazing and trampling by livestock in transmitting white clover mosaic and subterranean clover mottle viruses. *Ann Appl Biol* 1998; **132**: 91–105.

34. Morelli MJ, Wright CF, Knowles NJ, Juleff N, Paton DJ, King DP, et al. Evolution of foot-and-mouth disease virus intra-sample sequence diversity during serial transmission in bovine hosts. *Vet Res* 2013; **44**: 1–15.

35. Tromas N, Elena SF. The rate and spectrum of spontaneous mutations in a plant RNA virus. *Genetics* 2010; **185**: 983–989.

36. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, et al. Redefining the invertebrate RNA virosphere. *Nature* 2016; **540**: 539–543.

37. Wright CF, Morelli MJ, Thebaud G, Knowles NJ, Herzyk P, Paton DJ, et al. Beyond the Consensus: Dissecting Within-Host Viral Population Diversity of Foot-and-Mouth Disease Virus by Using Next-Generation Genome Sequencing. *J Virol* 2011; **85**: 2266–2275.

38. Orton RJ, Wright CF, Morelli MJ, Juleff N, Thébaud G, Knowles NJ, et al. Observing micro-evolutionary processes of viral populations at multiple scales. *Philos Trans R Soc B Biol Sci* 2013; **368**.

39. Roossinck MJ. Plants, viruses and the environment: ecology and mutualism. *Virology*

2015; **479–480**: 271–277.

40. Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD. Rapid evolution of the human gut virome. *Proc Natl Acad Sci U S A* 2013; **110**: 12450–12455.

41. Baker CA, Beckham K, Hiebert E. A virus related to clover yellow mosaic virus found east of the Mississippi River in *Verbena canadensis* in Florida. *Plant Dis* 2004; **88**: 223–223.

42. Mumford RA, Jarvis B, Harju V, Elmore J, Skelton A. The first identification of two viruses infecting trailing verbena in the UK. *Plant Pathol* 2005; **54**: 568.

43. Guy PL. First Record of Clover yellow mosaic virus from White Clover (*Trifolium repens*) in Australia . *Plant Dis* 2019; **103**: 1801.

44. Boonham N, Tomlinson J, Mumford R. Microarrays for Rapid Identification of Plant Viruses. *Annu Rev Phytopathol* 2007; **45**: 307–328.

45. Jo Y, Choi H, Kim SM, Kim SL, Lee BC, Cho WK. The pepper virome: Natural co-infection of diverse viruses and their quasispecies. *BMC Genomics* 2017; **18**: 1–12.

Figure 1. Summary of the read statistics for each sample. Plots A, B, and C were separated by 5 m with samples taken at 0, 2, 4, and 6 weeks. Bars represent the frequency of the read pairs that remained after filtering, matched an existing sequence in the RefSeq database, matched WCIMV, CIYMV, or failed to match to any sequence in RefSeq.

Figure 2. The rare virus families present in each of the samples after the dominant viral species (WCIMV and CIYMV in the *Alphaflexiviridae* family) were removed. Note that the *Alphaflexiviridae* present in this sample are comprised viral species other than WCIMV and CIYMV.

Figure 3. Phylogenetic tree for all samples of WCIMV combined with all complete genome sequences of WCIMV from GenBank.

Figure 4. Shannon entropy of the WCIMV of each of the samples. A) Shannon entropy (averaged over all genome positions for each sample) for each plot over time. Samples collected from plot A are in green, B in light green, and C in chartreuse. B) Mutation spectra of each sample. The mutation spectrum for FMDV (Orton et al 2013) is included as a reference. C) Haplotype reconstruction of WCIMV pooled across sample.