UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Viruses in Native California Perennial Plants: Advancing Virus Ecology by Developing New Study Systems for Manipulative Experiments

Permalink

https://escholarship.org/uc/item/4xz3r8f4

Author

Shates, Tessa Marie

Publication Date

2021

Supplemental Material

https://escholarship.org/uc/item/4xz3r8f4#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Viruses in Native California Perennial Plants: Advancing Virus Ecology by Developing New Study Systems for Manipulative Experiments

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Entomology

by

Tessa Marie Shates

December 2021

Dissertation Committee:
Dr. Kerry Mauck, Chairperson
Dr. Erin Wilson Rankin
Dr. Thomas Perring

Copyright by
Tessa Marie Shates
2021

The Dissertation of Tessa Marie Shates is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

I would first like to thank my amazing mentor and advisor, Dr. Kerry Mauck, for guiding me and motivating me through my doctoral studies. Through her mentorship, my experience completing my studies has been a fulfilling adventure.

I am also grateful to my dissertation and qualifying exam committee members – Dr. Erin Wilson Rankin, Dr. Thomas Perring, Dr. Richard "Rick" Redak, and Dr. Allison Hansen. I am especially grateful for support from Dr. Thomas Perring and his laboratory staff for allowing me use of their greenhouse space and for supporting my experiment's progress by watering and applying pest control treatments. I am also thankful for advice from Dr. Erin Wilson Rankin on statistical methods and analyses.

I would also like to thank the Entomology Department at UC Riverside. The office staff, graduate students, postdoctoral students, research staff, and faculty create an open environment that fosters support and collaboration. I am especially grateful to Dr. Jocelyn Millar, who co-advised me with Dr. Kerry Mauck for my first two and a half years of my graduate program.

I would like to thank my family and friends for supporting me and cheering me on throughout my graduate studies. I would especially like to thank my partner (Mike Stewart) and my sister (Julia Shates) for lending me a hand when I get stuck.

Chapter one was completed in collaboration with co-authors Dr. Penglin Sun, Chrysalyn (“Krissy”) Dominguez, Dr. Carolyn Marie Malmstrom, and Dr. Kerry Mauck. I am especially grateful for Dr. Penglin Sun for training me in techniques for this chapter, which I used throughout my graduate studies. I thank Krissy Dominguez for teaching basic concepts on sequence editing and analysis. And I thank Dr. Carolyn Malmstrom for contributing the network analysis. Funding for this project came from the Shipley Skinner Reserve - Riverside County Endowment

Chapter two was completed in collaboration with co-authors Dr. Penglin Sun, Oaksoe Aung, Amani Helo, Jaimie Kenney, and Dr. Kerry Mauck. I am especially grateful for the help collecting and processing samples from Oaksoe, Amani, and Jaimie. I also thank Dr. Penglin Sun for library preparation of samples before sequencing. I would also like to thank Dr. Marko J. Spasojevic, whose course Analysis of Ecological Communities provided the quantitative training and code for this project. Funding for this project came from the Shipley Skinner Reserve - Riverside County Endowment and the Howie Wier Memorial Conservation Grant – Anza-Borrego Foundation.

Chapter three was completed in collaboration with co-authors Dr. Marco Gebiola, Dr. Penglin Sun, Dr. Carolyn Marie Malmstrom, and Dr. Kerry Mauck. I am especially grateful to Dr. Marco Gebiola for the training and technical guidance he provided. I also thank Dr. Penglin Sun for library preparation of samples before sequencing. I am grateful for conversations with Dr. Carolyn

Malmstrom that gave me direction for approaching this chapter, and continuous help in troubleshooting analyses.

Chapter four was completed in collaboration with co-author Dr. Kerry Mauck. I am especially grateful for Dr. Kerry Mauck's work editing my writing and guiding my progress through this project. This work was funded by the remainder of Dr. Mauck's start-up funds through UC Riverside.

Dedication

I dedicate this dissertation to my grandparents, Olga and Michael Umansky.

ABSTRACT OF THE DISSERTATION

Viruses in Native California Perennial Plants: Advancing Virus Ecology by Developing New Study Systems for Manipulative Experiments

by

Tessa Marie Shates

Doctor of Philosophy, Graduate Program in Entomology
University of California, Riverside, December 2021
Dr. Kerry Mauck, Chairperson

Virus ecology in wild plants is a relatively new field of study that has emerged from disciplines of ecology, entomology, and plant pathology. In this dissertation, I critically review progress in plant virus ecology thus far to identify and address major research gaps, then perform a series of empirical studies to address these deficiencies. My analysis revealed that there is a deficit in research on perennial dicot plants, a weak understanding of the factors shaping virus community composition within and between hosts, and a lack of manipulative studies quantifying impacts of crop viruses on wild plant health. To address these understudied areas, I developed a wild, perennial plant study system consisting of three key hosts in local Riverside County reserves: *Cucurbita foetidissima*, *Cucurbita palmata*, and *Datura wrightii*. In Chapter 1 and Chapter 2, I characterized virus communities in several reserves then employed methods from community ecology to determine the factors that shape the virus communities in hosts, populations, and reserves. In Chapter Three, I determined the origins of the most prevalent aphid-transmitted virus in this system, *Cucurbit aphid-borne yellows virus*

(CABYV). And lastly, in Chapter 4, I used manipulative experiments combined with field observations to quantify CABYV effects on cucurbit performance and evaluated how virus infection modifies host interactions with a key aphid vector, *Aphis gossypii.* Overall, I found that crop-associated viruses are prevalent in the perennial target hosts, which often support multiple infections. Differences between hosts and reserves contribute to the structure of cucurbit virus communities, with some viruses emerging as key species. One of these species, CABYV, negatively affects its wild hosts in ways that are relevant for both virus and host persistence within the community.

Table of Contents

List of Figures

List of Tables

Introduction

Origins of plant virology

Plant viruses account for nearly half of the emerging infectious diseases in plants, and climate change is expected to exacerbate the impact of viral diseases on agriculture (Anderson et al., 2004; Ertunc, 2020). The first plant virus identified was *Tobacco mosaic virus* in 1935 (with earlier scientists in the 1890's ascribing symptoms to small bacteria or toxins). Since then, over 2000 plant viruses have been characterized, with more being discovered at a rapid pace (Agrios, 1997). Plant viruses are transmitted among hosts through a variety of pathways, some of which are driven by human agricultural activities. These include mechanical transmission through grafting stem tissue (often human-mediated) and natural root grafts, seed or vertical transmission where plant progeny carry viruses acquired from the mother plant, pollen transmission, transmission by soil-mobile fungi and nematodes, parasitic plants, and arthropods (mites and insects) (Agrios, 1997). Because of their economic impacts and the complexity of virus control through vector management, insect-transmitted viruses have been the focus of most plant virology research efforts to date.

Insects transmit viruses during foraging and feeding activities. Whether transmission occurs depends on where the virus resides in the plant host (e.g., mesophyll vs. vascular cells), where the virus attaches to or is sequestered in an insect, and whether the virus circulates or replicates within the insect vector (Hogenhout et al., 2008; Mauck et al., 2012; Ng & Falk, 2006). Stylet- and foregut-borne non-persistently transmitted (NPT) viruses are acquired and inoculated by short probes into leaf tissue performed

during host assessment and prior to long-term feeding (Martín et al., 1997). Persistently transmitted (PT) viruses are acquired during long ingestion and salivation periods in the plant phloem, followed by virus migration across gut wall membranes into the hemolymph and through salivary gland membranes, ultimately leading to virus accumulation in insect salivary glands (Prado & Tjallingii, 1994). Inoculation occurs when a viruliferous insect salivates into the phloem of a non-infected plant. Some persistently-transmitted viruses circulate harmlessly and accumulate in the to the salivary glands, while others can infect the vector, replicating in the salivary glands and other tissues.

The wide range of transmission modes and varying degrees of persistence in the vector make controlling insect-transmitted viruses in crops a challenging endeavor. Controlling viruses requires managing the vector at much lower economic thresholds relative to those for pests that damage crops directly, and therefore require different or more stringent actions for control (Broadbent, 1957; Perring et al., 1999; Stansly & Natwick, 2010). Compared to direct damage from feeding, damage from virus transmission requires fewer insects and therefore action must be taken at lower numbers. Some of the damage from viruses includes changes to the plant, such as reduced lifespan, stunting, yellowing and mosaic of leaves, and changes in fruit/seed output and quality. Epidemics of multiple viruses are associated with economic losses: up to 100% field losses of melons in the US southwest infected with *Cucurbit yellow stunting disorder virus* (CYSDV (Wintermantel et al., 2017); *Cucumber mosaic virus*, *Cowpea mild mottle virus*, and *Cowpea aphid-borne mosaic virus* in cowpea reducing yields by up to 95.4%

in Uganda (Amayo et al., 2012); yield losses up to 40-50% of a field infected with *Cucurbit aphid-borne yellows virus* (CABYV) (Lecoq et al., 1992), and many other examples. Global losses from viruses can be as high as 9% worldwide (for potato), and between 1-6% for wheat, rice, and maize (Oerke, 2006). On top of that, the regions hardest hit by crop pests and pathogens are those with growing human populations and food deficits (Savary et al., 2019).

Another unique aspect of insect-transmitted viruses is that they are mobile with their insect vectors and can travel beyond an original host, field, or region. Many insect vectors, such as aphids, whiteflies, and thrips have tremendous capability for dispersal and movement. For example, winged *Myzus persicae* (Sulzer) vector viruses to hosts that they do not colonize (Kennedy, 1950), and have been caught in low-level jet streams (Zhu et al., 2006). Most aphid migrations are local (<20 km distance), but there is evidence of further distances such as grain-associated aphids caught ~90-100km away from cultivated land in the Mojave desert (Loxdale et al., 1993). More recently, studies of vector dynamics in cucurbit agroecosystems demonstrated that landscapes between zero and five kilometers away from a crop field are significant predictors of the number of cucurbit-colonizing aphids, but landscapes five (or greater) kilometers were better predictors of the number of non-colonizing aphids trapped in cucurbit fields (Angelella et al., 2016). The sweet potato whitefly, *Bemisia tabaci* (Gennadius), is also capable of migration; laboratory flight chamber experiments show insects flying for an average of 15 minutes before responding to a vegetation cue, with some flying for as long as two hours (Byrne, 1999). Paired field experiments show whiteflies landing at distances of two

and five kilometers from release points (Byrne, 1999). When cultivated host plants are unavailable, these insects will move to wild plants, with viruses hitchhiking in stylets, foreguts, and salivary glands.

Insect Vectors & movement across the agro-ecological interface

Crop cultivation impacts viruses and their insect vectors, in both managed and natural systems. There is evidence that virus evolution and virulence are driven by repeated cycling and serial passage through crops (Gibbs et al., 2008; Mauck & Chesnais, 2020). New pathogen outbreaks in plant systems come from introductions of either a new plant, pathogen (or new isolate/strain), or vector insect (Jones, 2020). For example, Citrus tristeza virus in Riverside, California was not transmissible by the resident *Aphis gossypii* Glover, but changes in virus transmissibility allowed *A. gossypii* to spread the virus widely over a short period of time in California, and similarly in Israel with a strain from Sharon Plains (Roistacher et al., 1980). Movement and establishment of polyphagous invasive vectors can also drive the proliferation of multiple virus species. The global movement and emergence of viruses in the family *Geminiviridae* is driven by the spread of the invasive sweet potato whitefly, *B.tabaci* “MEAM1" or “B” biotype (Varma et al., 2011). Similarly, new species and resistance-breaking strains of tospoviruses, transmitted by thrips, have emerged in many locations around the world (Rojas & Gilbertson, 2008). The interactions between new pests, pathogens, and plants facilitated by agriculture is a major force in contributing to new disease outbreaks.

Beyond agriculture and global trade creating new opportunities for virus emergence and establishment, agriculture also creates different selection pressures on viral pathogens relative to wild habitats. In one example, heterogeneity of landscapes (and management by humans) influenced the prevalence, spatial genetic structure, and temporal dynamics of chiltepin-infecting begomoviruses. Managed chiltepin pepper fields had increased prevalence, symptoms, and co-infections relative to wild chiltepins in heterogeneous landscapes (Rodelo-Urrego et al., 2013). Disease risk in chiltepin was found to be positively correlated with human management, determined by a reduction in host genetic diversity and/or increase in host density, which supports the axiom in plant pathology that higher host density leads to higher disease risk (Pagán et al., 2012). In another example, simplified landscapes and monocrop systems had greater cereal and melon aphid abundances and virus prevalence (Hussein & Samad, 1993; Zhao et al., 2015). Viruses circulating in short-lived annual systems ultimately exist in dead-end hosts. Because of this, agricultural activity may select for viruses with higher virulence, replication rate, faster systematic colonization of hosts, and ability to manipulate hosts to increase the odds of vector transmission.

The impacts of agriculture on the traits of insect-transmitted viruses have clear negative consequences for crop yields. However, insect vectors are not bound by the borders of agriculture, and can fly between managed and natural systems, across the continuum of crop and wild plants commonly called the agro-ecological interface (Alexander et al., 2014) (Figure 1). Although only a handful of studies have explored crop virus infections in wild plants, the majority suggest that crop virus spillover

negatively affects plant health and fitness, especially when new viruses are introduced to a region (Malmstrom & Alexander, 2016). In one prominent example, introduced tospoviruses, potyviruses, luteoviruses, and poleroviruses reduced proxies for fitness and induced severe symptoms in native Australian vegetation (Vincent et al., 2014). Moreover, exposure of wild plant communities to crop viruses is increasing due to land-use changes. Most lands that were wild are now developed in some way and agriculture covers 40% of the ice-free land (Ramankutty et al., 2008); less than a quarter of ice-free land remains wild (Ellis & Ramankutty, 2008). This has led to smaller, fragmented wild/remnant lands that contain more edge habitat relative to interior habitat. Edges are particularly prone to invasions by plants, insects, and microbes, including insect vectors and the viruses they transmit (Bar-Massada et al., 2014). As invasions continue, and as we strive to protect these remnant lands, it is essential to understand the role of plant viruses in these ecosystems.

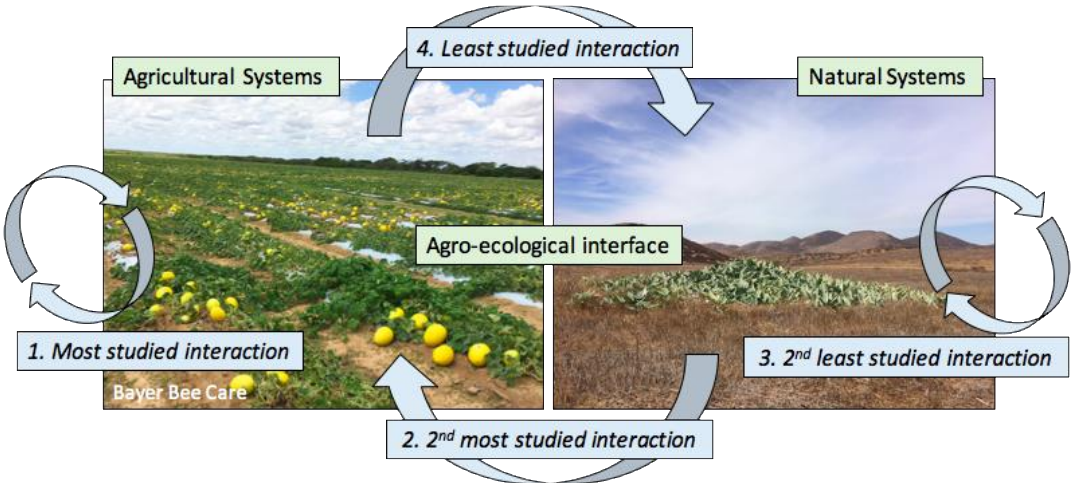

Figure 0.1. *The agro-ecological interface is a continuum of native and managed landscapes. The interactions of vectors and pathogens moving into wild systems from agriculture is the least studied interaction. This is followed by studies focused solely on wild systems. The most studied interaction is movement of vectors and pathogens within*

cultivated systems, followed by the movement of pathogens and vectors into crops from wild reservoirs. Figure adapted from Alexander et al., 2014.

Plant virus ecology: an emerging discipline aided by new technology

Most research on plant viruses in crops, and most ecology subfields (including restoration, conservation, and ecosystem ecology) have become established without consideration of virus communities (Malmstrom et al., 2011). However, studying virus dynamics within and among wild plants is important for basic research on pathogen-host-vector coevolution, disease emergence, and virus-vector interactions. There has been rapid progress in the field of plant virus ecology after the founding of an international research consortium known as the Plant Virus Ecology Network (PVEN), which was established in 2007 by Ulrich Melcher and Carolyn Malmstrom with funding from the United States National Science Foundation (NSF) (Malmstrom et al., 2011). Through this consortium, ecologists and plant virologists were brought together in meetings and workshops, facilitating collaboration. Fortuitously, this occurred alongside the emergence and broader usage of high throughput sequencing (HTS) technologies.

Previously, identifying viruses in wild plants has been difficult because symptoms are often not apparent or drastically different from those in domesticated crops. This makes it difficult to explore virus communities in wild plants using traditional diagnostic approaches created for use in crops. Some of these traditional approaches include enzyme-linked immunosorbent assays (ELISA) and reverse-transcription PCR (RT-PCR). The major drawback of these methods is that they require *a priori* knowledge of viruses present. ELISA is also less sensitive than RT-PCR, and both methods are

challenged by chemical and physical defenses in plants that inhibit accurate detection (Lacroix et al., 2016). Advances in HTS have alleviated challenges in diagnosis and discovery of viruses in hosts without symptoms. Common approaches for whole virome characterization using HTS includes 454 pyrosequencing (Bernardo et al., 2018; Roossinck et al., 2010) and sequencing using Illumina platforms (MiSeq, NextSeq, and HiSeq) (Shates et al., 2019), both of which provide sequences regardless of whether virus identities are known in advance. This allows researchers to find viruses that they may not expect, or those that may not be classified at all, ranging from viruses that are common pathogens of crops to those that are complete unknowns with minimal sequence similarity to any virus (Stobbe & Roossinck, 2014). Overall, the growing literature on plant virus ecology shows that viruses - known and unknown species - are present and abundant in wild systems, but relatively few studies have moved beyond virus discovery (Bernardo et al., 2018; Prendeville et al., 2012; Seabloom et al., 2010; Shates et al., 2019; Susi et al., 2019).

Purpose of this work

Even though progress is being made in the field of plant virus ecology, major research gaps remain. These are highlighted below along with hypotheses to be tested in this dissertation research.

Knowledge Gap 1: Deficits in research on perennial plants and dicot plants.

There is a bias for researching plants that are grasses (plant family Poaceae) and annuals (Shates et al., 2019). Although annuals are indeed prevalent, perennial plants are more common in wild plant communities (Alexander et al., 2014). We may expect different

impacts of viruses on perennials, and different host-virus interactions, due to perennials having a multi-season lifestyle. We also may expect different patterns to emerge from research on dicots because they are susceptible to other pests and pathogens relative to monocots. Most of the research on monocots is focused on one virus group: the Barley/Cereal yellow dwarf viruses (B/CYDVs), which do not infect dicots (BYDV Data Sheet, CABI). Furthermore, we may expect different patterns from perennial plants because they are exposed to vectors and pathogens across multiple years (Agrios, 1997).

Hypotheses

A. Drought-tolerant perennial, dicot hosts are frequently infected by at least one virus.

B. Infections in perennials are likely to be asymptomatically unapparent

C. Perennials are likely to host multiple infections by distinct virus species.

Knowledge Gap 2: Weak understanding of the factors shaping virus community composition within and between wild hosts.

Most research on virus ecology in wild plants addresses the "virus discovery" portion but rarely follows up with analysis of virus communities beyond species richness and diversity. Plants can be infected with multiple viruses, and the resulting assembly in a host is considered one community (virome) (Maclot et al., 2020). To advance virus ecology, it is important to incorporate methods from community ecology to understand how viruses are distributed in plant communities and the implications of these distributions for plant health and virus evolution. So far, we know that perennial grasses are more likely to be infected by B/CYDV than annual grasses, which shows that there are likely differences for plant life histories (Ingwell et al., 2017). The presence of

agriculture/cultivation of crops also influences the viruses found and their abundance (Bernardo et al., 2018; Rodelo-Urrego et al., 2013), and host diversity also plays a role in structuring virus communities (Sallinen et al., 2020). However, as this field is still new, there is much to be learned by studying new plant host systems in new regions.

Hypotheses

A. Perennial plants in more fragmented (or development-adjacent) habitats in southern California may accumulate infections by a wider diversity of crop-associated viruses.

B. Target perennial plants acquire and retain viruses over multiple seasons.

Knowledge Gap 3: Lack of manipulative studies quantifying the impacts of crop viruses on wild plant health.

Until recently, wild plants were believed to be asymptomatic reservoirs without notable consequences of infection. However, the few studies assessing virus impacts on wild plants show that infections by crop-associated viruses can be detrimental, even if symptoms do not mimic those in crop hosts (Malmstrom & Alexander, 2016). As with other areas of virus ecology, this evidence is heavily skewed toward B/CYDV-grassland systems. Additionally, traditional crop-relevant metrics of performance (like fruit yield) don't always translate to wild plants, especially for perennials that invest in structures for long term survival like larger roots and may not reproduce every year. One example is the seemingly asymptomatic CYDV infections in perennial switchgrass, which lead to reduced root production and length (Malmstrom et al., 2017). BYDV infected switchgrass in field trials experienced up to a 30% reduction in fitness despite appearing

asymptomatic, but this reduction was only apparent when quantified over multiple seasons of infection (Alexander et al., 2017). Determining the effects of crop viruses in key perennials is important for understanding how plant communities may be structured by virus infections, especially newly introduced crop-associated viruses arriving from agricultural areas.

Hypotheses

A. Infections by crop-associated viruses negatively affect drought-tolerant perennials in ways that are relevant for survival or fitness.

