Virus discovery in winter-growing perennial plants of southern California sage scrub habitat

Amani Helo, Undergraduate Student, Department of Biology
Tessa Shates, PhD Candidate, Department of Entomology
Kerry Mauck, Assistant Professor of Entomology, Department of Entomology

Abstract

Plant viruses cause serious disease and losses in domesticated crops. However, we know little about plant viruses outside of agriculture. One reason for this is the lack of symptoms of virus infection in wild plants to promptly diagnose and identify with targeted methods. This is now changing with the availability of untargeted "next-generation" sequencing technologies to analyze the viromes of asymptomatic wild plants and study virus impacts on plant health. In this study, we determined whether key winter-growing perennials in local reserves are infected with crop-associated viruses. A previous study in the same locations found that summer-growing perennials host co-infections by multiple crop-associated viruses, but winter-growing plants have not been explored. To enrich for virus sequences, we extracted doubled-stranded RNA, a unique feature of virus replication. We sequenced this material using the Illumina NextSeq platform, then assembled and identified viruses using Galaxy software. In contrast to summer-growing plants, we detected just one crop-associated virus in winter plants: the generalist, *Cucumber mosaic virus* (CMV). Phylogenetic analysis shows this CMV is in subgroup IA, the most prevalent in the United States. Our results suggest that winter-growing plant viromes have minimal overlap with summer-growing plant viromes and that winter plants may be less exposed to crop-associated viruses.

KEYWORDS: *Virome, Ecology, Diagnostics, Bioinformatics, Phylogeny*

Introduction

Plant viruses are important causal agents to infectious diseases of wild plants and cultivated crops. A better understanding of plant viruses and their origins is essential to manage emerging threats and to measure impacts on conservation efforts. Advancements in scientific technology now allow for the discovery and characterization of the communities of known and novel viruses in wild vegetation using high throughput, next generation sequencing that does not require prior knowledge of the targets. Such technological advancements are increasing the rate at which viruses are discovered and characterized, but there are still only a few studies focused on the communities of viruses infecting wild hosts, especially long-lived perennials that may be exposed to virus infections over multiple seasons. In contrast, there is a vast knowledge of cultivated plants and their corresponding viruses, as well as how they are transmitted among plants by insect vectors and human activities (Shates et al., 2019).

This study addresses the key knowledge gap of virus communities in wild perennial plants by using next-generation sequencing to characterize the viromes of *Solanum xanti* (SX), *Marah macrocarpa* (MM), and *Acmispon glaber* (AG), three cool-season species endemic to Southern California sage scrub habitats. These plants were selected for the study because they are common and abundant. They are also important resources for local insect pollinators, and they are large and apparent to insect herbivore vectors. We chose to sample populations in two natural reserves (Motte Rimrock Reserve [Reserve DOI: 10.21973/N31T0W] and Shipley Skinner Multispecies Reserve) as these areas are adjacent to both urban and agricultural spaces, and thus, plants in these preserved habitats may be exposed to pathogens from these sources. In prior work performed in the same reserves, it was found that plants in the same families as those chosen in this study, which only grow in the summer season, are exposed to insect vectors emigrating from crops, carrying crop-associated viruses (Shates et al., 2019). Crop-associated viruses are common in summer-growing perennials in both reserves, and plants also support large populations of insect vectors found in crops. Based on these results, we expected to find crop-associated viruses in winter season plants, but to potentially find fewer co-infections given that vector populations are much lower in the winter season (Shates et al., 2019).

By studying the viromes of wild plants on a seasonal basis, this study contributes to our understanding of viral interactions with hosts in non-cultivated systems. In the field of virus ecology, most studies focus on virus movement from wild to agricultural plants, with very few studies addressing the movement of viruses from agricultural plants to wild plants from the perspective of how this affects the wild community. A key knowledge gap is the many ways that non-cultivated species are affected by vector insects and their associated pathogens. Pathogens, such as plant viruses, may even impair the expression of functional traits within the host plant

that impact the stress response, fitness (Alexander et al., 2017), and interactions with other organisms (Malmstrom et al., 2017). Morphological effects include stunting of growth and distorted leaves, which are symptoms of viral infection. Virus build-up in plants may also affect our efforts of conserving native plant communities. This study will contribute to our knowledge of the virus communities present in wild plants and aid in the understanding of how agriculture impacts wild communities in ways that are less directly observable.