Approach

To test these hypotheses, I will use a study system consisting of three wild dicot species that are drought-adapted key perennials in the southwestern chaparral habitat type (*Datura wrightii* Regel, *Cucurbita foetidissima* Kunth, *and Cucurbita palmata* Wats.). Additionally, two of the target hosts, *C. foetidissima* and *C. palmata* are relatives of crop plants; both are the ancestral perennial form of domesticated crops in the genus *Cucurbita* (e.g., squash, pumpkins, gourds) (Tyack et al., 2020). These plants are sources of resting space, fruit, pollen, and nectar for various generalist insects such as ants, darkling and leaf beetles, and sweat bees as well as for rodents (personal observation). *Datura wrightii* are frequently fed on by hawkmoths adults and larvae, possibly due to their abundance (Alarcón et al., 2008). *Cucurbita spp.* are also visited by co-evolved, specialist pollinators important for squash and melon cultivation: squash bees *Xenoglossa spp.* and *Peponapis spp*. (Hurd et al., 1971; Castellanos-Morales et al., 2018). These plants are important to many organisms, but there is not sufficient evidence for them to

be considered keystone. However, based on their multifaceted roles in the ecosystem, from here on I call these hosts "key perennials." In my first chapter, I review the gaps in the literature in plant virus ecology to quantify the proportions of research to date on dicots, perennials, and studies that go beyond virome characterization. I also use untargeted sequencing to determine the identities of viruses infecting the target wild hosts in two reserves with significant edge habitat relative to interior habitat. With these data, I performed a network analysis to determine the relative importance of different viruses in target hosts, and phylogenetics to assess the relationship of known and novel viruses to other virus taxa. In my second chapter, I expanded virome sequencing to a third reserve in southern California, Anza-Borrego State Park, which, unlike the reserves targeted for chapter 1 sequencing efforts, features extensive interior habitat. Using virome data from all three reserves, as well as targeted sampling of select virus species, I performed community ecology analyses to determine the drivers of virus community structure and prevalence of two highly prevalent virus species (*Cucurbit aphid-borne yellows virus* (CABYV) and *Cucurbit yellow stunting disorder virus* (CYSDV)). In my third chapter, I analyzed the genetic relationships of the most prevalent virus across all three reserves: CABYV. I used 17 full genomes assembled from the sequencing in the first two chapters, combined with all GenBank contributed sequences submitted through 2019, to perform phylogenetics, recombination detection, and phylogeographic analyses and determine relationships and origins of what is likely a species complex. In my fourth chapter, I conduct greenhouse experiments to determine the impacts of CABYV on plant health. I manipulated infection status for the two cucurbits to evaluate symptom expression and

infection impacts as proxies for plant fitness between CABYV-infected and sham-inoculated plants. I paired these experiments with field studies where I use regression to assess trends between plant fitness proxies and symptom severity under natural conditions. Finally, I report preliminary experiments using the electrical penetration graph technique to compare *A. gossypii* (aphid vector) feeding behavior on sham and virus-inoculated *C. foetidissima* and *C. palmata*.

References

Agrios, G. N. (1997). *Plant Pathology, Fourth Edition (4th Edition)*. Harcourt Academic Press.

Alarcón, R., Davidowitz, G., & Bronstein, J. L. (2008). Nectar usage in a southern Arizona hawkmoth community. *Ecological Entomology*, *33*(4), 503–509.

Alexander, H. M., Bruns, E., Schebor, H., & Malmstrom, C. M. (2017). Crop-associated virus infection in a native perennial grass: reduction in plant fitness and dynamic patterns of virus detection. *The Journal of Ecology*, *105*(4), 1021–1031.

Alexander, H. M., Mauck, K. E., Whitfield, A. E., Garrett, K. A., & Malmstrom, C. M. (2014). Plant-virus interactions and the agro-ecological interface. *European Journal of Plant Pathology / European Foundation for Plant Pathology*, *138*(3), 529–547.

Amayo, R., Arinaitwe, A. B., Mukasa, S. B., Tusiime, G., Kyamanywa, S., Rubaihayo, P. R., & Edema, R. (2012). Prevalence of viruses infecting cowpea in Uganda and their molecular detection. *African Journal of Biotechnology*, *11*(77), 14132–14139.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, *19*(10), 535–544.

Angelella, G. M., Holland, J. D., & Kaplan, I. (2016). Landscape composition is more important than local management for crop virus–insect vector interactions. *Agriculture, Ecosystems & Environment*, *233*, 253–261.

Bar-Massada, A., Radeloff, V. C., & Stewart, S. I. (2014). Biotic and abiotic effects of human settlements in the wildland–urban interface. *Bioscience*, *64*(5), 429–437.

Bernardo, P., Charles-Dominique, T., Barakat, M., Ortet, P., Fernandez, E., Filloux, D., Hartnady, P., Rebelo, T. A., Cousins, S. R., Mesleard, F., Cohez, D., Yavercovski, N., Varsani, A., Harkins, G. W., Peterschmitt, M., Malmstrom, C. M., Martin, D. P., & Roumagnac, P. (2018). Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *The ISME Journal*, *12*(1), 173–184.

Broadbent, L. (1957). Insecticidal control of the spread of plant viruses. *Annual Review of Entomology*, *2*(1), 339–354.

Byrne, D. N. (1999). Migration and dispersal by the sweet potato whitefly, *Bemisia tabaci*. *Agricultural and Forest Meteorology*, *97*(4), 309–316.

Castellanos-Morales, G., Paredes-Torres, L. M., Gámez, N., Hernández-Rosales, H. S., Sánchez-de la Vega, G., Barrera-Redondo, J., Aguirre-Planter, E., Vázquez-Lobo, A., Montes-Hernández, S., Lira-Saade, R., & Eguiarte, L. E. (2018). Historical biogeography and phylogeny of Cucurbita: Insights from ancestral area reconstruction and niche evolution. *Molecular Phylogenetics and Evolution*, *128*, 38–54.

Ellis, E. C., & Ramankutty, N. (2008). Putting people in the map: anthropogenic biomes of the world. *Frontiers in Ecology and the Environment*, *6*(8), 439–447.

Ertunc, F. (2020). Emerging Plant Viruses. In *Emerging and Reemerging Viral Pathogens* (pp. 1041–1062). Academic Press.

Gibbs, A. J., Ohshima, K., Phillips, M. J., & Gibbs, M. J. (2008). The prehistory of potyviruses: their initial radiation was during the dawn of agriculture. *PloS One*, *3*(6), e2523.

Hogenhout, S. A., Ammar, E.-D., Whitfield, A. E., & Redinbaugh, M. G. (2008). Insect vector interactions with persistently transmitted viruses. *Annual Review of Phytopathology*, *46*, 327–359.

Hurd, P. D., Jr, Linsley, E. G., & Whitaker, T. W. (1971). SQUASH AND GOURD BEES (PEPONAPIS, XENOGLOSSA) AND THE ORIGIN OF THE CULTIVATED CUCURBITA. *Evolution; International Journal of Organic Evolution*, *25*(1), 218–234.

Hussein, M. Y., & Samad, N. A. (1993). Intercropping chili with maize or brinjal to suppress populations of *Aphis gossypii* Glov., and transmission of chili viruses. *International Journal of Pest Management*, *39*(2), 216–222.

Ingwell, L. L., Lacroix, C., Rhoades, P. R., Karasev, A. V., & Bosque-Pérez, N. A. (2017). Agroecological and environmental factors influence *Barley yellow dwarf viruses* in grasslands in the US Pacific Northwest. *Virus Research*, *241*, 185–195.

Jones, R. A. C. (2020). Disease pandemics and major epidemics arising from new encounters between indigenous viruses and introduced crops. *Viruses*, *12*(12). https://doi.org/10.3390/v12121388

Kennedy, J. S. (1950). Aphid migration and the spread of plant viruses. *Nature*, *165*(4208), 1024–1025.

Lacroix, C., Renner, K., Cole, E., Seabloom, E. W., Borer, E. T., & Malmstrom, C. M. (2016). Methodological guidelines for accurate detection of viruses in wild plant species. *Applied and Environmental Microbiology*, *82*(6), 1966–1975.

Lecoq, H., Bourdin, D., Wipf-Scheibel, C., Bon, M., Lot, H., Lemaire, O., & Herrbach, E. (1992). A new yellowing disease of cucurbits caused by a luteovirus, cucurbit aphid-borne yellows virus. *Plant Pathology*, *41*(6), 749–761.

Loxdale, H. D., Hardie, J., Halbert, S., Foottit, R., Kidd, N. A. C., & Carter, C. I. (1993). The relative importance of short- and long-range movement of flying aphids. *Biological Reviews of the Cambridge Philosophical Society*, *68*(2), 291–311.

Maclot, F., Candresse, T., Filloux, D., Malmstrom, C. M., Roumagnac, P., van der Vlugt, R., & Massart, S. (2020). Illuminating an ecological blackbox: using high throughput sequencing to characterize the plant virome across scales. *Frontiers in Microbiology*, *11*, 2575.

Malmstrom, C. M., & Alexander, H. M. (2016). Effects of crop viruses on wild plants. *Current Opinion in Virology*, *19*, 30–36.

Malmstrom, C. M., Bigelow, P., Trębicki, P., Busch, A. K., Friel, C., Cole, E., Abdel-Azim, H., Phillippo, C., & Alexander, H. M. (2017). 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*, 172–184.

Malmstrom, C. M., Melcher, U., & Bosque-Pérez, N. A. (2011). The expanding field of plant virus ecology: historical foundations, knowledge gaps, and research directions. *Virus Research*, *159*(2), 84–94.

Martín, B., Fereres, A., Tjallingii, W. F., & Collar, J. L. (1997). Intracellular ingestion and salivation by aphids may cause the acquisition and inoculation of non-persistently transmitted plant viruses. In *Journal of General Virology* (Vol. 78, Issue 10, pp. 2701–2705). https://doi.org/10.1099/0022-1317-78-10-2701

Mauck, K., Bosque-Pérez, N. A., Eigenbrode, S. D., De Moraes, C. M., & Mescher, M. C. (2012). Transmission mechanisms shape pathogen effects on host-vector interactions: evidence from plant viruses. *Functional Ecology*, *26*(5), 1162–1175.

Mauck, K. E., & Chesnais, Q. (2020). A synthesis of virus-vector associations reveals important deficiencies in studies on host and vector manipulation by plant viruses. *Virus Research*, *285*, 197957.

Ng, J. C. K., & Falk, B. W. (2006). Virus-vector interactions mediating nonpersistent and semipersistent transmission of plant viruses. *Annual Review of Phytopathology*, *44*, 183–212.

Oerke, E.-C. (2006). Crop losses to pests. *The Journal of Agricultural Science*, *144*(1), 31–43.

Pagán, I., González-Jara, P., Moreno-Letelier, A., Rodelo-Urrego, M., Fraile, A., Piñero, D., & García-Arenal, F. (2012). Effect of biodiversity changes in disease risk: exploring disease emergence in a plant-virus system. *PLoS Pathogens*, *8*(7), e1002796.

Perring, T. M., Gruenhagen, N. M., & Farrar, C. A. (1999). Management of plant viral diseases through chemical control of insect vectors. *Annual Review of Entomology*, *44*, 457–481.

Prado, E., & Tjallingii, W. F. (1994). Aphid activities during sieve element punctures. *Entomologia Experimentalis et Applicata*. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1570-

Prendeville, H. R., Ye, X., Morris, T. J., & Pilson, D. (2012). Virus infections in wild plant populations are both frequent and often unapparent. *American Journal of Botany*, *99*(6), 1033–1042.

Ramankutty, N., Evan, A. T., Monfreda, C., & Foley, J. A. (2008). Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, *22*(1). https://doi.org/10.1029/2007gb002952

Rodelo-Urrego, M., Pagán, I., González-Jara, P., Betancourt, M., Moreno-Letelier, A., Ayllón, M. A., Fraile, A., Piñero, D., & García-Arenal, F. (2013). Landscape heterogeneity shapes host-parasite interactions and results in apparent plant-virus codivergence. *Molecular Ecology*, *22*(8), 2325–2340.

Roistacher, C. N., Nauer, E. M., Kishaba, A., & Calavan, E. C. (1980). Transmission of citrus tristeza virus by *Aphis gossypii* reflecting changes in virus transmissibility in California. *International Organization of Citrus Virologists Conference Proceedings (1957-2010)*, *8*. https://escholarship.org/content/qt3w7436tq/qt3w7436tq.pdf

Rojas, M. R., & Gilbertson, R. L. (2008). Emerging Plant Viruses: a Diversity of Mechanisms and Opportunities. In M. J. Roossinck (Ed.), *Plant Virus Evolution* (pp. 27–51). Springer Berlin Heidelberg.

Roossinck, M. J., Saha, P., Wiley, G. B., Quan, J., White, J. D., Lai, H., Chavarría, F., Shen, G., & Roe, B. A. (2010). Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Molecular Ecology*, *19 Suppl 1*, 81–88.

Sallinen, S., Norberg, A., Susi, H., & Laine, A.-L. (2020). Intraspecific host variation plays a key role in virus community assembly. *Nature Communications*, *11*(1), 5610.

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution*, 1.

Seabloom, E. W., Borer, E. T., Mitchell, C. E., & Power, A. G. (2010). Viral diversity and prevalence gradients in North American Pacific Coast grasslands. *Ecology*, *91*(3), 721–732.

Shates, T. M., Sun, P., Malmstrom, C. M., Dominguez, C., & Mauck, K. E. (2019). Addressing research needs in the field of plant virus ecology by defining knowledge gaps and developing wild dicot study systems. *Frontiers in Microbiology*, *9*, 3305.

Stansly, P. A., & Natwick, E. T. (2010). Integrated systems for managing *Bemisia tabaci* in protected and open field agriculture. In P. A. Stansly & S. E. Naranjo (Eds.), *Bemisia: Bionomics and Management of a Global Pest* (pp. 467–497). Springer Netherlands.

Stobbe, A. H., & Roossinck, M. J. (2014). Plant virus metagenomics: what we know and why we need to know more. *Frontiers in Plant Science*, *5*, 150.

Susi, H., Filloux, D., Frilander, M. J., Roumagnac, P., & Laine, A.-L. (2019). Diverse and variable virus communities in wild plant populations revealed by metagenomic tools. *PeerJ*, *7*, e6140.

Tyack, N., Dempewolf, H., & Khoury, C. K. (2020). The potential of payment for ecosystem services for crop wild relative conservation. *Plants*, *9*(10). https://doi.org/10.3390/plants9101305

Varma, A., Mandal, B., & Singh, M. K. (2011). Global emergence and spread of whitefly (*Bemisia tabaci*) transmitted geminiviruses. In W. M. O. Thompson (Ed.), *The Whitefly, Bemisia tabaci (Homoptera: Aleyrodidae) Interaction with Geminivirus-Infected Host Plants: Bemisia tabaci, Host Plants and Geminiviruses* (pp. 205–292). Springer Netherlands.

Vincent, S. J., Coutts, B. A., & Jones, R. A. C. (2014). Effects of introduced and indigenous viruses on native plants: exploring their disease-causing potential at the agro-ecological interface. *PloS One*, *9*(3), e91224.

Wintermantel, W. M., Gilbertson, R. L., Natwick, E. T., & McCreight, J. D. (2017). Emergence and epidemiology of *Cucurbit yellow stunting disorder virus* in the American Desert Southwest, and development of host plant resistance in melon. *Virus Research*, *241*, 213–219.

Zhao, Z.-H., Sandhu, H. S., Gao, F., & He, D.-H. (2015). Shifts in natural enemy assemblages resulting from landscape simplification account for biocontrol loss in wheat fields. *Ecological Research*, *30*(3), 493–498.

Zhu, M., Radcliffe, E. B., Ragsdale, D. W., MacRae, I. V., & Seeley, M. W. (2006). Low-level jet streams associated with spring aphid migration and current season spread of potato viruses in the U.S. northern Great Plains. *Agricultural and Forest Meteorology*, *138*(1), 192–202.

Chapter One: Addressing research needs in the field of plant virus ecology by defining knowledge gaps and developing wild dicot study systems

Abstract

Viruses are ubiquitous within all habitats that support cellular life and represent the most important emerging infectious diseases of plants. Despite this, it is only recently that we have begun to describe the ecological roles of plant viruses in unmanaged systems and the influence of ecosystem properties on virus evolution. We now know that wild plants frequently harbor infections by diverse virus species, but much remains to be learned about how viruses influence host traits and how hosts influence virus evolution and vector interactions. To identify knowledge gaps and suggest avenues for alleviating research deficits, we performed a quantitative synthesis of a representative sample of virus ecology literature, developed criteria for expanding the suite of pathosystems serving as models, and applied these criteria through a case study. We found significant gaps in the types of ecological systems studied, which merit more attention. In particular, there is a strong need for a greater diversity of logistically tractable, wild dicot perennial study systems suitable for experimental manipulations of infection status. Based on criteria developed from our quantitative synthesis, we evaluated three California native dicot perennials typically found in Mediterranean-climate plant communities as candidate models: *Cucurbita foetidissima* (buffalo gourd), *Cucurbita palmata* (coyote gourd), and *Datura wrightii* (sacred thorn-apple). We used Illumina sequencing and network analyses to characterize viromes and viral links among species, using samples taken from multiple individuals at two different reserves. We also compared our Illumina workflow with

targeted RT-PCR detection assays of varying costs. To make this process accessible to ecologists looking to incorporate virology into existing studies, we describe our approach in detail and discuss advantages and challenges of different protocols. We also provide a bioinformatics workflow based on open-access tools with graphical user interfaces. Our study provides evidence that dicot perennials in xeric habitats support multiple, asymptomatic infections by viruses known to be pathogenic in related crop hosts. Quantifying the impacts of these interactions on plant performance and virus epidemiology in our logistically tractable host systems will provide fundamental information about plant virus ecology outside of crop environments.

Introduction

The integration of plant virology and ecology is recent, stimulated by the discovery that viruses are important but largely undescribed and understudied components of the ecology of wild plants. Since the inception of this field, ecologists and plant virologists have worked to illuminate two issues: (i) the ecological roles of plant viruses and vectors in unmanaged ecosystems, and (ii) the influence of ecosystem properties on the distribution and evolution of plant viruses and their vectors (Malmstrom et al., 2011). Thanks to these efforts, we now know that wild plants frequently harbor infections by a wide diversity of plant viruses, which almost certainly influence host traits in ways not previously quantified or considered by the field of plant ecology (Malmstrom and Alexander, 2016). Evidence further suggests that the characteristics of plant viruses can be influenced by selection in natural communities and across ag–wild and urban–wild interfaces (Bernardo et al., 2018; Rodelo-Urrego et al., 2015). But much remains to be learned about how ecosystem properties influence virus evolution and vector interactions.

The need for research in this area has never been greater. Among emerging plant diseases, viruses are the most common causal agents (Anderson et al., 2004). Natural areas have undergone significant fragmentation as agricultural production and urbanization has increased (Ellis et al., 2010; Fischer and Lindenmayer, 2007). A key issue is that the expanded array of agro-ecological boundaries has created new opportunities for microbes associated with wild and cultivated plants to interact and to move between host types. Thus, it is critical to ask to what degree these interactions will

influence microbial evolution and the probability that some microbes will emerge as novel pathogens (Alexander et al., 2013; Roossinck and García-Arenal, 2015).

At present, systems-level understanding of plant virus dynamics across agro-ecological interfaces is hampered by a dearth of information from wild communities, including those that border agricultural and urban land cover. Historically, most monitoring of plant virus dynamics has focused on cropping systems (Alexander et al., 2013; Roossinck and García-Arenal, 2015; Wren et al., 2006), although new geo-metagenomics approaches have begun to reveal virus distributions within non-cultivated systems (Bernardo et al., 2018). Here, we demonstrate that additional knowledge constraints arise because understanding of plant–virus dynamics is skewed towards annual host systems, which are more dominant in agriculture than in nature. While viruses have been tracked in some perennial crops such as stone fruit and grapes (e.g., plum pox virus), similar studies are rare for long-lived perennials in nature, even though perennials are dominant in most natural communities and may harbor diverse infections (Alexander et al., 2017).

To advance understanding in these critical research areas, it is essential that virus interactions within non-cultivated perennial systems be investigated. Perennials have potential to serve as modulators of viral virulence by exerting selection for longer-term host survival, or, in other cases, as sites for virus recombination through accumulation of co-infections or even as reservoirs for novel disease agents. In turn, non-cultivated perennial systems may be perturbed and affected by microbial movement from agricultural crops, which may also serve as reservoirs and exert selection pressures on

viruses in ways that are not fully recognized. From an ecological perspective, viruses from crop environments may modify the expression of host plant functional traits in ways that alter stress tolerance, fitness (Alexander et al., 2017), and competitive interactions (Malmstrom et al., 2017), with significant implications for conservation of threatened native plant communities.

To advance work in this area, several hurdles need to be surmounted. One need is to simplify methods of virus detection and identify likely pitfalls so that researchers without backgrounds in molecular virology can properly examine the viromes of wild plant study systems. A second need is to identify promising wild plant–virus systems that can serve as useful models for short- and long-term manipulative work that goes beyond surveys and correlational studies. For example, we found that a large fraction of plant virus ecology work involving factor manipulation has been conducted on Poaceae hosts in a single ecosystem type: invaded California grasslands (see results section). There is thus pressing need to identify new model systems involving dicot plant hosts beyond *Arabidopsis*. In the study presented here, we conducted a review of plant virus ecology literature to quantify resource gaps in the field, then use these results to define criteria for selecting the most useful new systems for future work. We then present a case study of three potential model hosts to demonstrate how investigators with limited budgets might initially evaluate the viromes of candidate wild plant systems, and tractability of various detection assays, with the aim of incorporating virus dynamics into ecological studies.

The case study presented here focuses on fragmented semi-arid plant communities in southern California (USA), which is a Mediterranean-climate biodiversity hotspot

(Myers et al., 2000). To explore the virus dynamics among key species within these communities, we used both Illumina sequencing and custom detection assays to study the viromes of three native dicot perennials (*Cucurbita foetidissima* and *Cucurbita palmata* [Cucurbitaceae], and *Datura wrightii* [Solanaceae]) that are summer-growing species within fragmented preserves of mixed grassland, xeric sage scrub and chaparral communities in the Southwestern U.S. (Fig. 1A). Grassland, chaparral, and sage scrub habitats are the core native plant communities in Southern California and are dominated by perennials. Communities with these features, such as the Tallgrass Prairie Preserve in Oklahoma, have contributed significantly to our emerging understanding of virus biodiversity in wild plants (Min et al., 2012; Muthukumar et al., 2009; Scheets et al., 2011; Thapa et al., 2012, 2015). Efforts are underway to conserve and maintain these habitats, which persist as fragmented preserves between agricultural and urban environments.