Methods

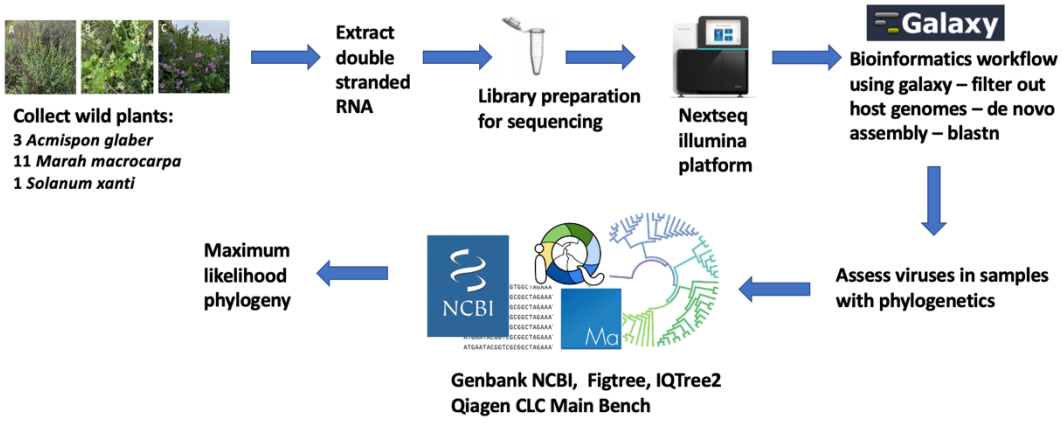
Target plant selection

To choose which plants to sample, preliminary surveys and online tracking were performed to identify important cool-season perennials in Southern California. User-generated species occurrence data found on iNaturalist (*iNaturalist*) and Calflora (*Calflora - Search for Plants*) were used to identify common species across locations and determine their leaf-out times. Within the list of common species, we further narrowed selections to focus on plants that are related to crops grown in Southern California (same family or same genus). We predicted that these hosts might be especially susceptible to vectors and viruses from crops and urban gardens. Based on these criteria, we selected three plant species for virome characterization **(Figure 1)**.

Figure 1 A-C: Selected perennial plant species. **(A)** *Acmispon glaber* (AG), **(B)** *Marah macrocarpa* (MM), and **(C)** *Solanum xanti* (SX).

The plant species chosen are prevalent in Southern California, as well as in the target reserves: Motte Rimrock Reserve (Reserve DOI: 10.21973/N31T0W) and Shipley Skinner Multispecies Reserve. These plants are confamilial and congeneric to locally grown crop plants. *Marah macrocarpa* is in the Cucurbitaceae family, which contains crops such as squash, pumpkin, and melons. *Acmispon glaber* is in the Fabaceae (legume) family, and is related to alfalfa, a common crop in the area of the reserves. *Solanum xanti* is in the same genus as tomato and potato, both of which are also grown in Riverside county. Following host plant selection, we proceeded with the workflow outlined in Figure 2 and detailed in subsequent sections.

Figure 2. Workflow diagram.



Collection & sample preparation methods

In March and April 2018, ten grams of leaf and stem tissues (and *A. glaber* flowers) were removed from each target plant by inverting a Ziplock bag and then cutting off the material before sealing. From Motte Rimrock, collection was as follows: one sample of SX, two of AG, and seven of MM. From Shipley Skinner, only four of MM and one of AG were collected. At the time of sample collection, there were more MM individuals leafed out and available for samplingthan AG and SX; therefore, an unequal amount of sampling was done. Samples were stored on dry ice before returning to the lab, where they were processed into 50-mL RNase-free

Falcon tubes and stored at −80°C until extracting. We extracted samples and performed sequencing following dsRNA extraction and library preparation protocols from Shates et al. (2019). The pooled library was sent to UCR Genomics Core for sequencing with the Illumina NextSeq platform with 75 base-pair paired ends with adaptors trimmed by Core staff, resulting in two raw data files per sample: one for "forward" and the other for "reverse" paired-end reads.

Virus identification

We used the online Galaxy platform (*Galaxy*) to perform a bioinformatics workflow on the sequencing outputs (originally developed by Shates et al., 2019) (Afgan et al., 2018). An important part of the workflow includes removing host genome sequences using the HISAT2 function, which maps sequences to an uploaded host genome. To filter out host plant sequences, we used the genomes of tomato for *Solanum xanti* (SX), pumpkin for *Marah macrocarpa* (MM), and birdsfoot trefoil for *Acmispon glaber* (AG). After removing potential host sequences, we performed de novo assembly using Trinity on unaligned forward and reverse reads (Grabherr et al., 2011). This is required in order to assemble the RNA sequence reads for alignment when lacking the reference genome. We imported a plant-infecting virus database from GenBank at NCBI into the Galaxy Platform and used the nucleotide blast (blastn) function to create a list of putative virus identity matches in our samples. We used an Excel macro in order to remove the duplicates generated by Galaxy and to create and sort a putative list of viruses. The cutoff for a species match was 90% or greater percent identity. To be included as a putative virus, the cutoff we used was 70% or greater percent identity to a known virus. However, we also used a cutoff of 70% of the genome to be represented in the sequencing to eliminate small fragments that match highly but are potentially false positives. We divided the size of the sequence fragment ("contig") by the full genome size of the virus to find that cutoff.

Phylogenetic analyses

The sample MRMM4 had 3/3 RNAs of the tripartite *Cucumber mosaic virus* (CMV). The most commonly analyzed region of the genome is the third RNA, which includes the coding region for the coat protein. We used an alignment of the coat protein from our sample and the coat proteins of CMV accessions publicly available on GenBank and cited by Lin et al. (2003), Nouri et al. (2014), and Roossinck (2002). We included two outgroups to root the analysis: *Peanut stunt virus* (PSV) & *Tomato aspermy virus* (TAV). Using Qiagen's CLC Main Workbench, we performed model testing on the alignment. The most fitting model for our alignment was the General Time Reversible with Rate and Topography Variation (GTR +G +T). To create the maximum likelihood phylogeny with 1000 bootstrap values, we used the fasta alignment file from CLC and uploaded to the publicly available platform IQTree (Trifinopoulos et al., 2016). We used the Newick Tree file generated from IQTree in subsequent tree visualizations. We used the software, FigTree, to visualize and modify the phylogeny (*FigTree*). When modifying the phylogenetic tree, we distinguished the different isolates (Outgroups, IA, IB, and II) in order to efficiently visualize the tree characteristics. Based on the clades of different isolate subgroups, we could see which subgroup the isolate was most similar to.

Results

Virus identification

Of all ten plants that were sequenced, only two contained sequences with a positive match to a known virus: MRMM4 and MRMM10 **(Table 1).** These samples of *M. macrocarpa* both contained *Cucumber mosaic virus*. The samples contain all three RNAs represented by this tripartite RNA-genome virus. Only one plant, MRSX1, has putative matches for novel viruses, one with a 76.9% nucleotide identity to *Verticillium dahliae partitivirus 1* isolate segment of

RNA1, and another with 77.1% percent identity to *Verticillium dahliae partitivirus 1* isolate segment of RNA2. This plant also has a match of 72.1% to *Botryotinia fuckeliana partitivirus 1* genomic RNA.

Table 1: Summary of virus detections by untargeted sequencing

Plant Code	Species	Reserve	Crop-Associated Viruses Detected	Novel Viruses Detected
MRAG1	*Acmispon glaber*	Motte Rimrock	0	0
MRMM1	*Marah macrocarpa*	Motte Rimrock	0	0
MRMM3	*Marah macrocarpa*	Motte Rimrock	0	0
MRMM4	*Marah macrocarpa*	Motte Rimrock	1	0
MRMM5	*Marah macrocarpa*	Motte Rimrock	0	0
MRMM8	*Marah macrocarpa*	Motte Rimrock	0	0
MRMM9	*Marah macrocarpa*	Motte Rimrock	0	0
MRMM10	*Marah macrocarpa*	Motte Rimrock	1	0
MRSX1	*Solanum xanti*	Motte Rimrock	0	2
SSAG1	*Acmispon glaber*	Shipley Skinner	0	0

Phylogenetic analysis

Phylogenetic analyses infer relationships based on the number of common ancestors shared. We found that the CMV coat protein sequence from MRMM4 from Motte Rimrock Reserve groups with the CMV IA genetic subgroup **(Figure 3)**. In the phylogeny **(Figure 3)**, the bootstrap support for the split between subgroups II and IA/IB is 99, which indicates high support. The split between IA and IB is less supported (bootstrap 52), but the clades are consistent with previous publications (Lin et al., 2003; Nouri et al., 2014; Roossinck, 2002).