For our study, we targeted two agriculture and urban-adjacent preserves in Riverside County, California, USA, that feature these Mediterranean climate communities (Fig. 1.1). Mediterranean-climate systems are characterized by cool winter growing seasons with some rain, followed by summer drought. Consistent with these seasonal patterns, the grassland, sage scrub, and chaparral environments within our selected sites include both winter- and summer-growing perennial species. Winter-growing species, while still adapted for xeric conditions, do not thrive in extreme heat and typically senesce by March or April each year. Summer species are drought- and heat-adapted plants capable of resisting severe abiotic stress. Due to these traits, the

summer-growing species targeted in our study (Fig. 1.1) are some of the few plants in leaf and bloom at the same time that insect vector populations peak in agricultural crops (Chu et al., 2001) and are likely exposed to both crop-associated and wild insect-transmitted viruses. Our summer-growing perennials also have congeneric and confamilial crop hosts grown in adjacent agricultural areas (squash and melons [Cucurbitaceae] and potatoes and peppers [Solanaceae]) and therefore may be susceptible to vectors and viral pathogens originating from monocultures of these crops. Thus, we hypothesized that we would frequently find infections by crop-associated viruses in our target species. We also hypothesized that these hosts could be important sources of novel emerging pathogens as infecting viruses evolve and adapt to exploit host tissues over multiple seasons. Therefore, we expected that if crop-associated viruses were present, some of them might have diverged substantially from those circulating in crop habitats while engaging in a persistent association with the same host individual across seasons. Finally, we expected to find novel plant viruses and viruses infecting plant fungal associates (endophytes, fungal pathogens), as these have been discovered in nearly all virus biodiversity studies to date that include steps for enrichment of virus nucleic acids followed by next-generation sequencing (Roossinck, 2014; Stobbe and Roossinck, 2014).

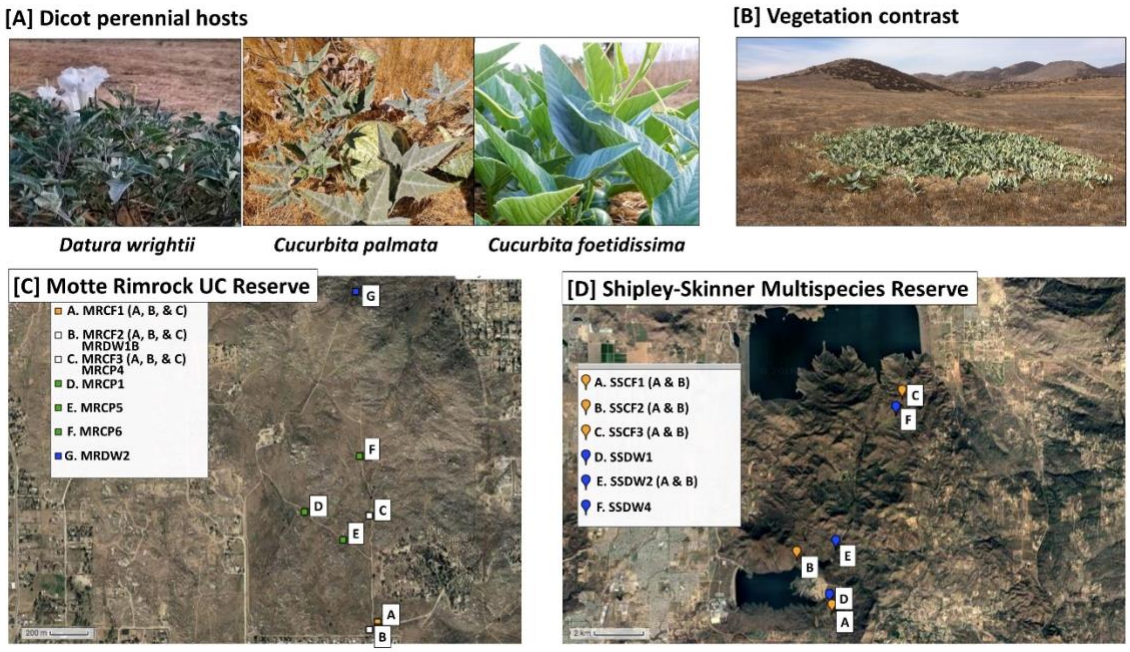

Figure 1.1. Study system and sampling locations. *[A] Images of asymptomatic host plants sampled from the two reserve sites. [B] Typical growth habit of* C. foetidissima *during arid summer months. Green foliage contrasts strongly with soil and dry, senescent vegetation. [C] Satellite image map of populations sampled from the University of California Motte Rimrock Reserve. [D] Satellite image map of populations sampled from Shipley-Skinner Multispecies Reserve. Maps created by GPSVisualizer online.*

Methods

Identifying knowledge gaps and resource needs in the field of virus ecology

Rapid progress in the field of plant virus ecology was ignited by the founding of an international research consortium known as the Plant Virus Ecology Network (PVEN), which was established in 2007 by Ulrich Melcher and Carolyn Malmstrom with funding from the U.S. National Science Foundation (NSF). PVEN supported international workshops that were instrumental in connecting molecular virologists, epidemiologists, ecologists, entomologists, and other researchers with diverse skill sets relevant to plant virus ecology. NSF records indicate that PVEN stimulated the production of over 75

peer-reviewed publications between 2007 and 2011. To analyze the impact of PVEN on the field of virus ecology, and identify knowledge and resource gaps, we quantified the content of papers arising directly from PVEN support (**precursor papers**) and subsequent papers that cited these precursor papers (**product papers**). Precursor papers consisted of all peer-reviewed publications listed in the NSF reporting website (NSF Award Abstract #0639139, 2011), excluding conference abstracts. We categorized each precursor paper as contributing to 1–3 possible research areas from among the following: *Virus evolution, Virus discovery, Virus description, Virus effects on host traits, Virus-vector interactions, Epidemiology, Methods, Host resistance, Review papers, Environmental virology, and Theoretical* (mathematical modeling). A full description of the categories, as well as a complete list and categorization of the precursor papers, can be found in Supplementary Data Sheet 1.1. We designated six of these categories (*Virus evolution, Virus discovery, Virus effects on host traits, Virus-vector interactions, Epidemiology, and Environmental virology*) as core research categories of virus ecology based on previously defined criteria (Malmstrom et al., 2011).

For precursor papers within each of these six categories, we assessed the ecological focus of research content by categorizing the domestication status of plant hosts studied (crop, wild, or crop + wild), the plant life history strategy (annual, perennial, or both annuals and perennials), and the characteristics of the viruses studied (crop-associated [pathogenic viruses prevalent in crop systems], wild [only known to associate with non-crop hosts], or both crop-associated and wild). To evaluate impact, we used Google Scholar to obtain a list of all subsequent papers (here, 'product papers')

citing each precursor paper since its publication. Each product paper was evaluated to determine if the content matched one or more of the six core virus ecology research areas. The ecological focus of product papers was determined in the same manner as for precursor papers. We assessed both literature sets for biases in favor of certain host and virus characteristics to identify knowledge gaps and resource needs within the field of virus ecology. We also evaluated characteristics of the most useful wild plant study systems in research categories that typically include experimental manipulations of infection status and other environmental conditions. This included all studies employing wild or crop/model + wild plants within the areas of *Environmental virology, Virus effects on host traits,* and *Virus-vector interactions*. Finally, we used this analysis, and the broader analysis of precursor and product papers, to generate a set of criteria for identifying and developing new study systems.

Host selection, site selection, and plant sampling

Based on the quantitative literature synthesis, we evaluated California native plant species typically found in fragmented natural communities as candidate model species for basic and applied virus ecology research. We focused on perennial dicot species and identified three promising candidates: *Cucurbita foetidissima* (buffalo gourd), *Cucurbita palmata* (coyote gourd), and *Datura wrightii* (sacred thorn-apple). We characterized individual viromes of representatives of these three species collected from two sites in southern California near the Los Angeles Urban Region in Riverside County: The Motte Rimrock Reserve (University of California Natural Reserve System) and the Shipley-Skinner Multispecies Reserve (Metropolitan Water District of Southern California).

These reserves serve as prime examples of fragmented, remnant natural communities that lie adjacent to urban and agricultural systems (Fig. 1.1). Vegetation at the Motte Rimrock Reserve (289 ha) is primarily inland coastal sage scrub. Mean annual precipitation is 33 cm; common soil types include coarse sandy loam, rocky sandy loam and sandy loam (Motte Rimrock Reserve Statistics, 2010; USDA Natural Resources Conservation Service, 2017). Dominant plant families include the Lamiaceae, Asteraceae and Poaceae, with Cucurbitaceae, Solanaceae, and Cactaceae occurring in discrete, often large, populations (Motte Rimrock Species List, 2009). Vegetation at the larger Shipley-Skinner Reserve (1012 ha) includes native inland coastal sage scrub and chaparral communities, as well as grasslands dominated by non-native species. Mean annual precipitation ranges from 25.4–40.64 cm throughout the reserve; common soil types include sandy loam, rocky loam, cobbly clay and rocky sandy loam (personal communication with rangers Robert Williams and Tom Ash). Dominant plant families include the Lamiaceae (particularly *Salvia mellifera* and *Salvia apiana*), Asteraceae (*Artemisia* spp.), Polygonaceae (*Eriogonum* spp.), and Poaceae (*Bromus* spp., *Avena* spp., *Stipa pulcra*, *Stipa cernua*, and *Melica imperfecta)* (personal communication with rangers Robert Williams and Tom Ash). As at Motte Rimrock, Cucurbitaceae and Solanaceae species are interspersed among the dominant vegetation. During the hot, dry summer months at both reserves (June–September), our target species are particularly green and apparent, as most other vegetation is senescent or dormant (Fig. 1B).

All samples were collected without regard to expression of infection symptoms between 7:00-10:00 am over the course of several days in August 2017. At Motte

Rimrock, we collected tissue from three *C. foetidissima* individuals from each of three populations 200–500 meters apart, from four *C. palmata* individuals also about 200–500 meters apart, and from two *D. wrightii* individuals at different ends of the reserve, 1600 m distant from each other. At Shipley-Skinner Reserve, we found individuals of only two of the three species (*C. foetidissima* and *D. wrightii*). These plants were distributed unevenly and so were sampled along a north-south transect through the reserve. In total, we collected tissue from 25 plants: 9 *C. foetidissima,* 4 *C. palmata*, and 2 *D. wrightii* from Motte Rimrock Reserve, and 6 *C. foetidissima* and 4 *D. wrightii* from the Shipley-Skinner Reserve (Fig. 1). We collected 10 g of leaf and stem tissue from each plant by inverting a clean plastic ziplock bag over the tissue and removing it from the plant, then sealing the bag. Bagged tissue was placed on dry ice and transported back to the laboratory within 2-3 hours. Samples were carefully partitioned into 50-mL RNase-free Falcon tubes and stored at -80℃ until processing.

Extraction and next-generation sequencing of viral nucleic acids

In our case study, a primary goal was to illustrate how useful information could be gained by cost-effective approaches suitable for initial data collection. We emphasized next-generation sequencing (NGS) methods because of their capability to detect even previously unknown viruses. Numerous NGS platforms exist and new models become available quickly. We focused on a less expensive bench-top model available at a broad range of institutions: The Illumina NextSeq 500. To further reduce costs, we used the Mid Output flow cell (v2) with a paired-end 75 bp read-length. This configuration can

produce up to 260M paired-end reads. For a subset of viruses, we additionally used virus-specific RT-PCR and Sanger sequencing to characterize virus infections.

Sequencing plant-infecting RNA viruses typically requires extraction of total RNA (host plus associated microbes), small RNA (21–24 nt), or double-stranded RNA (dsRNA) (Kesanakurti et al., 2016). There are advantages and disadvantages to each approach, which vary with application. If total RNA is extracted without further purification, sequencing resources will be spent on host nucleic acids, which may decrease the probability of detecting low-abundance viruses. This disadvantage can be mitigated by first conducting virion-associated nucleic acid (VANA) semi-purification (Bernardo et al., 2018) or depleting ribosomal RNA fractions. In some cases, it is effective to target small RNAs, which are often abundant in infected plants following degradation of virus genomes by conserved, anti-viral silencing mechanisms (Pooggin, 2018). However, this approach varies in feasibility with plant silencing efficiency, coverage of viral genomes may be uneven, and it is difficult to detect viral sequence variants due to the very short reads (21–24 nt). We chose to focus on extractions of dsRNA. Few plant RNA viruses have dsRNA genomes, but dsRNA forms of viral single stranded (ssRNA) genomes are present in all infected plants during the replication cycle. High molecular weight dsRNA can also be found in tissue infected with some DNA viruses (Weber et al., 2006), suggesting that dsRNA extracts can detect RNA viruses along with at least some DNA viruses, although this has not been rigorously tested. Next-generation sequencing of dsRNA extracts is also well suited to low-abundance viruses

because dsRNA extracted from infected plants is primarily of virus origin, not host origin.

The tissue of wild plants typically is tougher than that of crop and model species and contains chemically diverse metabolite profiles, which can interfere with nucleic acid extraction and downstream molecular analysis (Lacroix et al., 2016). To evaluate the effectiveness of genomic analyses in this study, we tracked virus recovery from sampled plants by spiking each sample with a leaf punch of bell pepper cv. California Wonder containing *Bell pepper endornavirus* (BPEV, *Endornaviridae*) prior to dsRNA extraction (Kesanakurti et al., 2016). One leaf punch was added to 4 g of sample tissue contained in a 50mL RNase-free falcon tube with 12 3-mm diameter stainless steel grinding balls. Spiked samples were homogenized using a Geno/Grinder® (SPEX SamplePrep) by shaking at 1700 rpm for 30 min (necessary for complete homogenization of 4 g of tissue). Samples were kept frozen during grinding by pre-cooling solid metal racks containing the tubes in liquid nitrogen for five minutes. We isolated double stranded RNA from 4 g of plant tissue using a low-cost extraction procedure involving binding and enrichment of dsRNA using Sigmacell cellulose type 101 powder (Kesanakurti et al., 2016; Tzanetakis and Martin, 2008). New tubes, grinding balls, and barrier tips were used throughout the extraction procedure and downstream library preparation work to prevent cross-contamination. Purified dsRNA was dissolved in 25 µl of nuclease-free water.

To construct libraries (one per sample), we denatured an aliquot of 5 µl of dsRNA solution from each extraction by incubating at 99 ℃ for 5 min, then placed the solution on ice and used the NEBNext® Ultra™ II Directional RNA Library Prep Kit for

Illumina® by following the manufacturer's protocol. The average insert size is 250–300 bp. We barcoded 27 libraries, which were pooled and sequenced in two separate runs (one with 12 libraries, one with 15, including two pepper controls) on an Illumina NextSeq 500 instrument using Mid Output v2 kit. We elected to use 75 bp paired end reads for cost savings. Sequencing, adapter removal and quality checks were performed by the UC Riverside genomics core facility. Sequences that could not be assigned an index based on the adapter sequence were discarded (see raw read statistics in Supplementary Data Sheet 1.2; adapter-removed raw reads are available in the sequence read archive of GenBank with accession number SRP149013). Reads were assembled into contigs with the Trinity assembly and analyzed with "NCBI BLAST+ blastn" using a Galaxy-based workflow (Supplementary Data Sheet 1.2). This workflow does not require any specialized coding skills, Linux knowledge, or data storage expenses, and thus may be more accessible to researchers new to this work. For our main analysis, we chose to subtract reads that matched the available genomes of closely related host species (*Cucurbita maxima* and *Solanum lycopersicum*). For a subset of samples, we compared this approach to a direct analysis without host genome filtering. Details of both approaches are available in Supplementary Data Sheet 1.2.

Virus identification and phylogenetic analysis

To identify known and novel virus sequences in our samples, we performed BlastN on the Illumina-based contigs using the "NCBI BLAST+ blastn" tool in Galaxy (Supplementary Data Sheet 1.2). Assemblies were compared to a previously prepared database of plant and fungal viruses

(https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239&host=plants). We chose this database since it includes complete nucleotide sequences for plant viruses, fungal viruses, and viruses infecting both plants and fungi. To minimize false or misleading hits, we did not include partial nucleotide or amino acid virus sequences available in GenBank.

The contigs assembled included nearly complete genome sequences from four of six detected crop-associated viruses (*Tomato chlorosis virus,* [ToCV] *Closteroviridae; Cucurbit aphid-borne yellows virus,* [CABYV] *Luteoviridae*; *Zucchini yellow mosaic virus,* [ZYMV] *Potyviridae*; and *Papaya ringspot virus*, [PRSV] *Potyviridae*). We performed further phylogenetic analyses to identify possible origins and presence of host-specific variants. All sequences used for alignments and phylogenetics are from the Illumina sequencing of dsRNA libraries except for ToCV, for which we amplified and Sanger-sequenced the complete coat protein coding sequence from total RNA cDNA (for sample numbers SSCF1A, SSDW4 and SSDW2B). This was necessary because our Illumina sequencing did not yield complete genomes for all samples. All other sequences used are from previously reported crop-associated virus isolates (Supplementary Data Sheet 1.3). For CABYV, PRSV, and ZYMV we selected published virus sequences based on their geographic origins (preference for adjacent geographic regions within the US, followed by representative sequences from regions outside the US where the virus is established). For ToCV, which was not previously known to occur in California, we constructed phylogenies using coding sequences for coat protein, minor coat protein, and

RNA-dependent RNA polymerase to determine the possible origin of the ToCV detected in *D. wrightii* and *C. foetidissima*.

For analysis, we aligned coat protein coding sequences from contigs identified as representing crop-associated viruses using the freely available ClustalW tool in Mega X software (Kumar et al., 2018). Sequences in alignments were uniformly trimmed to ORF regions of interest for which we had obtained corresponding sequences from the field. The trimmed alignments were manually checked to verify the absence of mis-alignments, using translated amino acid sequences as a guide, and then exported to fasta format. We then conducted maximum likelihood phylogenetic analysis on each alignment in CLC Main Workbench (version 8), a low-cost, user-friendly software platform available from Qiagen. For nucleotide analyses, we first used the CLC model testing tool to determine the best substitution model for each analysis, which varied with the alignment. This tool evaluates the suitability of substitution models using four different statistical analyses, including hierarchical likelihood ratio test (hLRT), Bayesian information criterion (BIC), minimum theoretical information criterion (AIC), and the corrected minimum theoretical information criterion (AICc). The base tree for model evaluation was created by the neighbor-joining method. We selected the model with the greatest support across all four statistical evaluations. Selected models included General Time Reversible (GTR) + G + T (Yang, 1994), Kimura 80 + G + T (Kimura, 1980), and HKY + T (Hasegawa et al., 1985), where G indicates rate variation with estimated gamma distribution parameter and T indicates topology variation (see Figure legends for specific selections). For amino acid analysis, we used the Whelan and Goldman (WAG) model of protein substitution

(Whelan and Goldman, 2001). For the maximum likelihood phylogenetic analysis, the starting trees were created with the neighbor-joining method. One hundred bootstrap re-samples were performed for each analysis. The results are presented in unrooted radial format, drawn to scale, with branch lengths measured in the expected number of substitutions per site. Only nodes with bootstrap values >= 70 are shown; nodes with bootstrap values < 70 are collapsed and not shown.

For putative novel viruses (PV1-4) we selected and aligned the longest coding sequence with a minimum of 400 bp matching the RNA-dependent RNA polymerase gene (RdRp) with corresponding RdRp coding sequences from viruses having a BlastN hit in the plant virus database. Alignments were performed using the ClustalW tool in Mega X software (Kumar et al., 2018). We chose RdRp amino acid sequences for alignments because this genome region received the most BlastN hits across samples. For *partitivirus* 1 (PV1), we excluded GenBank sequences NC003885.1 and NC004018.1 in the alignment as they were highly dissimilar to other sequences returned in the BlastN analysis. Alignment trimming, manual checks for mis-alignments, and maximum likelihood phylogenetic analysis were performed as described above for known viruses. For all analyses, we used the WAG model of protein substitution.

Comparison of Illumina sequencing to RT-PCR on dsRNA or total RNA extracts

For a subset of detected viruses (*Cucurbit aphid-borne yellows virus* [CABYV, *Luteoviridae*] and *Tomato chlorosis virus* [ToCV, *Closteroviridae*]), we performed targeted detection assays (RT-PCR) using dsRNA and total RNA isolations. Tissue used for both isolations experienced only one freeze-thaw cycle. We isolated total RNA from

~100mg leaf tissue using a phenol-based method (RiboZol™ ,VWR) and dissolved extracted RNA in 100 µl of nuclease-free water. Total RNA concentrations were 100–400 ng/µl; we repeated extractions until we obtained RNA from each sample with an A260/A280 ratio of greater than 1.5. Prior work with wild plant species demonstrated that low A260/A280 ratios are associated with inhibition of reverse transcription and PCR reactions (Lacroix et al., 2016). Therefore, for reverse transcription, we used an enzyme (Superscript IV, Invitrogen) that is robust against most plant-derived inhibitors according to the manufacturer's product description, and added RNase inhibitor (Ribolock, Thermo Fisher) to prevent RNA degradation. 5 µl total RNA was reverse transcribed, resulting in 20 µl of cDNA. We used both random hexamers (low cost, one RT reaction for all viruses) and gene-specific primers (high cost, unique RT reactions for each virus) to determine how these two approaches perform. When dsRNA extracts were processed, we first denatured the material by incubating 2 µl at 99 ℃ for 5 min. We then placed the solution on ice and performed reverse transcription as for total RNA.

Subsequent PCR reactions (20 µl reaction volume) consisted of 1 µl of template cDNA, 4 µl of 5X HF buffer (ThermoFisher Scientific), 1 µl of each 10µM primer, 2 µl of dNTP mix (2mM each) and 0.2 µl of Phusion DNA polymerase (ThermoFisher Scientific). Primers used for PCR detection were previously developed by others. One PCR protocol was used for both CABYV and ToCV viruses: 98℃ for 3 min for initial denaturation, followed by 40 cycles of 98℃ for 10 s, 60℃ for 30 s, 72℃ for 1 min, then 72℃ for 10 min. PCR was conducted in a Bio-Rad T100™ Thermal Cycler. For Sanger sequencing

of PCR products, amplicons were purified with Mag-Bind® Total Pure NGS magnetic beads (Omega Bio-tek). Sanger sequencing was performed by Retrogen Inc. (San Diego, CA) and primer sequences were removed from contigs prior to analysis.

Results

Knowledge gaps and resource needs in the field of plant virus ecology

Among precursor papers (Fig. 1.2A), studies focusing exclusively or partially on wild plant hosts have equal or greater representation relative to studies focusing exclusively on crop hosts within the areas of *Environmental virology*, *Epidemiology*, *Virus effects on host traits*, *Virus evolution*, and *Virus discovery*, while studies focusing exclusively or partially on wild hosts are somewhat underrepresented within the area of *Virus–vector interactions* (5/11 studies) (Fig. 1.2B). Studies focusing exclusively on annual hosts are overrepresented in the areas of *Environmental virology* (5/7 studies, with none on perennials alone), *Virus effects on host traits* (7/9 studies), and *Virus–vector interactions* (7/11 studies) (Fig. 1.2C). *Virus discovery* is the only research area in which studies focusing exclusively on perennial hosts (5/9) outnumber those focusing exclusively on annual hosts (1/9) (Fig. 1.2C). Nearly all studies, regardless of research area, focus on crop-associated viruses, except for the *Virus discovery* category, in which 7/9 studies included wild viruses that are not causative agents of disease in crops (Fig. 1.2D).