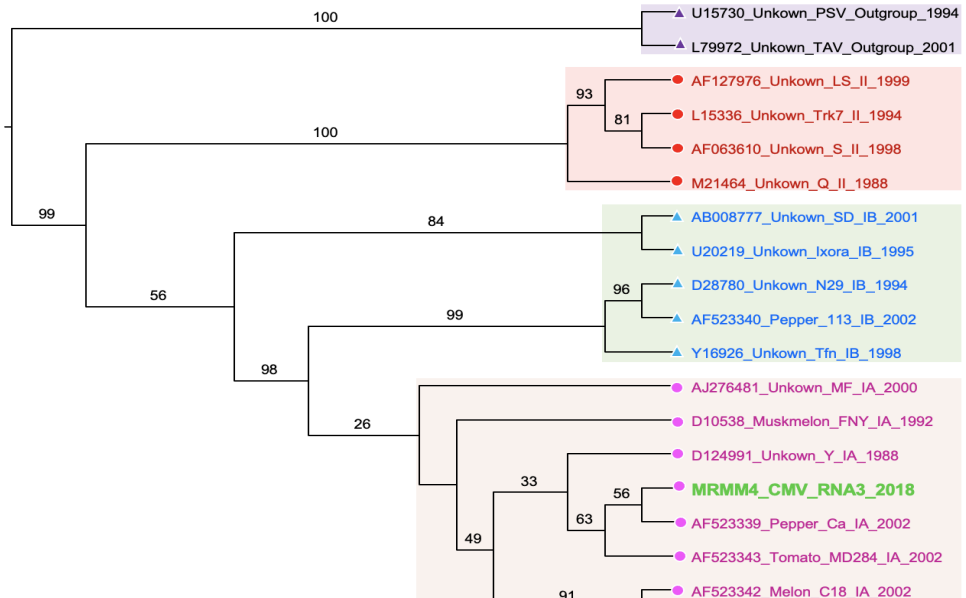


Figure 3. Rooted Phylogenetic tree illustrating results of the maximum likelihood analysis of CMV coat protein nucleotide sequences. The GTR + G +T model was used for tree construction. Color coding and designated shapes have been applied to aid interpretation of isolate origin (purple = unknown outgroups, red = CMV isolate II, green = CMV isolate IB, orange = CMV isolate IA).

Discussion

Using non-targeted virome sequencing, we were able to characterize the viromes of cool-season perennials in California sage scrub habitats. Consistent with our prediction that cool-season plants would harbor fewer crop-associated viruses, we only detected one virus CMV in two samples of one host species (*M. macrocarpa*). These detections suggest that *M. macrocarpa* is susceptible to CMV (Ng and Perry, 2004). *Cucumber mosaic virus* (family Bromoviridae) is an RNA plant virus with a tripartite genome and with the broadest host range of all characterized plant viruses (over 1200 known hosts). It is transmitted by over 80 species of aphids through brief probes of plant tissue—meaning the aphids just need to taste the plant to infect it and do not need to stay and feed or be able to reproduce on it. Based on this, it is not surprising that CMV was present in at least one of the target species, as it is probably one of the most prevalent viruses in both wild and cultivated habitats worldwide and has the most opportunities for transmission. Previous evolutionary analyses with the coat protein of RNA3 suggest that CMV is subdivided into two main genetic groups (I and II), with group I further divided into subgroup IA, IB (Nouri et al., 2014). We found that the isolate from MRMM4 likely belongs to subgroup IA, which is the most prevalent isolate found in the United States (Nouri et

al., 2014). The separation between IA and IB bootstrap support of 56 is low. However, it is still consistent with the split in published literature. This support might be low because of the sequences chosen for this analysis. If other accessions were chosen, the value might be different.

Beyond CMV, very few viruses were detected in these hosts. The only other putative viruses found were in *S. xanti* and matched to sequences of known fungi-infecting viruses in the family Partitiviridae **(Table 1)**. Thus, the only plant virus we detected is CMV, which is in strong contrast with a study performed on hot-season (summer-growing) perennials from the same locations, where all plants were found to be infected with multiple viruses, and a total of six crop-associated viruses were detected (Shates et al., 2019). This contrast may be due to how summer-growing plants are more exposed to crop viruses. These summer-growing plants may also be benefiting from some members of the virus community that aid in tolerating drought conditions, Partitiviridae being among these viruses present (Shates et al., 2019).