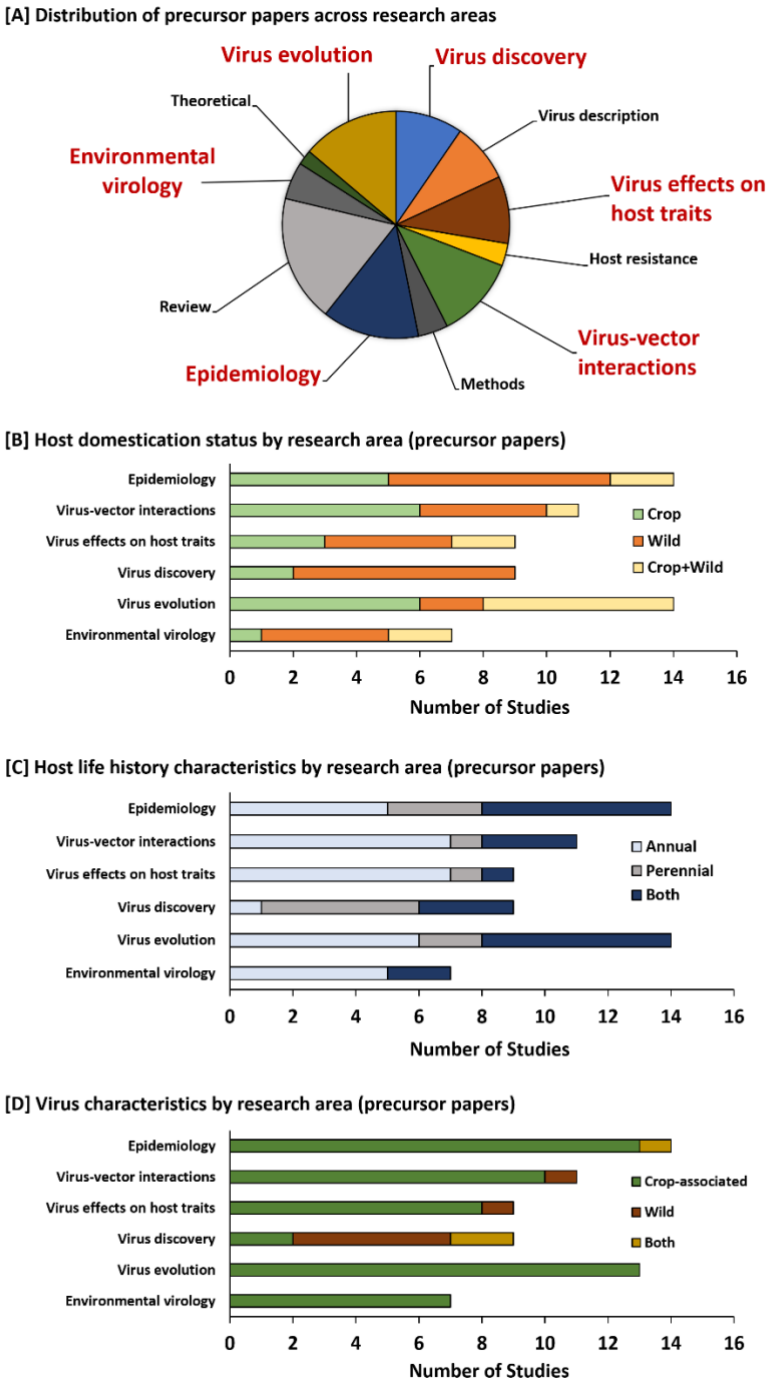

Figure 1.2. Precursor paper research topic categories. *The pie graph [A] illustrates the proportion of precursor papers (direct outputs of PVEN) falling into each category. The six categories most relevant to the field of virus ecology are emphasized in large, red*

text. Graphs [B-D] summarize the number of studies utilizing hosts and viruses with specific categorical characteristics.

Among product papers falling into one or more of the six core virus ecology research areas (Fig. 1.3A), there is a clear shift toward greater proportional representation of crop and model host plants relative to wild hosts (Fig. 1.3B). When each exclusive category (crop, model, or wild) is considered alongside the combined category, studies employing wild hosts only enjoy equal representation with crops or models in the *Virus discovery* and *Epidemiology* categories – both of which are heavily based on surveys rather than manipulative studies. In contrast, studies on wild hosts are lacking in areas that typically employ experimental manipulations (*Virus effects on host traits*, *Virus-vector interactions*, and *Virus evolution*) (Fig. 1.3B). When host domestication status is examined along an axis of plant family rather than research category, additional biases are apparent (Fig. 1.3C). Nearly half of all studies focusing exclusively on wild plants, and about one third of studies that include both wild and cultivated plants, were performed with monocots from a single plant family (Poaceae) (Fig. 1.3C). Solanaceous hosts are also popular, but only a quarter of studies on Solanaceae include wild species (Fig. 1.3C). About 80% of studies focused exclusively on crop-associated viruses, with only the *Virus discovery* research area (which is primarily based on survey work) including wild viruses in about 60% of studies (Fig. 1.3D).

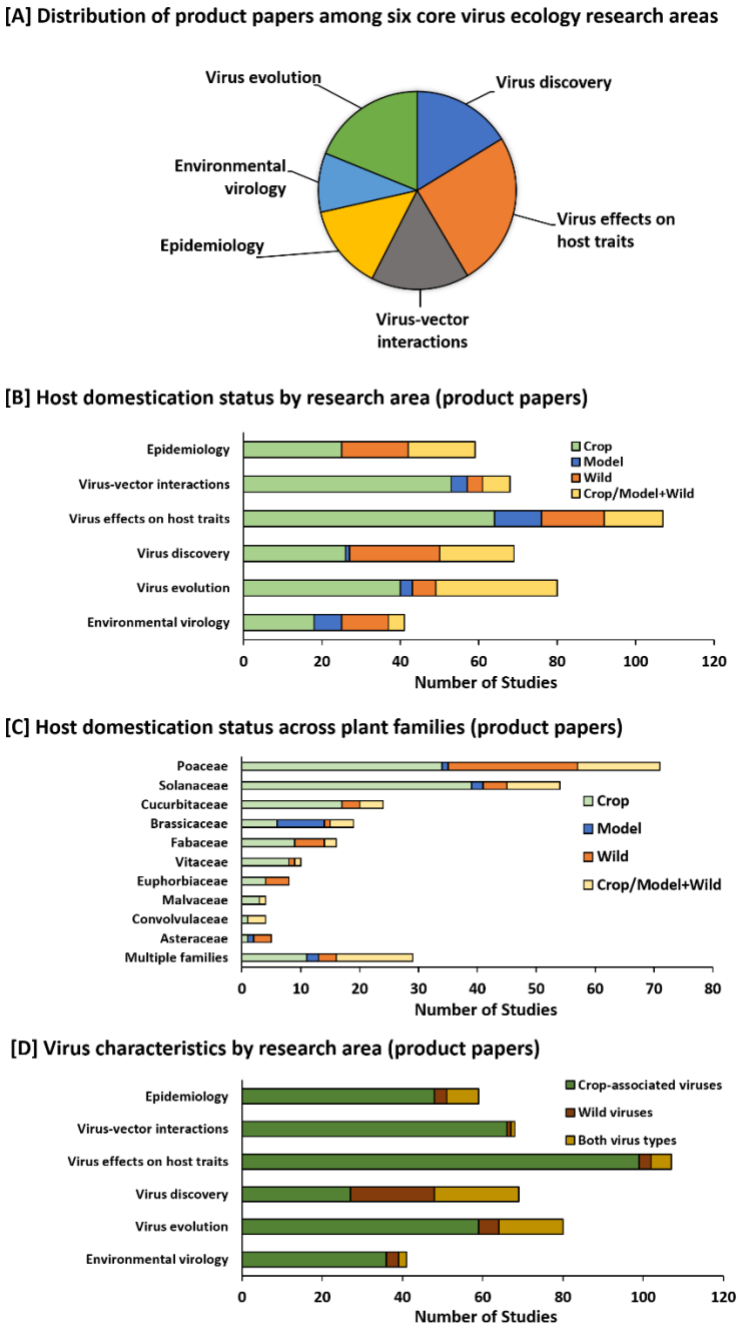

Figure 1.3. Product paper analysis. *The pie graph [A] illustrates the proportion of product papers (those citing direct outputs of PVEN) falling into each core plant virus ecology research areas. Graph [B] illustrates the domestication status of hosts studied in product papers organized by research area. Graph [C] illustrates domestication status of hosts studied in product papers organized by plant family and pooled across research areas. Families depicted are those represented by >3 studies. Families not depicted*

include Actinideaceae (one study, crop), Amaranthaceae (one study, crop), Amaryllidaceae (two studies, crop/model+wild, one study, wild), Apocynaceae (one study, wild), Orchidaceae (one study, crop/model+wild, one study, wild), Rutaceae (three studies, one each in crop, wild, and crop/model+wild), and Zingiberaceae (one study, crop). Graph [D] illustrates virus characteristics organized by research area.

Among product papers within the areas of *Virus effects on host traits*, *Virus–vector interactions*, and/or *Environmental virology* that included wild hosts, we detected overrepresentation of monocots (specifically, Poaceae), overrepresentation of annuals (most wild plant communities are dominated by perennials), a preference for uncomplicated seed-based propagation methods, and a tendency to prefer exotic/invasive hosts with weedy characteristics (Fig. 1.4). Despite the fact that only a fraction of wild plants are ancestral to crops, more than one third of wild hosts studied are in the same genus as a crop species, and about one quarter to one third are themselves cultivated in some context or selected for use in landscape manipulation (restoration or generation of fodder for grazing by ruminants) (Fig. 1.4).

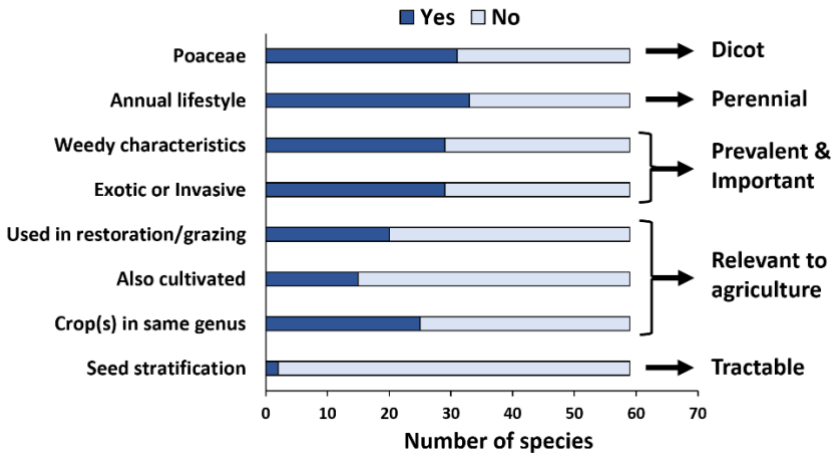

Figure 1.4. Characteristics of wild hosts used as focal research organisms in experimental virus ecology studies. *In total, 59 wild plant species were studied across all product papers in three research areas that frequently involve factorial or manipulative*

empirical studies (Virus-vector interactions, Virus effects on host traits, and Environmental virology). Arrows and text to the right indicate desirable characteristics of potential model hosts for experimental approaches.

Virus detections in model wild dicot hosts: Illumina sequencing

We detected six crop-associated viruses via dsRNA extraction and Illumina sequencing followed by host genome filtering (Fig. 1.5A). For 65% of these detections we recovered greater than 90% of the virus genome. All of the detected crop-associated viruses are known to be present in California, with the exception of ToCV, a whitefly-transmitted virus that is established in the Southeastern U.S. (Wintermantel and Wisler, 2006). We confirmed infections by ToCV in at least two *D. wrightii* plants via RT-PCR and sanger sequencing, all of them from the Shipley-Skinner reserve site. We also identified one instance of ToCV infection in a cucurbit host (*C. foetidissima*) growing in the same site as *D. wrightii* plants infected with ToCV (Fig. 1.1 & 1.5). Although it is primarily a pathogen of cucurbits, we also detected CABYV infections in *D. wrightii* (Fig. 1.5A). In addition to crop-associated viruses, we detected several viruses most closely related to members of the *Partitiviridae*, which we partitioned into four groups (PV1–PV4) based on sequence similarity among isolates (Fig. 1.5A). Most of the plants we sampled harbored viruses in groups PV1–3, while viruses in group PV4 were less frequently detected. A network analysis based on Illumina detections (Fig. 1.6) illustrates the number and strength of interactions among viral players across host species and sampling locations. This analysis suggests that one crop-associated virus, CABYV, is strongly associated with both cucurbit species, with potential transmission to the solanaceous host. The other crop-associated viruses are peripheral, appearing only sporadically within

select populations. The putative *Partitiviridae* within each of the PV1-PV3 groups also have strong connections with all hosts, and are present in both sampling sites, while viruses in the PV4 group occupy an intermediate position between the center and the periphery.

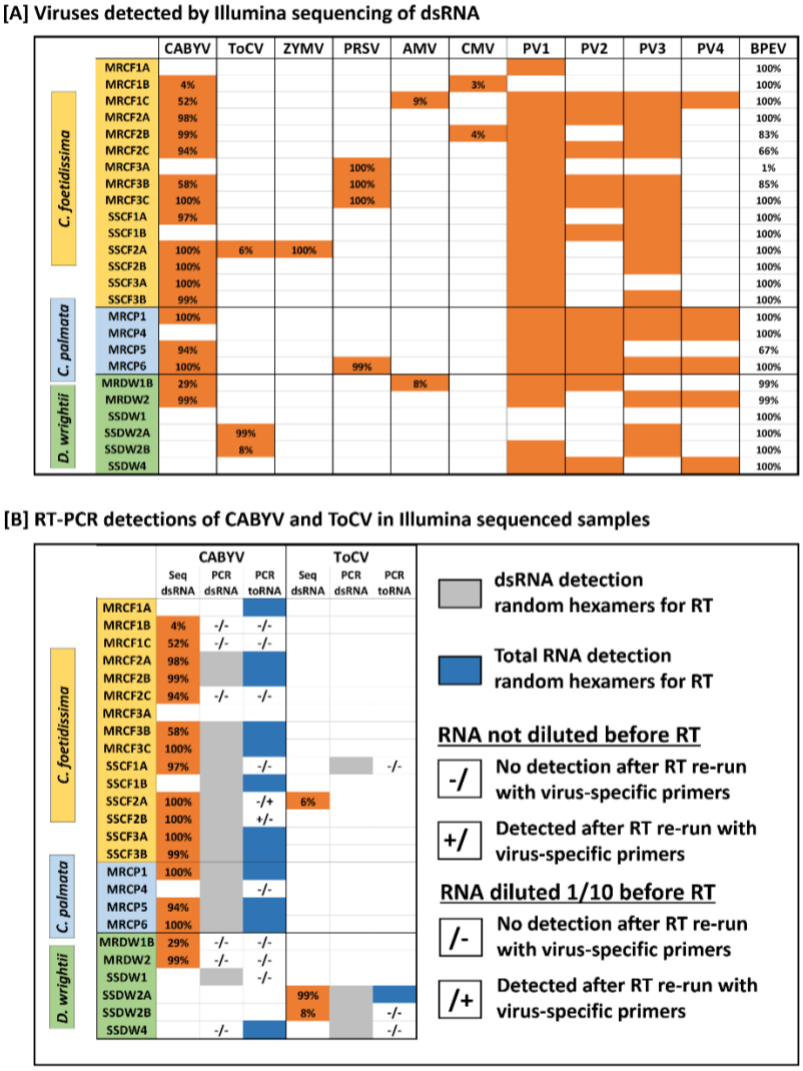

Figure 1.5. *Viromes of individual plants [A] and summary of detections by different diagnostic methods for select crop-associated viruses [B]. Sample codes within each host plant section indicate species (CF, CP, or DW), site of collection (MR = Motte Rimrock, SS = Shipley-Skinner), numbers (1, 2, and 3) indicate populations within a site, and letters (A, B, and C) indicate individuals within each population. Viruses are as follows, from left to right:* Cucurbit aphid-borne yellows virus, Tomato chlorosis virus, Zucchini

yellow mosaic virus, Papaya ringspot virus, Alfalfa mosaic virus, and Cucumber mosaic virus. *PV1-4 refer to novel viruses related to members of the family* Partitiviridae. *The last column shows the coverage of the internal control (*Bell pepper endornavirus*) for each sample. Results are from assemblies constructed after subtraction of host genomes (see Methods). We found no effect of host subtraction on the identity and number of viruses detected (see Discussion).*

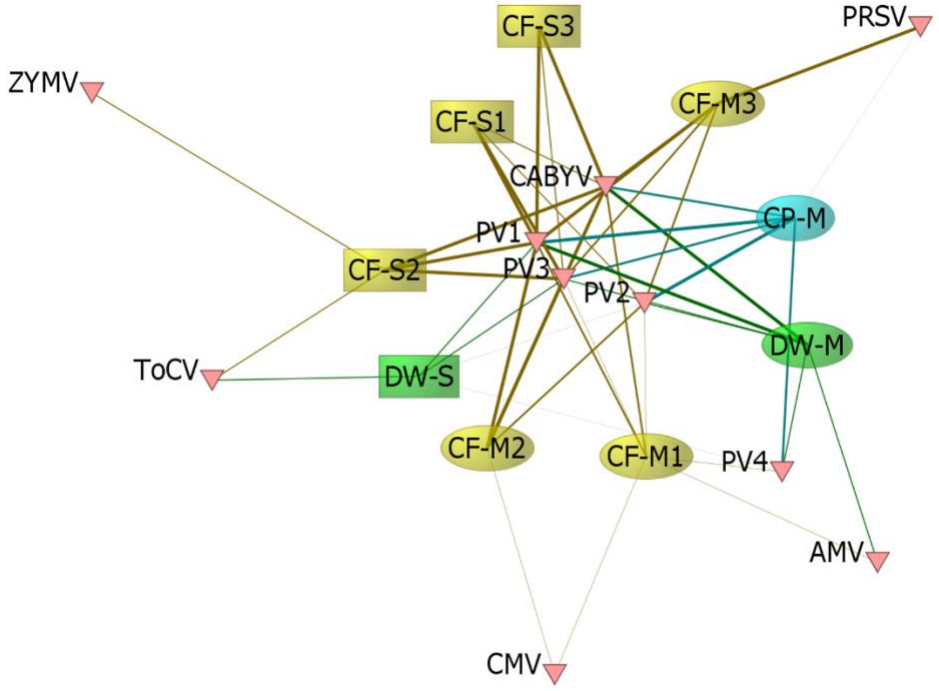

Figure 1.6. Viral links among hosts. *We constructed a matrix of the nine population groups sampled x ten viral groups detected, where values for each population within the matrix represented the proportion of individuals sampled in which Illumina sequencing detected the corresponding virus group. We visualized the network structure with the network analysis software ORA-Lite version 11 (Carley et al., 2018), where the network was sized by betweenness centrality. The width of the edge between a given population and virus group represents the proportion of individuals with Illumina-detected infection. Sample codes indicate species (CF, CP, or DW), site of collection (MR = Motte Rimrock, SS = Shipley-Skinner), and population number within a site (1, 2, and 3).*

Detection efficiency for crop-associated viruses

The internal control of pepper tissue infected with BPEV worked well as a metric to assess dsRNA recovery and sequencing depth. In most samples, we recovered the entire

genome of BPEV. But in one case, we recovered only 1% of the genome (sample MRCF3A – Fig. 1.5A). This sample was left within our data set to illustrate how recovery of the BPEV genome tracks with detection of other viruses. Sample MRCF3A had fewer co-infections than most other samples based on Illumina sequencing. RT-PCR was performed to amplify a 381 bp fragment of BPEV in the pepper-spiked dsRNA extraction from this individual. The fragment was detected, suggesting that dsRNA isolation was successful, but possibly compromised in a way that lowered abundance of virus dsRNA below the detection threshold for sequencing. Use of BPEV as an internal control enables attribution of the source of this variation to processing errors rather than biological factors.

Wild plants often contain inhibitors of RT and PCR reactions (Lacroix et al., 2016). If our candidates are particularly recalcitrant for targeted detection assays, this will alter their suitability as models for virus ecology work and/or require additional method optimization. For CABYV and ToCV, we found good congruence between detection via Illumina sequencing and detection via RT-PCR using random hexamers during the RT reactions for both dsRNA and total RNA (Fig. 1.5B). Additionally, two samples tested positive for ToCV and five samples tested positive for CABYV without testing positive via Illumina sequencing (Fig. 1.5B). For a subset of samples with discrepancies among the three detection methods, we repeated the protocol with virus-specific primers, and with or without RNA dilution (Fig. 1.5B) (Lacroix et al., 2016). Detection was only improved for two samples: SSCF2A, where CABYV was detected in the 1/10 dilution

with specific primers, and SSCF2B, where CABYV was detected in the undiluted sample with specific primers.

Phylogenetic analyses of crop-associated and novel viruses

To determine relationships among virus isolates and explore geographic origins, we performed a phylogenetic analysis for all crop-associated viruses for which we recovered all or most of the genome (ToCV, CABYV, PRSV and ZYMV). ToCV is not known to be present in California and our detection represents the first instance of this virus being present in this state. To determine possible ToCV origins, we used sequences on Genbank for the major and minor coat proteins and the RNA-dependent RNA-polymerase. All trees suggest that our ToCV sequences are most similar to those of isolates collected within the continental U.S. (Florida and Colorado), and to those of isolates collected in China (Fig. 1.7). There is no evidence that the ToCV infecting *C. foetidissima* is distinct from ToCV infecting *D. wrightii* hosts in the same reserve site (Fig. 1.7A). Phylogenetic analysis of the CABYV coat protein sequence suggests the presence of two genotypes (Fig. 1.8). The first CABYV genotype is most similar to isolates collected in Asia (China, Japan, Korea) and the midwestern U.S. (Oklahoma). This genotype infects both *C. palmata* and the solanaceous host, *D. wrightii*. The second CABYV genotype groups with isolates from Europe, and we found no evidence of this genotype infecting *D. wrightii*. Phylogenetic analysis of our PRSV isolates suggests a single genotype present in the Motte Rimrock Reserve site (Fig. 1.9A). Origins are not clear, as our PRSV group is equivalently similar to isolates from the U.S., Mexico, and Australia (Fig. 1.9A). In contrast, based on the coat protein sequence, our ZYMV isolate

is likely of U.S. origin, with the most similar sequence being from another California isolate collected from an agricultural field (Fig. 1.9B).

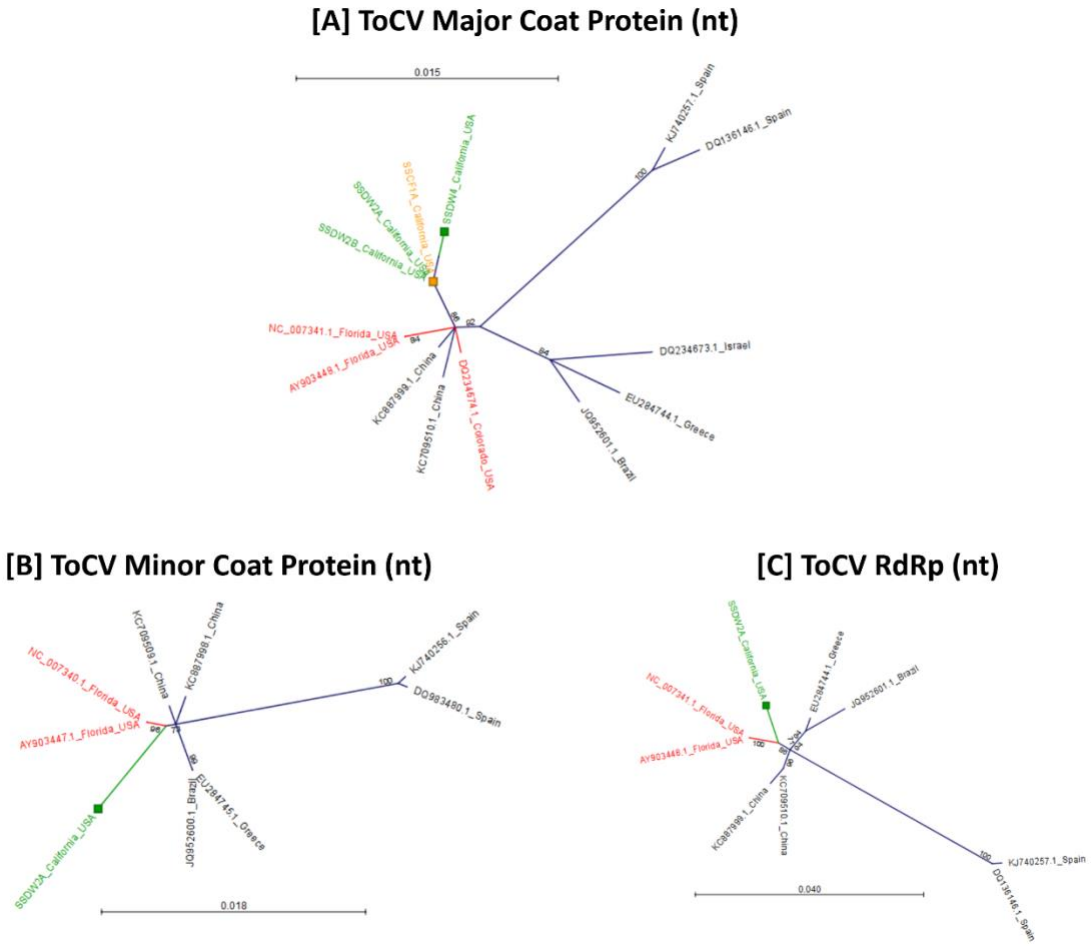

Figure 1.7. Radial, unrooted trees illustrating results of the maximum likelihood phylogenetic analyses of ToCV nucleotide sequences. *For the major coat protein [A], the HKY + T model was used. For the minor coat protein [B], the GTR + G + T model was used. For the RNA-dependent RNA polymerase [C], the HKY + T model was used. Color coding has been applied to aid interpretation of sequence origins (green =* D. wrightii, *orange =* C. foetidissima, *red = other sequences from U.S. isolates). One hundred bootstrap re-samples were performed for each analysis. Branch lengths are measured in the expected number of substitutions per site. Only nodes with bootstrap values >= 70 are shown; nodes with bootstrap values < 70 are collapsed and not shown.*

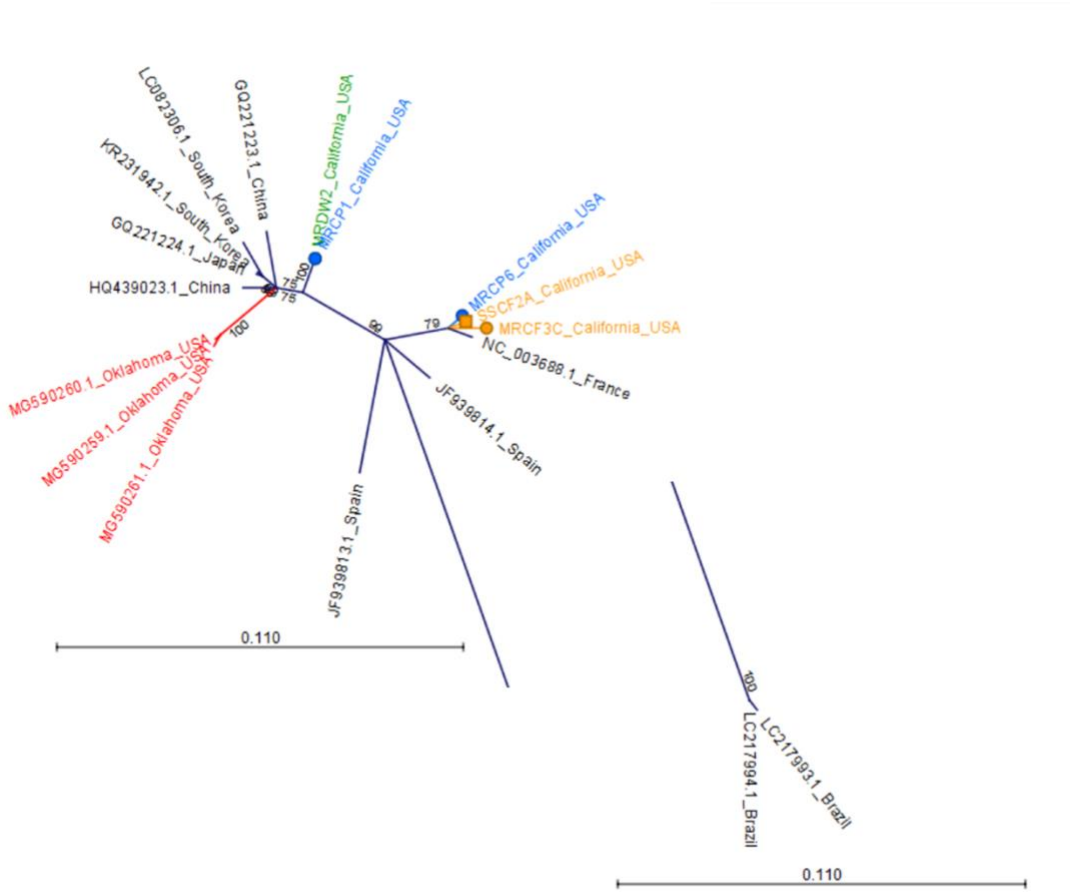

Figure 1.8. Radial, unrooted trees illustrating results of the maximum likelihood phylogenetic analyses of CABYV coat protein nucleotide sequences. *The GTR + G + T model was used for tree construction. Two Brazilian CABYV isolates exhibit extreme divergence from all other sequences and the branch containing these isolates has been trimmed to improve readability of the tree. Color coding has been applied to aid interpretation of sequence origins (green* = D. wrightii, *orange* = C. foetidissima, *blue* = C. palmata, *red = other sequences from U.S. isolates). One hundred bootstrap re-samples were performed for each analysis. Branch lengths are measured in the expected number of substitutions per site. Only nodes with bootstrap values >= 70 are shown; nodes with bootstrap values < 70 are collapsed and not shown.*

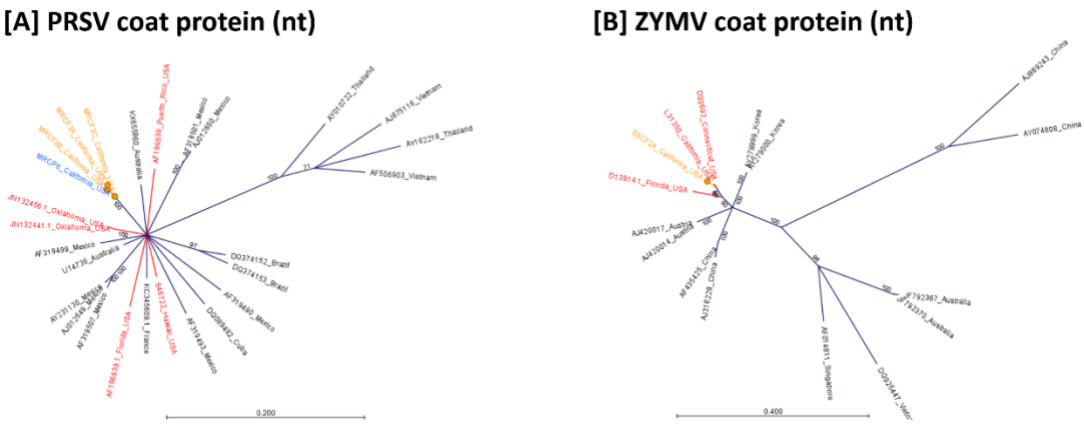

Figure 1.9. Radial, unrooted trees illustrating results of the maximum likelihood phylogenetic analyses of* Potyvirus *nucleotide sequences. *For PRSV [A], and ZYMV [B] the GTR + G + T model was used. Color coding has been applied to aid interpretation of sequence origins (orange =* C. foetidissima*, blue =* C. palmata*, red = other sequences from U.S. isolates). One hundred bootstrap re-samples were performed for each analysis. Branch lengths are measured in the expected number of substitutions per site. Only nodes with bootstrap values >= 70 are shown; nodes with bootstrap values < 70 are collapsed and not shown.*

Phylogenetic analysis of RdRp amino acid sequences from detected *Partitiviridae*-like viruses revealed two possible virus lifestyles, most of which are consistent with patterns of detection among our three hosts and sampling sites. The PV1 group appears to contain sequences from a single virus species, as the sequences detected shared 100% identity. The PV1 group drew BlastN hits only for fungus-infecting *Partitiviridae* in the genus *Gammapartitivirus*, which does not include any plant-infecting viruses (Fig. 1.10A). A fungus-infecting lifestyle is consistent with the finding that PV1 is frequently present in both cucurbits and *D. wrightii* hosts across two sites (Fig. 1.5A). Plants have numerous fungal endophytes and other associates. PV1 viruses could infect a generalist fungus associate of all three hosts, but this requires additional validation. The PV2 group contains at least two virus genotypes, which possibly

represent two species, as the distance between the two clades is greater than that between different characterized viruses included in the same tree (Fig. 1.10B). Members of each group are present in both *C. foetidissima* and *C. palmata* tissues. The PV2 group members are most similar to plant-infecting members of the genus *Alphapartitivirus,* with fungus-infecting members of this genus being slightly more distant in branch length. However, PV2 sequences were also detected in *D. wrightii,* although a large enough sequence was not available from *D. wrightii* to include in this analysis (Fig. 1.5A). Based on this, a mycovirus lifestyle should not be ruled out. Analysis of PV3 sequences also suggests that this group may contain multiple species with different lifestyles. *C. palmata* derived sequences group closely with plant-infecting *betapartitiviruses*, while the isolates from *C. foetidissima* and *D. wrightii* group with fungus-infecting *betapartitiviruses* (Fig. 1.10C). In contrast, group PV4 appears to contain three distinct *Partitiviridae*-like viruses that are most similar to plant-infecting species (Fig. 1.10D). The only classified member in this tree is PCV2 (*Pepper cryptic virus 2*, genus *Deltapartitivirus*).

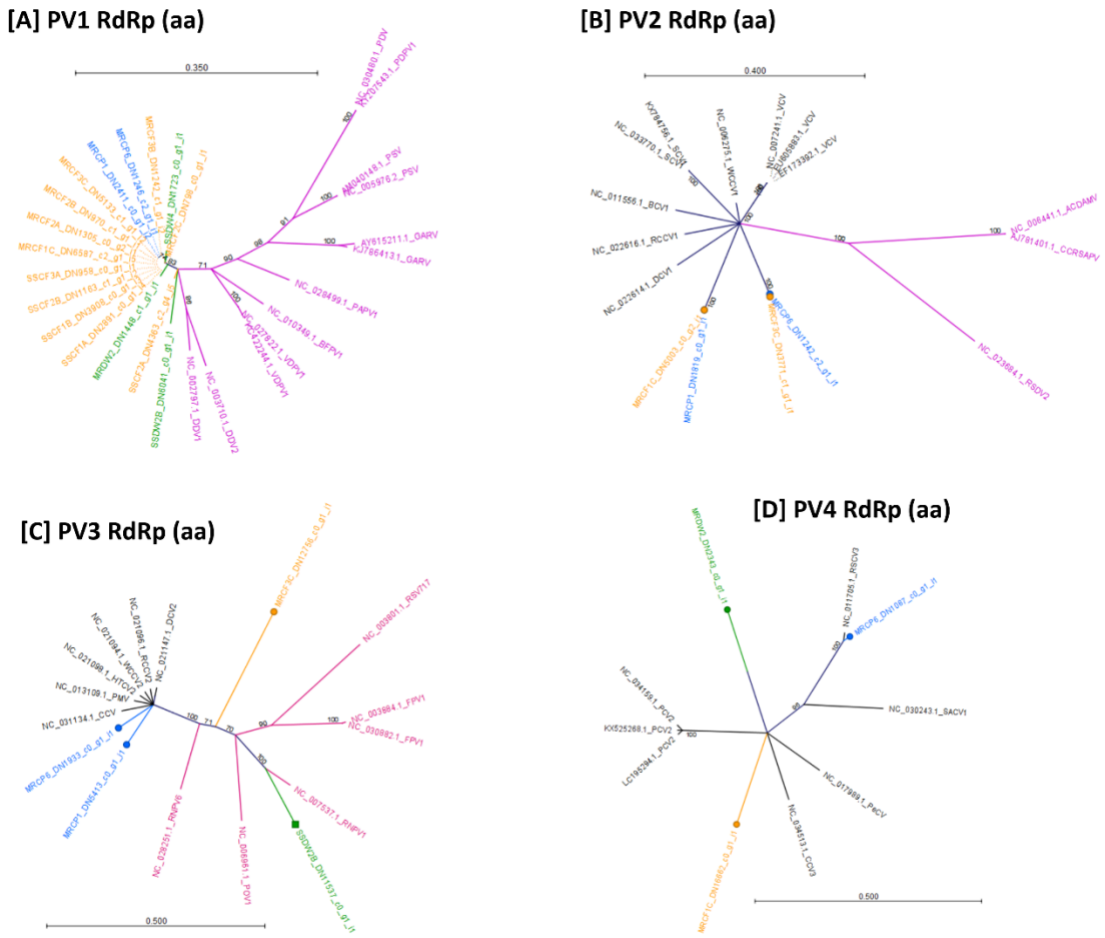

Figure 1.10. Radial, unrooted trees illustrating results of the maximum likelihood phylogenetic analyses of* Partitiviridae *amino acid sequences. *Trees in [A] through [D] illustrate relationships among putative* Partitiviridae *detections reported in Figure 5. Color coding has been applied to aid interpretation of sequence origins (Trinity assemblies → green =* D. wrightii, *orange =* C. foetidissima, *blue =* C. palmata*; Genbank sequences → purple = fungus-infecting viruses, black = plant-infecting viruses). Full virus names and genera for Genbank sequences are available in Supplementary Data Sheet 8. One hundred bootstrap re-samples were performed for each analysis. Branch lengths are measured in the expected number of substitutions per site. Only nodes with bootstrap values >= 70 are shown; nodes with bootstrap values < 70 are collapsed and not shown.*

Discussion

Knowledge gaps and resource needs in the field of plant virus ecology

Plant virus ecology is an expanding discipline. Its origins are rooted in the managed plant systems that served as the basis for advancements in plant virology, but its strengths lie in the translation of this research to ecological questions, and model systems, outside of agriculture. To quantify its development, we evaluated outputs of the NSF-funded Plant Virus Ecology Network (PVEN), which ignited the field of plant virus ecology by bringing together research teams and propagating new syntheses (NSF Award Abstract #0639139, 2011). We further quantified the impact of these outputs (precursor papers) by evaluating the ecological scope of studies that built upon them (product papers). Our quantitative synthesis of precursor papers suggests that the efforts of PVEN succeeded in promoting studies on virus interactions with non-cultivated (wild) host plants (Fig. 1.2A); most research categories had equal or greater representation of wild hosts relative to crop hosts. Although the majority of hosts were annuals, a sizeable proportion of studies included perennials as well (Fig. 1.2C). And within the *Virus discovery* research area, several studies focused on detecting and describing viruses not associated with disease in crops (Fig. 1.2D), which could serve as the focus of future studies in research areas outside of *Virus discovery*.

These patterns reflect a desire and need to expand outside of agricultural systems while maintaining study system feasibility. Annual wild plants are often more suitable for manipulative experiments than perennials due to faster generation times and greater greenhouse tractability. Crop-associated viruses are readily available as pure isolates and research on these viruses is fundable from both basic and applied grant sources. It is logical that initial efforts to expand the field of virus ecology would do so by hybridizing

features of agricultural systems with wild systems. However, our analysis of product papers suggests that this hybrid approach is becoming less common and there is a reversion to reliance on crop systems for addressing ecological questions. Among product papers, the only research areas having equitable proportional representation of wild plants with crops or models (either alone or in combination) are those in which collection surveys dominate over experimental approaches (*Epidemiology*, *Virus discovery,* and *Virus evolution*) (Fig. 1.3B). More troubling, crops and laboratory models (e.g., *Nicotiana benthamiana*) are preferred as hosts in the core research areas that are most likely to use manipulative experimental approaches to understand the roles of viruses in plant ecology (*Virus-vector interactions, Virus effects on host traits,* and *Environmental virology*) (Fig. 1.3B). Within these studies, wild plant hosts are not diverse—about half of all studies on wild plants use hosts from a single family: the Poaceae (Fig. 1.3C, Fig. 1.4).

Our quantitative synthesis suggests that there are persistent trade-offs in virus ecology research based largely on pathosystem logistics. This assessment is supported by our more detailed analysis of wild hosts used in studies within the *Virus effects on host traits, Virus-vector interactions*, and *Environmental virology* research areas (Fig. 1.4). Researchers prefer annuals, hosts that are common and prevalent (e.g., those exhibiting weedy characteristics or invasive tendencies), hosts that are related to crops or marginally domesticated, and hosts with easy propagation (Fig. 1.4). As a result of these biases in host choice, and a preference for using crop-associated viruses in manipulative studies, there are, and will continue to be, lingering gaps in our understanding of the ecological

roles of plant viruses in unmanaged systems. But these selections are not driven by neglect, but rather the need to learn what we can from what we have available in a time frame that aligns with funding sources. Altering this entrenched framework is not feasible for a single discipline. However, our quantitative synthesis reveals several avenues for enhancing the breadth of tractable study systems available for virus ecology research given these constraints. For example, we found that there is a pressing need for more studies on dicot perennials. We also found that researchers prefer to use hosts that are easy to find in the environment, relevant to agriculture, and simple to grow and propagate in the greenhouse. Therefore, the plant virus ecology research community should be working to identify and develop new wild dicot perennial systems that meet these criteria (Fig. 1.4). This goal formed the basis of our case study to characterize virus infections in several native, perennial dicots that are key species within an important Mediterranean-climate biodiversity hotspot—the Southern California Floristic Province (Myers et al., 2000).

Viruses detected via Illumina sequencing of dsRNA from candidate model hosts

Based on the criteria developed from our quantitative synthesis, we selected three co-occurring perennials, *C. foetidissima, C. palmata,* and *D. wrightii,* as candidate model systems. All three hosts satisfy the criteria laid out in Fig. 1.4. The two cucurbits are prevalent in Mediterranean-climate plant communities within California and in other arid grassland/shrub communities throughout the U.S. They are in the same genus as important crop hosts (summer and winter squash, gourds, and pumpkins), which have long served as models for understanding chemically mediated interactions among plants,

microbes, and insects in agroecosystems (Mauck and Shapiro, 2018). *Datura wrightii*, while not as closely related to crops, is one of the few wild perennial plants developed as a model for field-based chemical ecology research (Hare and Sun, 2011; van DAM and Hare, 1998). Expanding use of this model to include plant virus ecology is a natural extension of this research.

Using low-cost Illumina sequencing, basic molecular techniques, and open-source bioinformatics tools, we demonstrated that our candidate model hosts frequently harbor infections by viruses common in agricultural crops, as well as crop-associated viruses not known to be present in the region (ToCV). We also found evidence of novel host family associations (CABYV in *D. wrightii* and ToCV in *C. foetidissima*) (Fig. 1.5). Along with crop-associated virus detections, we uncovered multiple plant-associated viruses that appear related to known members of the family *Partitiviridae*, several of which are most similar to vertically-transmitted plant-infecting viruses (Fig. 1.5A, Fig. 1.10D). Our network analysis reveals the importance of certain viruses in the sampled host populations, as well as the ways in which hosts are connected by their virus associates (Fig. 1.6). Surprisingly, most crop-associated viruses are peripheral within the network. Because observational studies such as ours focus on the residual plants that have persisted in the face of infection pressure, further ecological studies are needed to determine the causes of this pattern. It may indicate low levels of infection pressure, strong control of infection by hosts, or high mortality among infected hosts. The interesting exception is CABYV, which has strong connections to cucurbit hosts at both sampling sites and may be adapting to infect co-occurring *Datura* hosts. CABYV is

transmitted in a persistent circulative manner by several aphid species, with the most efficient vector being the generalist cotton-melon aphid, *Aphis gossypii*, a global crop pest and vector of many viral plant pathogens. During subsequent field work in the 2018 season, we observed that *A. gossypii* is abundant on both cucurbit hosts. Populations build over the summer and peak in late August to early September. Density on individual hosts can become so high that tissue necrosis and premature senescence occur (Fig. 1.11). Based on these observations, and our finding that CABYV is prevalent across both sampling sites, we hypothesize that this virus–vector association is a significant part of the ecology of both cucurbit hosts. This virus may modify host traits, such as growth rates and flowering patterns (Alexander et al., 2017), or tolerance of abiotic stress (Davis et al., 2015; Westwood et al., 2013; Xu et al., 2008) in ways that alter survival and multi-year fitness. CABYV can also be transmitted by *Myzus persicae*, another generalist vector that can feed on both cucurbitaceous and solanaceous crop hosts (UC Statewide IPM Program, 2016). We observed *M. persicae* in low numbers on *Datura* and the invasive tree tobacco, *Nicotiana glauca*, and we hypothesize that this vector may be responsible for CABYV transmission from perennial cucurbits to *D. wrightii.* Using tissue from our CABYV positive *C. foetidissima* and laboratory-reared *A. gossypii*, we recently transmitted this pathogen to cultivated melons and future work will experimentally test host associations detected via Illumina sequencing.