Conclusion

Viruses are one of the most common microbial effectors in cellular life, including the infectious diseases of plants and commercialized crops. We did an untargeted search for all viruses in three cool-season perennial plants and found only one crop-associated virus (CMV): a species with a very broad host range. This finding prompted further analysis to determine the relationship of the virus to other isolates taken from crops. Our finding that our isolate is likely in subgroup IA was expected, as the subgroup I is most prevalent in the U.S. according to Nouri et al. (2014). Although we did not sample a large number of plants, the lack of other plant virus detections in the target hosts suggests that winter-growing plants may be less exposed to crop-associated viruses (perhaps due to lower vector numbers in winter) and are not routinely acquiring viruses from summer-growing plants that host many co-infections (Shates et al., 2019).

Future studies on these undomesticated plants will further close knowledge gaps as we learn more about viromes of wild communities across seasons. This work will also aid in assessing threats to our native plants from local food production at the same time as monitoring for agricultural threats.

Acknowledgements

I would like to extend my sincere gratitude to my incredible mentor, Dr. Mauck, for her guidance. The constant support I have received in the past three years in Mauck research lab has allowed me to gain a wealth of knowledge and has enabled me to grow as a researcher, student, and individual. I would also like to thank my amazing PhD candidate graduate student mentor, Tessa Shates, for all her patience in helping me gain a stronger understanding of this project. With her dedication and hard work, she helped me bring this project to life.

References

Afgan, Enis, et al. “The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2018 Update.” *Nucleic Acids Research*, vol. 46, no. W1, July 2018, pp. W537–44.

Alexander, Helen M., et al. “Crop‐associated Virus Infection in a Native Perennial Grass: Reduction in Plant Fitness and Dynamic Patterns of Virus Detection.” *The Journal of Ecology*, vol. 105, no. 4, Wiley, July 2017, pp. 1021–31.

Calflora. https://www.calflora.org. Accessed 28 Feb. 2021.

FigTree. http://tree.bio.ed.ac.uk/software/figtree/. Accessed 1 Mar. 2021.

Galaxy. https://usegalaxy.org/. Accessed 18 Feb. 2021.

Grabherr, Manfred G., et al. “Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome.” *Nature Biotechnology*, vol. 29, no. 7, May 2011, pp. 644–52.

iNaturalist. https://www.inaturalist.org/. Accessed 28 Feb. 2021.

Lin, Han-Xin, et al. “Genetic Diversity and Biological Variation among California Isolates of Cucumber Mosaic Virus.” *The Journal of General Virology*, vol. 84, no. Pt 1, Jan. 2003, pp. 249–58.

Malmstrom, Carolyn M., et al. "Crop-associated virus reduces the rooting depth of non-crop perennial native grass more than non-crop-associated virus with known viral suppressor of RNA silencing (VSR)." *Virus Research,* 241 (2017): 172-184.

Ng, James C. K., and Keith L. Perry. “Transmission of Plant Viruses by Aphid Vectors.” *Molecular Plant Pathology*, vol. 5, no. 5, Sept. 2004, pp. 505–11.

Nouri, Shahideh, et al. “Genetic Structure and Molecular Variability of Cucumber Mosaic Virus Isolates in the United States.” *PloS One*, vol. 9, no. 5, May 2014, p. e96582.

Roossinck, Marilyn J. “Evolutionary History of Cucumber Mosaic Virus Deduced by Phylogenetic Analyses.” *Journal of Virology*, vol. 76, no. 7, Apr. 2002, pp. 3382–87.

Shates, Tessa M., et al. “Addressing Research Needs in the Field of Plant Virus Ecology by Defining Knowledge Gaps and Developing Wild Dicot Study Systems.” *Frontiers in Microbiology*, vol. 9, 2019, p. 3305.

Trifinopoulos, Jana, et al. “W-IQ-TREE: A Fast Online Phylogenetic Tool for Maximum Likelihood Analysis.” *Nucleic Acids Research*, vol. 44, no. W1, July 2016, pp. W232–35.