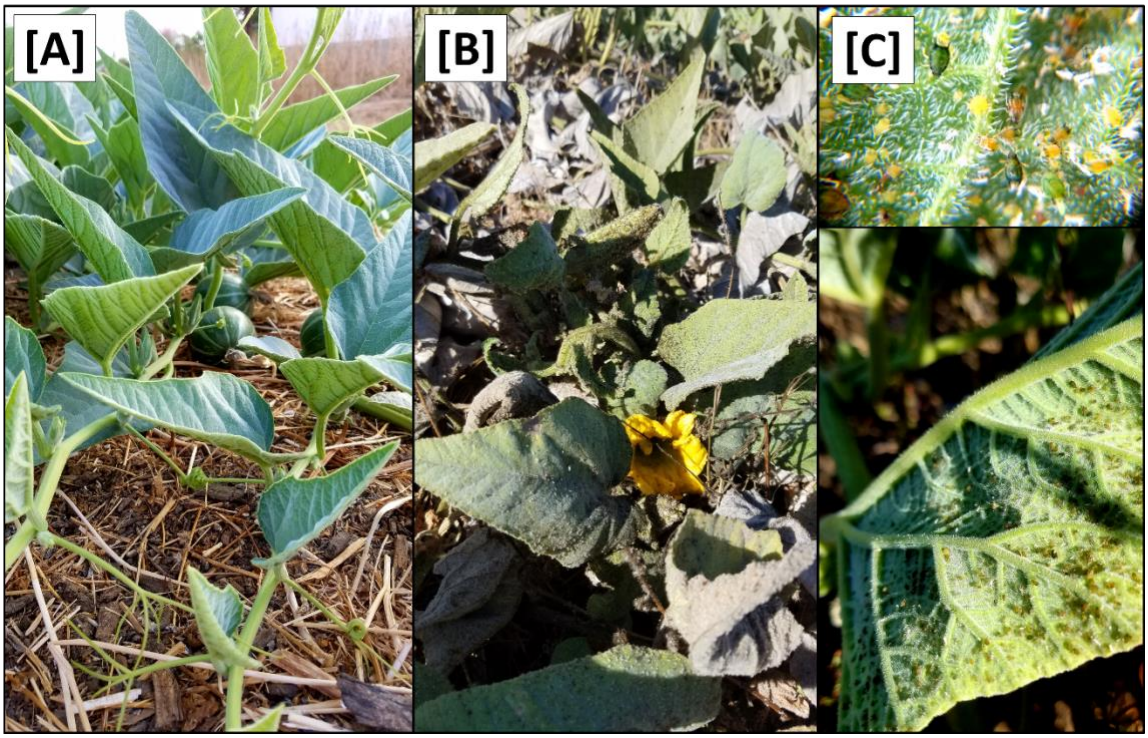

***Figure 1.11. Effect of* Aphis gossypii *feeding on* C. foetidissima.** *[A] Uninfested mature* C. foetidissima *plant. [B]* C. foetidissima *with heavy infestation by* A. gossypii. *Leaves are covered in sooty mold and exhibit premature senescence. [C] Typical A. gossypii densities on* C. foetidissima *leaves during heavy infestations.*

Ecological insights from phylogenetic analyses of crop-associated viruses

ToCV infections in *D. wrightii* and *C. foetidissima* confirm that this virus is present in California, even if it is not currently reported infecting solanaceous crops. Our ToCV isolates group together in the coat protein nucleotide sequence tree, with good bootstrap support for divergence from other U.S. isolates (Fig. 1.7A). And phylogenies based on the minor coat protein and RdRp sequences both indicate that the closest related isolates from Florida are still distinct from the ToCV detected in our study (Fig. 1.7B&C). However, more sequence data and host range studies are needed to assess whether our California ToCV isolates represent novel variants circulating in wild plant populations

and not recent introductions from crops. ToCV is present in Mexico within greenhouse production of solanaceous crop seedlings and illegally imported plant material from this region is one potential source of the virus. At present, there are no publicly available coding sequences for the coat protein, minor coat protein, or RdRp of Mexican ToCV isolates, but our results suggest that greater sampling in this region would help to clarify potential pathways for ToCV introduction to California.

There is no evidence that the ToCV isolate from the novel host (*C. foetidissima*) has diverged from isolates infecting solanaceous hosts (Fig. 1.7A). The coat protein nucleotide sequence of the ToCV isolate from *C. foetidissima* (SSCF1A) is not distinct from sequences of two other *D. wrightii* isolates collected at the same site (SSDW2B and SSDW2A). Deep sequencing also revealed instances of CABYV infecting *D. wrightii* (Fig. 1.5A). Based on studies with crops, this virus is known to infect numerous cucurbit hosts and a few diagnostic species but failed to infect crops in the Solanaceae in host range studies (Lecoq et al., 1992). Again, there is no evidence that the CABYV infecting the novel host is distinct from CABYV infecting a host that is related to known susceptible crop species (*C. palmata*), although two groups of CABYV are present in our populations, and only one contains an instance of infection in *D. wrightii* (Fig. 1.8). Additional targeted sampling of isolates from a larger number of hosts within our sites is necessary to determine if more host-specific genotypes co-occur alongside those capable of infecting hosts across the Cucurbitaceae and Solanaceae.

Even with limited sampling, our results demonstrate that virome characterization in wild perennial hosts can serve to expand our understanding of crop-associated virus host ranges while monitoring for new pathogens. Unusual associations might not occur in short-lived annual systems, but perennial hosts could undergo repeated exposure to inoculum from infected heterospecific perennial neighbors or adjacent agricultural crops over successive seasons. At present, we lack information on the effects of repeated inoculum exposures in perennials, and whether these lead to more virus diversity and mixing than in annuals. We might also expect that infection by one virus could modify susceptibility to subsequent pathogen inoculation, allowing novel associations to occur (Mukasa et al., 2006; Syller, 2012). For example, one cucurbit host infected with ToCV (genus *Crinivirus*) was also infected with ZYMV (Genus *Potyvirus*). Synergistic interactions among criniviruses and potyviruses have been documented in crop systems (Mukasa et al., 2006) and are one mechanism by which viruses overcome resistance traits. Additionally, while the perennial host genotype remains constant, the virus genotype will vary, creating opportunities for propagation and persistence of novel genotypes in new hosts if compatible insect vectors are present. In our pathosystems, *Cucurbita foetidissima* and *D. wrightii* are both highly suitable wild hosts for a major ToCV vector (*Bemisia tabaci* MEAM1) (van DAM and Hare, 1998; Wintermantel et al., 2016) and CABYV is transmitted by generalist aphids that feed on both cucurbit and solanaceous hosts. *C. foetidissima* and *D. wrightii* are also some of the only plants that are actively growing in our sites during the arid summer months, making them targets for vector activity in the absence of other options.

Descriptions of novel viruses

The "wild" viruses we detected are all putative members of the family *Partitiviridae*, which includes fungal, plant, and protozoan viruses (Nibert et al., 2014). PV1 sequences are all closely related and are most similar to fungus-infecting members of the genus *Gammapartitivirus*. In our survey, nearly identical PV1 sequences were associated with all three host plant species (Fig. 1.5A, Fig. 1.10A), suggesting that this virus could be infecting one or more generalist fungi capable of living in or on leaf tissue of diverse hosts. Groups PV2-PV3 contain several possible *Partitiviridae* species, some of which group most closely with mycoviruses (Fig. 1.10B&C). Complete characterization of putative mycoviruses is beyond the scope of our study and will require isolation of individual fungal associates from plant tissue and subsequent sequencing of the viromes of these pure isolates (Ong et al., 2017). Given that fungal viruses are also known to modify the outcomes of fungus-plant associations (Márquez et al., 2007), a deeper exploration of the viromes of plant-associated fungi could be of equal importance to expanded efforts to understand plant virus diversity and prevalence in more host species.

Unfortunately, far less is known about the ecological roles of plant-infecting *Partitiviridae* despite growing evidence that such viruses are common in wild plants (Roossinck, 2014; Thapa et al., 2015; Wren et al., 2006). Our study is no exception, with all three plant species hosting at least one possible plant-infecting member of the *Partitiviridae* (sequence group PV4 - Fig. 1.10D). Plant-infecting *Partitiviridae* are dsRNA viruses transmitted vertically only during host reproduction. They are not transmissible via grafting, mechanical inoculation, or vectors (Valverde and Navas-

Castillo, 2013). Because of the difficulty in obtaining genetically identical virus-free lines, little information is available about the impact of exclusively vertically-transmitted viruses on wild plant hosts.

In a subset of the crop systems for which virus-infected and virus-free lines are available, significant changes in phenotype and disease resistance due to infection with vertically-transmitted viruses are evident (Valverde and Navas-Castillo, 2013). Performing such studies in wild systems will be challenging, as vertical transmission is nearly 100 percent. However, our prevalence data for the most well-sampled species (*C. foetidissima*) (Fig. 5A) indicates that not all individuals are infected with putative plant-infecting *Partitiviridae* (PV4), even within a population. Future studies will leverage this natural variation, tractability of *C. foetidissima* for greenhouse research, and genomic resources for the genus *Cucurbita* (Paris, 2017; Sun et al., 2017), to begin exploring the biological relevance of dsRNA virus infections for plant fitness, stress tolerance, and resistance to infection by insect-vectored viruses.

Evaluating detection biases in bioinformatic workflows and downstream applications

We devised and tested a virome profiling workflow consisting of an inexpensive dsRNA extraction protocol (Kesanakurti et al., 2016), economical and rapid Illumina mid-output 75bp paired-end sequencing using the NextSeq 500 platform, and a user-friendly Galaxy-based bioinformatics protocol for virus discovery. This approach worked well in in the context of our study, the goal of which was detection of known crop-associated viruses and initial discovery of putatively novel viruses. But we acknowledge that longer reads are preferred for these purposes. New technologies with the capability to produce long

reads, such as the Nanopore platforms, are being validated as tools for virus discovery, and we expect these tools will become popular for plant virus ecology work. To explore the efficacy of our approach as a tool for enabling researchers from non-virology disciplines to perform virus discovery and prevalence studies with wild hosts, we compared options for pre-assembly host genome filtering and performed RT-PCR assays using previously published primer sets. Pre-assembly host genome sequence subtraction is often included as a step in virus discovery workflows to reduce the computational load during assembly (Daly et al., 2015). However, if this process removes virus sequences, it could hinder rather than help downstream analyses. In our study, read filtering against a host genome in the same genus (*Cucurbita maxima*) or subfamily (*Solanum lycopersicum*) significantly reduced the number of reads going into the BlastN analysis for both cucurbits (up to 83 % of reads subtracted), but had little effect when applied to the more distantly related *D. wrightii* host (up to 14 % of reads subtracted). Meanwhile, the identity of detected viruses and number of detections did not change. Our comparison suggests that a significant proportion of reads are host-derived, even when dsRNA is the extraction material, but sequence subtraction using publicly available genomes becomes less useful with increasing phylogenetic distance.

Feasibility of non-NGS detection methods is another important consideration when identifying candidate wild hosts for virus ecology studies, as wild plant tissues typically contain more inhibitors of RT and PCR reactions (Lacroix et al., 2016) which could make subsequent work challenging and halt progress at the virome characterization stage. At the same time, costs must be kept low, which is equally challenging when

dealing with frequent co-infections and the need to detect multiple viruses from the same sample. For our candidate hosts, we explored the use of general and virus-specific primers in RT reactions and looked for evidence of inhibitors by diluting RNA as previously recommended (Lacroix et al., 2016). For CABYV, use of random hexamers in the RT reaction resulted in detection of 12/17 of the Illumina detections, and additionally detected five possible infections that were missed by Illumina sequencing (Fig. 5B). The same pattern was seen for ToCV detections (Fig. 5B). Use of previously published, virus-specific primers on undiluted or 1/10 diluted RNA (to reduce the possible influence of inhibitors (Lacroix et al., 2016)) only slightly improved detection efficiency of RT-PCR vs. Illumina for CABYV, and only for the total RNA extractions. These data indicate that for detection of multiple viruses identified in our system, use of random hexamers is an acceptable, lower-cost alternative to virus-specific RT reactions. Additionally, our target hosts are not overly problematic with regard to inhibitors. Use of a RT reagent specifically labeled for use with inhibitor-prone samples was sufficient for robust detection.

Conclusions

Viruses are ubiquitous microbial associates within all habitats that support cellular life and represent the most important emerging infectious diseases of plants. By quantifying pathosystem characteristics across a representative sample of plant virus ecology literature, we identified key knowledge gaps that hinder our understanding of the diversity, prevalence, and ecological roles of plant viruses outside of agricultural systems. In particular, we found major need to increase the number of studies focusing on the

undomesticated, dicot perennial plants that form the basis of most native communities. Our case study directly addresses this need by evaluating the tractability of three keydicot perennials in fragmented semi-arid plant communities in southern California (USA), which is a Mediterranean-climate biodiversity hotspot (Myers et al., 2000).

Mediterranean-climate regions like this experience a unique off-set of temperature and precipitation peaks (winter rains, summer drought) that offers advantages for agriculture (e.g., ripening fruit experience less fungal disease) and human populations. As a result, these regions provide dramatic and clear examples of interfaces between annual crops and diverse native perennial vegetation, as likewise seen in South Africa and France (Bernardo et al., 2018) and Australia (Vincent et al., 2014), as well as with urbanized areas. Here, we expand the suite of possible dicot perennial study systems while providing a first glimpse into the ecology of plant viruses in North American Mediterranean-climate region communities. In doing so, we found evidence that some crop-associated viruses (e.g., CABYV) are strongly associated with our candidate hosts, while others (PRSV and ZYMV) were detected only sporadically. Our detection of ToCV demonstrates the usefulness of perennials as subjects for virus monitoring. And multiple instances of unusual virus-plant associations (based on host range studies with crops) indicate that our three candidate hosts may be strongly connected via their shared insect vector communities. Additionally, our study lays the groundwork for controlled field and greenhouse studies to explore virus effects on wild plant performance under variable environmental conditions - all of which are logistically feasible because our hosts were selected based on the criteria defined in Fig. 1.4. Perennials can suffer multi-year fitness

impacts of viruses, especially viruses whose prevalence in wild communities is driven by amplification in adjacent annual cropping systems (Alexander et al., 2013; Malmstrom et al., 2017; Malmstrom and Alexander, 2016). Alternatively, known and novel viruses may contribute to the drought and heat tolerance characteristics of their hosts (Carr, 2017; Davis et al., 2015; Xu et al., 2008). Both possibilities remain underexplored for wild plants, but manipulative studies with our candidate hosts and their pathogens are now possible with the resources provided by our case study.

References

Alexander, H. M., Bruns, E., Schebor, H., and Malmstrom, C. M. (2017). Crop-associated virus infection in a native perennial grass: reduction in plant fitness and dynamic patterns of virus detection. *J. Ecol.* doi:10.1111/1365-2745.12723.

Alexander, H. M., Mauck, K. E., Whitfield, A. E., Garrett, K. A., and Malmstrom, C. M. (2013). Plant-virus interactions and the agro-ecological interface. *Eur. J. Plant Pathol.* 138, 529–547. doi:10.1007/s10658-013-0317-1.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., and Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.* 19, 535–544. doi:10.1016/j.tree.2004.07.021.

Bernardo, P., Charles-Dominique, T., Barakat, M., Ortet, P., Fernandez, E., Filloux, D., et al. (2018). Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *ISME J.* 12, 173–184. doi:10.1038/ismej.2017.155.

Carley, K., Pfeffer, J., Reminga, J., Storrick, J., and Columbus, D. (2018). ORA User's Guide. Carnegie-Mellon University. Center of the Computational Analysis of Social and Organizational Systems (CASOS).

Carr, J. P. (2017). Exploring how viruses enhance plants' resilience to drought and the limits to this form of viral payback. *Plant Cell Environ.* 40, 2906–2908. doi:10.1111/pce.13068.

Chu, C. C., Henneberry, T. J., Natwick, E. T., Ritter, D., and Birdsall, S. L. (2001). Efficacy of CC traps and seasonal activity of adult *Bemisia argentifolii* (Homoptera: Aleyrodidae) in Imperial and Palo Verde Valleys, California. *J. Econ. Entomol.* 94, 47–54.

Daly, G. M., Leggett, R. M., Rowe, W., Stubbs, S., Wilkinson, M., Ramirez-Gonzalez, R. H., et al. (2015). Host subtraction, filtering and assembly validations for novel viral discovery using next generation sequencing data. *PLoS One* 10, e0129059. doi:10.1371/journal.pone.0129059.

Davis, T. S., Bosque-Pérez, N. A., Foote, N. E., Magney, T., and Eigenbrode, S. D. (2015). Environmentally dependent host–pathogen and vector–pathogen interactions in

the *Barley yellow dwarf virus* pathosystem. *J. Appl. Ecol.* 52, 1392–1401. doi:10.1111/1365-2664.12484.

Ellis, E. C., Klein Goldewijk, K., Siebert, S., Lightman, D., and Ramankutty, N. (2010). Anthropogenic transformation of the biomes, 1700 to 2000. *Glob. Ecol. Biogeogr.* 23, no–no. doi:10.1111/j.1466-8238.2010.00540.x.

Fischer, J., and Lindenmayer, D. B. (2007). Landscape modification and habitat fragmentation: a synthesis. *Glob. Ecol. Biogeogr.* 16, 265–280. doi:10.1111/j.1466-8238.2007.00287.x.

Hare, J. D., and Sun, J. J. (2011). Production of induced volatiles by *Datura wrightii* in response to damage by insects: effect of herbivore species and time. *J. Chem. Ecol.* 37, 751–764. doi:10.1007/s10886-011-9985-5.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22, 160–174.

Kesanakurti, P., Belton, M., Saeed, H., Rast, H., Boyes, I., and Rott, M. (2016). Screening for plant viruses by next generation sequencing using a modified double strand RNA extraction protocol with an internal amplification control. *J. Virol. Methods* 236, 35–40. doi:10.1016/j.jviromet.2016.07.001.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi:10.1093/molbev/msy096.

Lacroix, C., Renner, K., Cole, E., Seabloom, E. W., Borer, E. T., and Malmstrom, C. M. (2016). Methodological guidelines for accurate detection of viruses in wild plant species. *Appl. Environ. Microbiol.* 82, 1966–1975. doi:10.1128/AEM.03538-15.

Lecoq, H., Bourdin, D., Wipf-Scheibel, C., Bon, M., Lot, H., Lemaire, O., et al. (1992). A new yellowing disease of cucurbits caused by a luteovirus, cucurbit aphid-borne yellows virus. *Plant Pathol.* 41, 749–761. doi:10.1111/j.1365-3059.1992.tb02559.x.

Malmstrom, C. M., and Alexander, H. M. (2016). Effects of crop viruses on wild plants. *Curr. Opin. Virol.* 19, 30–36. doi:10.1016/j.coviro.2016.06.008.

Malmstrom, C. M., Bigelow, P., Trębicki, P., Busch, A. K., Friel, C., Cole, E., et al. (2017). 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 Res.* 241, 172–184. doi:10.1016/j.virusres.2017.07.006.

Malmstrom, C. M., Melcher, U., and Bosque-Pérez, N. A. (2011). The expanding field of plant virus ecology: historical foundations, knowledge gaps, and research directions. *Virus Res.* 159, 84–94. doi:10.1016/j.virusres.2011.05.010.

Márquez, L. M., Redman, R. S., Rodriguez, R. J., and Roossinck, M. J. (2007). A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. *Science* 315, 513–515. doi:10.1126/science.1136237.

Mauck, K. E., and Shapiro, L. R. (2018). "Chemically-mediated Interactions among Cucurbits, Insects and Microbes," in *Chemical Ecology of Insects* (CRC Press), 55–90.

Min, B.-E., Feldman, T. S., Ali, A., Wiley, G., Muthukumar, V., Roe, B. A., et al. (2012). Molecular characterization, ecology, and epidemiology of a novel *Tymovirus* in *Asclepias viridis* from Oklahoma. *Phytopathology* 102, 166–176. doi:10.1094/PHYTO-05-11-0154.

Motte Rimrock Reserve Statistics (2010). Available at: https://motte.ucr.edu/reservestats.html [Accessed November 20, 2018].

Motte Rimrock Species List (2009). Available at: https://motte.ucr.edu/specieslists.html [Accessed November 20, 2018].

Mukasa, S. B., Rubaihayo, P. R., and Valkonen, J. P. T. (2006). Interactions between a crinivirus, an ipomovirus and a potyvirus in coinfected sweetpotato plants. *Plant Pathol.* 55, 458–467. doi:10.1111/j.1365-3059.2006.01350.x.

Muthukumar, V., Melcher, U., Pierce, M., Wiley, G. B., Roe, B. A., Palmer, M. W., et al. (2009). Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. *Virus Res.* 141, 169–173. doi:10.1016/j.virusres.2008.06.016.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858. doi:10.1038/35002501.

Nibert, M. L., Ghabrial, S. A., Maiss, E., Lesker, T., Vainio, E. J., Jiang, D., et al. (2014). Taxonomic reorganization of family *Partitiviridae* and other recent progress in

partitivirus research. *Virus Res.* 188, 128–141. doi:10.1016/j.virusres.2014.04.007.

NSF Award Abstract #0639139 (2011). *Plant Virus Ecology Research Coordination Network*. Available at: https://www.nsf.gov/awardsearch/showAward?AWD_ID=0639139&HistoricalAwards=false [Accessed November 18, 2018].

Ong, J. W. L., Li, H., Sivasithamparam, K., Dixon, K. W., Jones, M. G. K., and Wylie, S. J. (2017). The challenges of using high-throughput sequencing to track multiple bipartite mycoviruses of wild orchid-fungus partnerships over consecutive years. *Virology* 510, 297–304. doi:10.1016/j.virol.2017.07.031.

Paris, H. S. (2017). "Genetic Resources of Pumpkins and Squash, Cucurbita spp," in *Genetics and Genomics of Cucurbitaceae* Plant Genetics and Genomics: Crops and Models., eds. R. Grumet, N. Katzir, and J. Garcia-Mas (Cham: Springer International Publishing), 111–154. doi:10.1007/7397_2016_3.

Pooggin, M. (2018). Small RNA-omics for plant virus identification, virome reconstruction and antiviral defense characterization. *Front. Microbiol.* 9, 2779. doi:10.3389/fmicb.2018.02779.

Rodelo-Urrego, M., García-Arenal, F., and Pagán, I. (2015). The effect of ecosystem biodiversity on virus genetic diversity depends on virus species: A study of chiltepin-infecting begomoviruses in Mexico. *Virus Evol* 1, vev004. doi:10.1093/ve/vev004.

Roossinck, M. J. (2014). Metagenomics of plant and fungal viruses reveals an abundance of persistent lifestyles. *Front. Microbiol.* 5, 767. doi:10.3389/fmicb.2014.00767.

Roossinck, M. J., and García-Arenal, F. (2015). Ecosystem simplification, biodiversity loss and plant virus emergence. *Curr. Opin. Virol.* 10, 56–62. doi:10.1016/j.coviro.2015.01.005.

Scheets, K., Blinkova, O., Melcher, U., Palmer, M. W., Wiley, G. B., Ding, T., et al. (2011). Detection of members of the *Tombusviridae* in the Tallgrass Prairie Preserve, Osage County, Oklahoma, USA. *Virus Res.* 160, 256–263. doi:10.1016/j.virusres.2011.06.023.

Stobbe, A. H., and Roossinck, M. J. (2014). Plant virus metagenomics: what we know and why we need to know more. *Front. Plant Sci.* 5, 150. doi:10.3389/fpls.2014.00150.

Sun, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y., et al. (2017). Karyotype stability

and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes. *Mol. Plant* 10, 1293–1306. doi:10.1016/j.molp.2017.09.003.

Syller, J. (2012). Facilitative and antagonistic interactions between plant viruses in mixed infections. *Mol. Plant Pathol.* 13, 204–216. doi:10.1111/j.1364-3703.2011.00734.x.

Thapa, V., McGlinn, D. J., Melcher, U., Palmer, M. W., and Roossinck, M. J. (2015). Determinants of taxonomic composition of plant viruses at the Nature Conservancy's Tallgrass Prairie Preserve, Oklahoma. *Virus Evol* 1, vev007. doi:10.1093/ve/vev007.

Thapa, V., Melcher, U., Wiley, G. B., Doust, A., Palmer, M. W., Roewe, K., et al. (2012). Detection of members of the Secoviridae in the Tallgrass Prairie Preserve, Osage County, Oklahoma, USA. *Virus Res.* 167, 34–42. doi:10.1016/j.virusres.2012.03.016.

Tzanetakis, I. E., and Martin, R. R. (2008). A new method for extraction of double-stranded RNA from plants. *J. Virol. Methods* 149, 167–170. doi:10.1016/j.jviromet.2008.01.014.

UC Statewide IPM Program (2016). UC IPM Pest Management Guidelines: Cucurbits. University of California, Davis Available at: http://ipm.ucanr.edu/PDF/PMG/pmgcucurbits.pdf.

United States Department of Agriculture, Natural Resources Conservation Service (2017). Web Soil Survey. Available at: https://websoilsurvey.nrcs.usda.gov/app/ [Accessed November 20, 2018].

Valverde, R. A., and Navas-Castillo, J. (2013). 1-1 Are persistent plant viruses novel endophytes? *Endophytes for plant protection: the state of the art*, 4.

van DAM, N. M., and Hare, D. J. (1998). Differences in distribution and performance of two sap-sucking herbivores on glandular and non-glandular *Datura wrightii*. *Ecol. Entomol.* 23, 22–32.

Vincent, S. J., Coutts, B. A., and Jones, R. A. C. (2014). Effects of introduced and indigenous viruses on native plants: exploring their disease-causing potential at the agro-ecological interface. *PLoS One* 9, e91224. doi:10.1371/journal.pone.0091224.

Weber, F., Wagner, V., Rasmussen, S. B., Hartmann, R., and Paludan, S. R. (2006). Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. *J. Virol.* 80, 5059–5064. doi:10.1128/JVI.80.10.5059-5064.2006.

Westwood, J. H., Mccann, L., Naish, M., Dixon, H., Murphy, A. M., Stancombe, M. A., et al. (2013). A viral RNA silencing suppressor interferes with abscisic acid-mediated signaling and induces drought tolerance in *Arabidopsis thaliana*. *Mol. Plant Pathol.* 14, 158–170.

Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18, 691–699. doi:10.1093/oxfordjournals.molbev.a003851.

Wintermantel, W. M., Gilbertson, R. L., McCreight, J. D., and Natwick, E. T. (2016). Host-specific relationship between virus titer and whitefly transmission of *Cucurbit yellow stunting disorder virus*. *Plant Dis.* 100, 92–98. doi:10.1094/PDIS-11-14-1119-RE.

Wintermantel, W. M., and Wisler, G. C. (2006). Vector specificity, host range, and genetic diversity of *Tomato chlorosis virus*. *Plant Dis.* 90, 814–819. doi:10.1094/PD-90-0814.

Wren, J. D., Roossinck, M. J., Nelson, R. S., Scheets, K., Palmer, M. W., and Melcher, U. (2006). Plant virus biodiversity and ecology. *PLoS Biol.* 4, e80. doi:10.1371/journal.pbio.0040080.

Xu, P., Chen, F., Mannas, J. P., Feldman, T., Sumner, L. W., and Roossinck, M. J. (2008). Virus infection improves drought tolerance. *New Phytol.* 180, 911–921. doi:10.1111/j.1469-8137.2008.02627.x.

Yang, Z. (1994). Estimating the pattern of nucleotide substitution. *J. Mol. Evol.* 39, 105–111.

Chapter 2. Geographic and ecological drivers of virus prevalence and community composition in drought-adapted perennial plants

Abstract

Plant viruses have historically been studied in the context of agriculture, but the advancement of sequencing technology provides new opportunities to study plant virus ecology in wild plants is expanding. Within virus ecology, most studies on wild plants to date focus on the discovery of viruses. More studies are needed to understand how virus communities are structured in wild systems. In the present study, we combine community ecology analyses with high throughput sequencing and virome characterization to understand species and landscape-level drivers of virus diversity in wild plants. We focus on three native, perennial species that in drought and heat-adapted chaparral communities: *Cucurbita foetidissima*, *C. palmata*, and *Datura wrightii*. Both cucurbit species grow along roads, washes, and arroyos, and *C. foetidissima* a prevalent road-side weed while *C. palmata* grows more sporadically but can be found in large stands (Kates, 2019; Bolley et al., 1950). *Datura wrightii* also commonly occurs in river washes, slopes, and roadsides (Van Dam and Hare, 1998; Personal Observation). We sampled these plants from three ecological reserves (preserved wildlands/remnant vegetation) that varied in both size and adjacent land cover types. We found 12 crop-associated, insect-transmitted viruses infecting target plants across the reserves including *Alfalfa mosaic virus*, *Beet western yellows virus*, *Cucurbit aphid-borne yellows virus*, *Cucumber mosaic virus*, *Cucurbit yellow stunting disorder virus*, *Papaya ringspot virus*, *Squash mild leaf*

curl virus, *Squash vein yellowing virus*, *Tomato chlorosis virus*, *Tomato necrotic dwarf virus*, *Watermelon mosaic virus*, and *Zucchini yellow mosaic virus*. Based on virome results, we selected two economically important cucurbit-infecting viruses for a broader surveillance effort: *Cucurbit aphid-borne yellows virus* (CABYV) and *Cucurbit yellow stunting disorder virus* (CYSDV). Virus prevalence varied among reserves and host species; prevalence was highest at 88% for CABYV at one reserve, but as low as zero detections for a different reserve. CYSDV was highest at 37% in one reserve, but rare or nonexistent in other locations. We also found that CABYV is retained across seasons in both *Cucurbita spp.* hosts. Overall, we found that space between reserves, host identity, and indicator species are driving factors of virus communities, and that host identity also drives the prevalence of two of the most common crop-viruses in these wild plants.

Introduction

Human development increasingly encroaches on natural areas. This leads to fragmented and patchy wild lands that are immediately adjacent to agricultural, suburban, and urban environments. Exposure of wild land edges to these areas creates opportunities for the transfer of organisms between developed and undeveloped plant communities (Bar-Massada et al., 2014). This includes exposure to plant-infecting viruses via the movement of insect vectors (Alexander et al., 2014). Effects of these viruses on plant communities outside of agriculture are only beginning to be explored, but evidence to date suggests that crop virus infections negatively affect wild plants, especially long-lived perennial hosts for which lifetime fitness occurs over multiple seasons (Alexander et al., 2017; Malmstrom & Alexander, 2016).

Much of what we know about viruses in wild plant communities comes from studies using high throughput sequencing (HTS) to characterize viromes of wild hosts (Maclot et al., 2020). Plant virology in the pre-PCR and pre-sequencing eras relied on studying viruses in crop hosts and indicator hosts, where symptoms were apparent. But symptoms are not always apparent in wild plant hosts or may manifest differently relative to crop hosts (Malmstrom et al., 2017). Affordable HTS technologies, combined with growing economic resources, have enabled targeted detection of viruses infecting wild plants regardless of symptoms and without *a priori* knowledge of what might be present (Stobbe & Roossinck, 2014).

Studies employing HTS to characterize wild plant viromes have revealed that viruses are common and likely play important roles in natural ecosystems. Wild plants

support infections by viruses that cause disease in crops (crop-associated viruses) as well as novel viruses that may be vertically transmitted, long-term symbionts of their host plants (Maclot et al., 2020; Roossinck, 2010). Furthermore, virus infections in wild plants are highly prevalent; wild plants are commonly infected with at least one, and often several viruses (Prendeville et al., 2012; Shates et al., 2019; Susi et al., 2019). Limited virome work at the landscape scales suggests possible connections between the composition of the landscape surrounding wild lands and the virus diversity within individual non-crop hosts and plant communities (Bernardo et al., 2018). These studies suggest that viruses are important players in wild plant ecology, and that there are as-yet undescribed factors driving the assembly of virus communities within host individuals, populations, and communities.

In the present study, we combine community ecology analyses with HTS and virome characterization to understand species and landscape-level drivers of virus diversity in wild plants. As a study system, we focused on three key perennial species that thrive in the summer period in Southern California chaparral and grassland ecosystems (Fig. 1). We chose two plants that are congeneric to pumpkins and summer squash, *Cucurbita foetidissima* Kunth and *Cucurbita palmata* Wats., and one plant in the nightshade family, *Datura wrightii* Regel. All three plants grow in the height of summer when other plants are dormant, creating a strong visual signal of green foliage against a brown background. They senesce in the winter before regrowing the following year, surviving dormancy and hot summers using large taproot systems (Dittmer and Talley, 1964; Bye, 2001). They can grow in large populations, but also as sprawling individuals

with vines several meters long (e.g., *C. foetidissima*), which are particularly apparent to flying insect vectors traveling on air currents or foraging locally (Schröder et al., 2017). Based on these characteristics, we hypothesized that these three key perennial species are frequently exposed to both vectors and crop-associated viruses. To explore this in the context of habitat fragmentation, we sequenced viromes of plants growing in three reserves of varying size and likely exposure to human-associated pests (such as crop viruses). The smallest reserve, Motte Rimrock Reserve is 736 acres and embedded in a patchwork of areas developed for housing and businesses. Shipley Skinner Multispecies Reserve is 14,000 acres and is adjacent to both housing, wildlands, and ranchland/crops. Anza-Borrego Desert State Park, the largest state park in California, is 585,930 acres. There are some small buildings and roads through the park. However, compared to the other two reserves, the land is significantly more intact with little adjacent urbanization. Each site houses mixed populations of the target hosts which were sampled and processed for virus community characterization using HTS.

Figure 2.1. *A: The three reserves are located in Riverside County, with Anza-Borrego extending to San Diego County. Motte Rimrock is the smallest and is embedded in a suburban setting. Shipley Skinner is the next largest and is surrounded by vineyards, homes and ranches. Anza-Borrego is the largest state park in California, with the least fragmentation and development activity within or adjacent to the land area. B-C: Images of the target perennial hosts, Cucurbita foetidissima (B), Datura wrightii (C), and Cucurbita palmata (D). All three hosts are heat and drought adapted, growing primarily from May through August.*

To understand how virus communities varied by host species and reserve, we selected methods common for analysis of ecological community data, including calculating species diversity and richness, PERMANOVA, principal coordinate analysis (PcoA), redundancy analysis, indicator species analysis, and distance-based Moran's Eigenvector Maps. Beyond understanding the diversity of viruses within individual hosts and across communities, we also leveraged virome data to inform follow-up targeted detection studies over multiple years. Time is an overlooked aspect of virus ecology in

wild hosts. This reflects the historical focus on annuals, where only a single time point or year is often presented (Prendeville et al., 2012; Shates et al., 2019). However, the fitness of perennial hosts involves multiple seasons, between which the aboveground tissue dies back and completely regrows from roots. Virus infections may have cumulative effects if retained from one season to the next (Alexander et al., 2017). But year-to-year virus dynamics within hosts are unknown for most wild plants. To address this in our system, we tracked the year-to-year prevalence of the crop-associated viruses most commonly detected with HTS.

Methods

Field sampling and tissue handling

We collected plant tissue from three perennial dicot target species (*C. foetidissima*, *C. palmata*, and *D. wrightii*) growing in three reserves in southern California: Anza-Borrego Desert State Park (June 2019), Motte Rimrock Reserve, and Shipley Skinner Multispecies Reserve (August 2017) (Fig 1) (Shates et al., 2019) . Samples from Anza-Borrego were collected approximately 5-50 meters from roads that allowed access to interior regions of the park. For all collections, we sampled both leaf and stem tissue from each plant by inverting a clean plastic Ziplock bag over the tissue and pulling with force, then sealing the bag. Approximately 10 grams of leaf tissue were taken for dsRNA extraction, and 4 grams of leaf tissue were carefully partitioned into 50-mL Rnase-free Falcon tubes and stored in −80◦C until further processing. In total, we collected 20 *C.*

foetidissima, 14 *C. palmata* and nine *D. wrightii* for deep sequencing from all three reserves, with each plant sampled once.

We also collected leaves from additional individuals to estimate the prevalence of select viruses within each reserve. At Anza-Borrego, we collected tissue from an additional 20 *C. foetidissima*, 44 *C. palmata*, and 10 *D. wrightii* individuals from each sampling location within the reserve (seven larger populations and four scattered individuals) at the same time that we sampled hosts for HTS in 2019. Each individual contributed a single sample. For these collections, which required less tissue for total RNA extraction, we only collected one leaf per plant. At Motte Rimrock and Shipley Skinner reserves, we returned to the plant populations and collected leaves from ~15-20% of the plants (estimated after initial counts early in the summer) in 2018, 2019 and 2020. All surveyed plants and metadata are in Supplementary Table 2.1.

GPS coordinates for sampled plants were recorded by photographing plants using phones and accessing photograph metadata. All three hosts are long-lived and perennial with large tap roots, which is ideal for resampling. However, the cucurbits can grow large and vines can grow many meters long. We attempted to return to the same individuals year-to-year, but due to variation in plant growth each year, only a subset of plants sampled could conclusively be used for assessing year-to-year virus retention (Supplementary Table 2.1).

Land cover assessment

We used QGIS, LecoS (Landscape Ecology Statistics, University

of Copenhagen, Copenhagen, Denmark) plugin (Jung, 2016), and the 2019 National Land Cover Database (https://www.mrlc.gov/data/nlcd-2019-land-cover-conus) to quantify the proportion of land cover types at varying distances from populations. Sampled plants are considered within one population if they were within three kilometers of each other based on *Cucurbita* spp. Outcrossing distances (Kirkpatrick & Wilson, 1988), similar to other work sampling wild cucurbits for pathogens (Prendeville et al., 2012). We quantified the land cover at one, three, and five kilometer buffers from the population because previous work showed that land cover can have predictive values for virus infections at one km and five km distances (Angelella et al., 2016; Ingwell et al., 2017).

Nucleic acid extractions, sequencing, and virus detection

We extracted double-stranded RNA (dsRNA) for non-targeted HTS for all samples from which we collected 10g of tissue (Supplementary Table 2.1) (Kesanakurti et al., 2016; Ma et al., 2019; Shates et al., 2019). For quality control, we spiked one leaf punch of bell pepper cv. California Wonder into each sample before extraction. This tissue was infected with bell pepper endornavirus (BPEV, *Endornaviridae*). Before library preparation, we performed reverse transcription-PCR for BPEV to evaluate virus recovery. To construct libraries (one per sample), we denatured an aliquot of 5µl of dsRNA solution from each extraction by incubating at 99◦C for 5 min, then placed the solution on ice and used the NEBNextR UltraTM II Directional RNA Library Prep Kit for Illumina by following the manufacturer's protocol. The average insert size was 250–300 bp. Sequencing was done at the UCR Genomics Core on the Illumina NextSeq platform on a Mid Output flow cell (v2) with a paired-end 75 bp read-length, which can

produce 260M paired-end reads. Adaptors were trimmed by Core staff before we processed the output using a workflow we previously designed for the Galaxy Platform (Afgan et al. 2018; Shates et al. 2019). We filtered out host genomes, performed de novo assembly using Trinity (Grabherr et al., 2011), and nucleotide blast to find putative virus identities from a database we generated using 2017 NCBI GenBank plant virus listing by plant host. Sequencing results for Motte Rimrock Reserve and Shipley Skinner Multispecies Reserve were previously published in (Shates et al., 2019) (GenBank read archive accession SRP149013) and results for collections from Anza-Borrego are published here.

For targeted detection of select viruses identified through HTS, we extracted total RNA from the leaf tissue of plants subjected to HTS and the additional plants sampled to confirm HTS directions and estimate virus prevalence in each location. We separated 100mg of leaf tissue into 2mL Eppendorf tubes with two stainless steel grinding balls (4mm from SpexSamplePrep), then dipped in liquid nitrogen. Then, we homogenized samples in the Geno/GrinderR (SPEX SamplePrep) for one minute at 1100 RPM. We used TRI Reagent (Sigma-Aldrich), with manufacturer's protocols for tissue and a LegendMicro21R microcentrifuge (ThermoScientific) at 4ºC for all steps. At the RNA precipitation step, we added 250µL isopropanol and 250µL of 0.8M sodium citrate and 1.2M sodium chloride solution. Following TRI Reagent extractions, RNA was purified to remove excess polysaccharides and other inhibitors that are frequent contaminants accompanying nucleic acid extractions from wild plants (Lacroix et al., 2016). For purifications, we added 500µL water to each sample, followed by 50µL 3M Sodium

acetate (pH 5.2) and 500µL of room temperature isopropanol. Samples were mixed well and incubated at room temperature for 20 minutes. Next, we pelleted RNA by centrifuging at 12,000 x g and 4°C, then washed the pellet with 500µL ice cold 70% ethanol and centrifuged each sample for 30 seconds at 7,500 x g, twice. After carefully removing the ethanol, the pellets were air-dried for 15 minutes, then resuspended in 50µL ultrapure water. In order to facilitate dissolving the pellet, we incubated the samples for 10 minutes at 55°C before storing in -80°C or proceeding with reverse-transcription PCR.

For the viruses present in plants from Motte Rimrock and Shipley Skinner Multispecies reserves, we previously confirmed putative virus identities (Shates et al., 2019). Here, we repeated these steps (as follows) for the plants from Anza-Borrego reserve plants. We used targeted reverse-transcription PCR with virus-specific PCR primers for the coat protein for known virus species to confirm virus identities from double-stranded RNA extracts. All reverse transcription and PCR was conducted in a Bio-Rad T100TM Thermal Cycler. Primers are listed in Supplementary Table 2.1. When dsRNA extracts were processed, we first denatured the material by incubating 2 µl at 99°C for 5 min. We then placed the solution on ice and performed reverse transcription as for total RNA. The reaction components include Ultrapure water, 10mM dNTP mix, random hexamers, 100mM DTT, and 5x SSIV Buffer. We used an enzyme (Superscript IV, Invitrogen) that is robust against most plant-derived inhibitors, and added an Rnase inhibitor (Ribolock, Thermo Fisher) to prevent RNA degradation. We used random hexamers in the reverse transcription instead of gene-specific primers because this method worked best in trial tests and because cDNA products can be used for detecting

any RNA virus present in the sample but otherwise followed the manufacturer's instructions. The subsequent PCR reactions (20 µl reaction volume) consisted of 1 µl of template cDNA, 4 µl of 5X HF buffer (Thermo Fisher Scientific or NEB), 1 µl of each 10 µM primer, 2 µl of dNTP mix (2 mM each) and 0.2 µl of Phusion DNA polymerase (Thermo Fisher Scientific & NEB). One PCR protocol was used for all viruses: 98°C for 3 min for initial denaturation, followed by 40 cycles of 98°C for 10 s, 55°C for 30 s, 72°C for 1 min, then 72°C for 10 min. For Sanger sequencing on subsets of PCR products, amplicons were purified using a Zymo Gel Recovery Kit. Sanger sequencing was performed by Retrogen Inc. (San Diego). For total RNA samples, we followed the same reverse transcription-PCR steps with the exception of the amount of RNA solution used in each reaction (approximately 0.5-10µL of RNA from each extract depending on concentration).

Statistical Analysis

Analysis of all data was performed using R version 4.0.3 (R Core Team, 2021).

Statistics for land cover, prevalence of two viruses, and retention

HTS revealed that two crop-associated viruses dominated plant communities in different sites: *Cucurbit aphid-borne yellows virus* (CABYV) and *Cucurbit yellow stunting disorder virus* (CYSDV). These were targeted for detection in additional tissue sampled to estimate prevalence of select virus species. We calculated the proportion of detections for each year, at each reserve, for each virus x host species. Then, we used two separate generalized linear models to determine significant factors in the prevalence dataset for either CABYV or CYSDV detection using the package *lme4* (Bates et al., 2013), package

emmeans (Lenth, 2019) and false discovery rate adjustment for post hoc tests. We used CABYV or CYSDV detection as the response variable, and as factors we included the sample reserve, kilometers by land cover proportion surrounding those populations in the reserves, plant species, and year sampled (only possible for Shipley Skinner Multispecies and Motte Rimrock reserves). Post hoc tests were performed on factors that were significant, shown by an analysis of deviance from the *car* (Fox and Weisberg, 2019) package on the model. To explore differences in virus retention by host species, the proportion of positive detections per plant species was compared between 2019 and 2020 using a two-tailed Z-test of two proportions.

Analysis of virus communities by site and host species

This analysis, and all of the following, use data from full virome datasets. To determine likely drivers for the assembly of virus communities and identify key differences between regions, we calculated beta diversity and used permutational multivariate analysis of variance (PERMANOVA), principal coordinate analysis (PcoA), redundancy analyses (RDA), indicator species analysis (ISA), and distance-based Moran's Eigenvector Maps (dbMEM). For these analyses, we removed plant individuals that did not have any detected crop-associated viruses (Shipley Skinner DW1, DW4, CF1B; Motte Rimrock CP4, CF1A) because many of these analyses do not tolerate zero values. The ~6 undescribed viruses (<70% nucleotide identity match to a known virus species) found across all samples were removed from analyses because their identities could not be confirmed putting them beyond the scope of this study.

Diversity metrics

Gamma diversity represents the total number of virus species found across all the hosts. Alpha diversity was calculated as species richness and Pilou's J for evenness between plant pieces and sites. Species richness is the number of species per site. Our virome data set consists of presence-absence data. Therefore, species diversity (Shannon or Simpson diversity, or Hill numbers) was not calculated because a presence-absence data set will produce the same abundance for all species. There are many metrics to calculate beta diversity, the link between alpha (local) and gamma (regional) diversity. We chose Sorenson's dissimilarity metric for beta diversity calculations because the virus species data are present-absence data, and this is a more robust metric with extreme differences between communities. We used the *vegan* (Oksanen et al., 2020) and *ade4* (Drey et al., 2021) packages to calculate beta diversity. The output values of beta diversity are intermediates and are used in the PcoA analysis.

Pattern testing

To test for differences in community composition, we used a permutational multivariate analysis of variance that employs distance matrices (*PERMANOVA*). This procedure tests for similarities among objects belonging within and across groups in the context of environment variables. The overall PERMANOVA tests for location effects of differences among communities in vector space. In addition to this procedure, we also evaluated the groups as pairs to determine which groups are significantly different from each other. We used the R packages *ade4* (Drey et al., 2021A), *vegan* (Oksanen et al., 2020), and *devtools* (pairwiseAdonis) (Martinez, 2020).

Ordinations

To visualize the composition of the three communities, we created a principal coordinate analysis (PcoA) plot using the packages *rgl* (Murdoch and Adler, 2021), *MASS* (Ripley, 2021), *vegan* (Oksanen et al, 2020), *ecodist* (Goslee and Urban 2020), *magick* (Ooms, 2021), *ggplot2* (Wickham et a., 2021a), and *dplyr* (Wickham et a., 2021b),. We also performed a redundancy analysis (RDA) to produce an ordination that shows the variation of response variables explained by explanatory variables (R packages *vegan* and *ggvegan* (Oksanen et al., 2020; Simpson, 2021)). We included reserve, plant species, plant family, plant community type (mixed species, single species, or individual plant), and site elevation as potential explanatory variables for community composition.

Drivers of virus community structure

We used an indicator species analysis (ISA) to determine differences among virus communities by reserve and plant host species (R package *indicspecies* (De Cáceres et al., 2020)). The analysis computes a *p*-value using permutation and gives an ISA value of 0-1, 1 being the strongest indicator value. Then, we used distance-based Moran's Eigenvector Maps (dbMEM) to calculate how much space/distance between communities contributes to differences in community composition. To do this, we first plotted spatial correlograms to find evidence for spatial correlations based on diversity (R package *vegan* (Oksanen et al., 2020)). Next, we calculated euclidean distances among plots and constructed dbMEM eigenfunctions (R package *adespatial* (Dray et al., 2021B)). We used an RDA for the forward selection process to isolate the positive and significant spatial variables (R package *adesptial*).

Results

Virus detection by HTS

We found 12 crop-associated viruses across all plant species and reserves (regional diversity/gamma diversity), and at least six unknown viruses. These unknowns were not considered in subsequent analyses because without further characterization, we could not confirm that the same unknown virus was present across multiple host individuals (Fig. 2.2). The crop-associated viruses were all at least 90% nucleotide identity match to a known virus species, and were all insect-transmitted by sap-sucking insects in the order Hemiptera, either family Aphididae or Aleyrodidae (aphids and whiteflies) (see Supplementary Table 2.1).

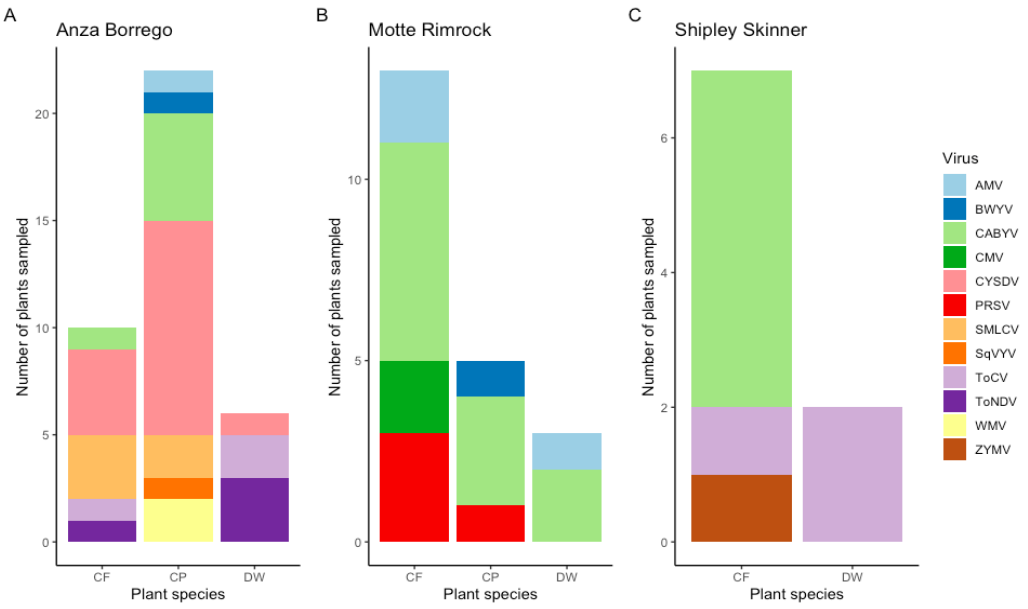

Figure 2.2. Gamma diversity *Crop-associated viruses detected in this study. Virus acronyms: AMV*= Alfalfa mosaic virus, *BWYV*= Beet western yellows virus, *CABYV* = Cucurbit aphid-borne yellows virus, *CMV* = Cucumber mosaic virus, *CYSDV*= Cucurbit yellow stunting disorder virus, *PRSV* = Papaya ringspot virus, *SMLCV* = Squash mild leaf curl virus, *SqVYV* = Squash vein yellowing virus, *ToCV* = Tomato chlorosis virus, *ToNDV* = Tomato necrotic dwarf virus, *WMV* = Watermelon mosaic virus, *ZYMV* = Zucchini yellow mosaic virus.

Diversity metrics

From full virome sequencing, we detected 12 crop-associated viruses across all plant species and reserves (regional diversity/gamma diversity), and at least six unknown viruses removed from subsequent analyses (gamma diversity). Five plants were removed before the analysis because no crop viruses were detected. For species richness, most of the 37 plants were infected with between one and three crop viruses, and one plant was infected with four viruses (Fig. 2.3). Based on analysis by ANOVA, beta diversity values for sites by reserve (P=0.95) and by host species (P=0.08) did not differ significantly from each other (Table 2.1).

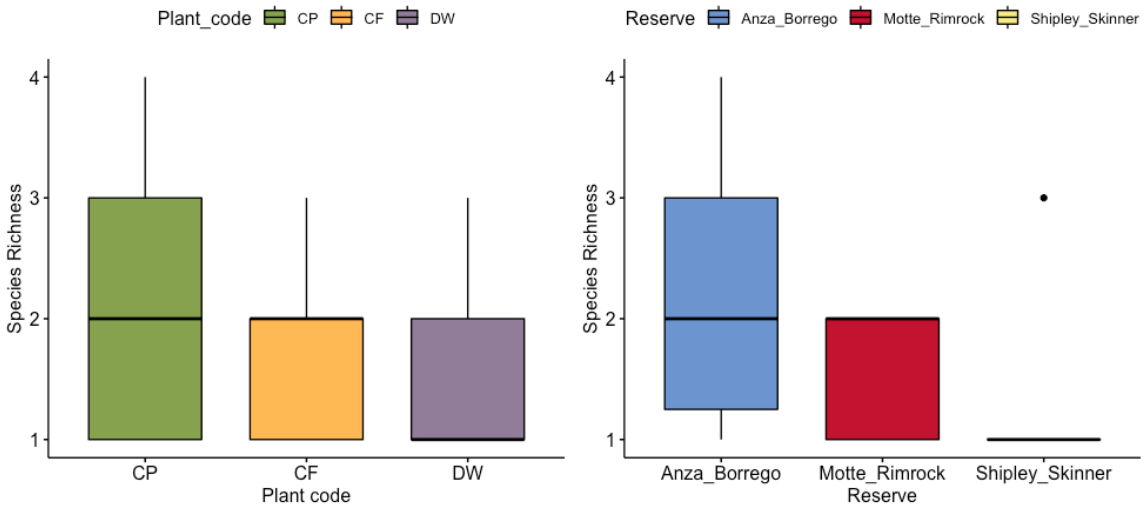

Figure 2.3. *Species richness of each site visualized by plant host species and reserve. Plant codes are CP*= C. palmata, *CF*= C. foetidissima, *and DW*= D. wrightii.

Type	Factors	Beta
Reserve	Anza Borrego	0.405
Reserve	Motte Rimrock	0.337
Reserve	Shipley Skinner	0.408
Plant species	C.foetidissima	0.482
Plant species	C.palmata	0.352
Plant species	D.wrightii	0.498

Table 2.1. *Beta diversity values based on Sorenson matrix by reserve and by plant species.*

Pattern testing

We used a permutational multivariate analysis of variance using distance matrices (PERMANOVA) to test for virus community differences by reserve, plant species and reserve by plant species interactions. This analysis was followed by a pairwise PERMANOVA for each pair. The overall PERMANOVA and model fit were all significant, but pairwise tests showed that not all pairs were significantly different from each other. Neither the reserve pair Motte Rimrock and Shipley Skinner Multispecies Reserve nor both *Cucurbita sp.* differ from each other (Table 2.2).

Group	Test	R2	P
Reserve	PERMANOVA	0.39	0.001
Plant species	PERMANOVA	0.13	0.001
Reserve:Plant	PERMANOVA	0.13	0.001
Overall model	MODEL FIT	0.66	0.001
Anza vs Motte	Pairwise	0.35	0.003
Anza vs Shipley	Pairwise	0.28	0.003
Motte vs Shipley	Pairwise	0.14	0.102
CP vs CF	Pairwise	0.42	0.126
CP vs DW	Pairwise	0.001	0.003
CF vs DW	Pairwise	0.011	0.033

Table 2.2. *Overall PERMANOVA and pairwise PERMANOVA R^2 values and P-values. Significant and nonsignificant interactions reported.*

Ordinations

The PcoA plot shows some degree of grouping by reserve, with Motte Rimrock Reserve and Shipley Skinner Multispecies Reserve being most similar to each other and Anza-Borrego being distant from those two reserves (Fig. 2.4A). There are less obvious groups by species among the two smaller reserves. But for Anza-Borrego, there is clear clustering of *C. palmata* separately from *C. foetidissima* and *D. wrightii*. The RDA, which included reserve, plant species, and other biotic and environmental variables (plant family, plant community type [mixed species, single species, or individual plant]) shows that many of the variables group together in the center (most similar) but others spread out along axes (Fig. 2.4B). For host species *C. palmata* individuals from only CP as well as those with CP intermixed with other species separate out in space from other hosts and are associated with CABYV and CYSDV. For host species *D. wrightii* individuals from a plant community made up of only *D. wrightii* are associated with the viruses ToCV and ToNDV and go in the opposite direction in coordinate space from the *C. palmata*-associated group. The original R^2 value was 0.5636597, but after adjusting was 0.402, indicating that variables input into the analysis explain about 40% of the variation in virus communities.

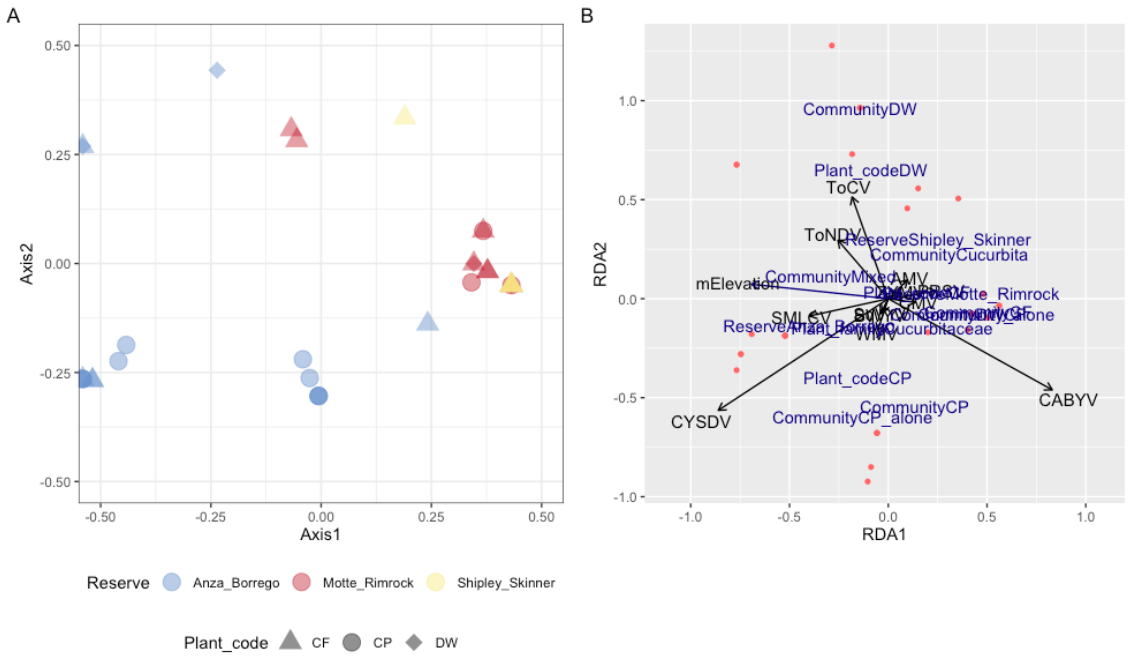

Figure 2.4. *A) PcoA plot made with* ggplot2 *(Wickham et al., 2021A) to explain variation by each axis. Points color coded by reserve, and assigned shapes based on host plant species of the community. B) RDA plot made with* ggvegan *(Simpson, 2021). Virus acronyms are in black font, variables in the RD analysis are in blue font, and sites/plant hosts are represented by red points.*

Drivers of virus community structure

The results from the indicator species analysis (Table 2.3) suggest that for differences between hosts in different reserves, there are four virus species driving differences among reserves and three virus species impacting the differences between host species. CYSDV is an indicator species for Anza-Borrego. *Papaya ringspot virus* (PRSV) is an indicator species for Motte Rimrock Reserve. *Tomato chlorosis virus* (ToCV) is an indicator species for Shipley Skinner. CABYV is an indicator for both Motte Rimrock and Shipley Skinner Multispecies reserves. Indicator species for communities in different plant

species are CYSDV for *C. palmata* and ToCV and *Tomato necrotic dwarf virus* (ToNDV) for *D. wrightii.*

Type	Group	Virus_species	ISA_value	P_value
Reserve	Anza Borrego	CYSDV	0.913	0.001
Reserve	Motte Rimrock	PRSV	0.555	0.027
Reserve	Shipley Skinner	ToCV	0.555	0.048
Reserve	Motte + Shipley	CABYV	0.812	0.013
Plant species	CP	CYSDV	0.722	0.009
Plant species	DW	ToCV	0.692	0.002
Plant species	DW	ToNDV	0.616	0.009

Table 2.3. *Indicator species analysis values and P-values; only significant (alpha >0.05) species are reported. Two separate analyses were run: by reserve and by plant species. Both analyses are reported in this table.*

The dbMEMs tests for whether distance between communities (latitude, longitude) drives differences in communities. The plotted correlogram shows one positive spatial correlation and evidence for one spatial variable – "1MEM". After running an RDA as a forward selection model, and adjusting the R^2 value, this test shows that 20% of the variation in these virus communities comes from the spatial variable.

Select virus prevalence among host populations within each site

Diversity metrics and indicator species analyses suggested that CABYV and CYSDV are key virus species in target host populations. Targeted analyses of more individuals from each population therefore focused on these two virus species. In 2018-2020, CABYV was detected in all field sites surveyed (Fig. 2.5, Table 2.4). In 2019 and 2020, CYSDV was detected at Anza-Borrego and to a lesser extent at Motte Rimrock, but was not detected in

plants from Shipley Skinner reserve (Fig. 2.5, Table 2.4). At Anza-Borrego, coinfection between both viruses was more common than single infections for both cucurbit species, but not for *D. wrightii*. At Motte Rimrock, where both viruses are present, CYSDV only occurred in co-infection with CABYV, which is highly prevalent. One *D. wrightii* plant was sampled at Motte Rimrock in 2020, and it was also infected with CYSDV. The distributions of plants by year, plant species, and virus infection are shown in Figures S2.1-S2.5 (Figure S2.1 Anza-Borrego, Figure S2.2 Motte Rimrock, Figure S2.3 Shipley Skinner 2018, Figure S2.4 Shipley Skinner 2019, Figure S2.5 Shipley Skinner 2020.

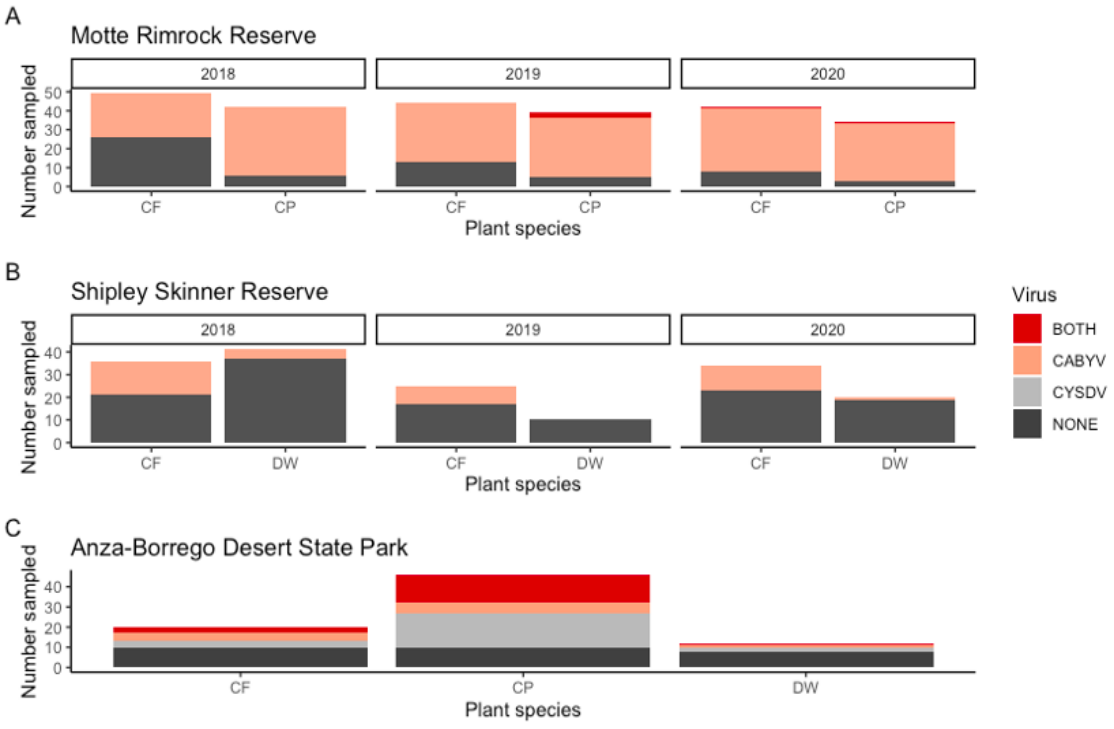

Figure 2.5. Numbers of virus detections at each reserve by host species. *CYSDV was only targeted for detections in 2019 and 2020 following discovery through HTS in Anza-Borrego samples. Anza-Borrego was sampled during 2019 only.*

	Anza Borrego			Motte Rimrock			Shipley Skinner		
2018	*C. foetidissima*	*C. palmata*	*D. wrightii*	*C. foetidissima*	*C. palmata*	*D. wrightii*	*C. foetidissima*	*C. palmata*	*D. wrightii*
Coinfection				na	na	na	na	na	na
CABYV				0.47	0.86	na	0.42	na	0.1
CYSDV				Untested	Untested	na	Untested	na	Untested
Neither				0.53	0.14	na	0.58	na	0.9
N sampled				49	42	0	36	0	41
2019	*C. foetidissima*	*C. palmata*	*D. wrightii*	*C. foetidissima*	*C. palmata*	*D. wrightii*	*C. foetidissima*	*C. palmata*	*D. wrightii*
Coinfection	0.15	0.3	0.08	9	0.08	na	0	na	0
CABYV	0.2	0.11	0.08	0.7	0.79	na	0.32	na	0
CYSDV	0.15	0.37	0.17	0	0	na	0	na	0
Neither	0.5	0.22	0.67	0.3	0.13	na	0.68	na	1
N sampled	20	46	12	44	39	0	25	0	10
2020	*C. foetidissima*	*C. palmata*	*D. wrightii*	*C. foetidissima*	*C. palmata*	*D. wrightii*	*C. foetidissima*	*C. palmata*	*D. wrightii*
Coinfection				0.02	0.03	0	0	na	0
CABYV				0.79	0.88	0	0.32	na	0.05
CYSDV				0	0	1	0	na	0
Neither				0.19	0.09	0	0.68	na	0.95
N sampled				42	34	1	34	0	20

Table 2.4*. Numbers of hosts testing positive for CABYV and CYSDV at each reserve by year and plant species.*

Out of the total surveyed plants, a subset (N=55) was repeatedly sampled over multiple seasons to determine year-to-year retention of the most prevalent virus (CABYV) at Motte Rimrock Reserve. In 2019, 20/29 *C. foetidissima* plants tested positive for CABYV. In 2020, 26/29 were positive (Fig. 2.6). In 2019, 22/26 *C. palmata* plants tested positive for CABYV. In 2020, that increased to 24/26 plants with positive detections (Fig. 2.6). Using a Z test of two proportions, we found that both species did not significantly vary in proportion of infections between the years when significance is $p<0.05$ (*C. foetidissima*: z score -1.9449 and $p = 0.05238$; *C. palmata*: z score -0.8681 and $p = 0.3843$). Overall, eight plants that did not have detectable CABYV in 2018 tested positive in re-sampling years, while only two had undetectable virus after testing positive in 2018. For CYSDV, one of the plants that tested positive in 2019 was also positive in 2020.

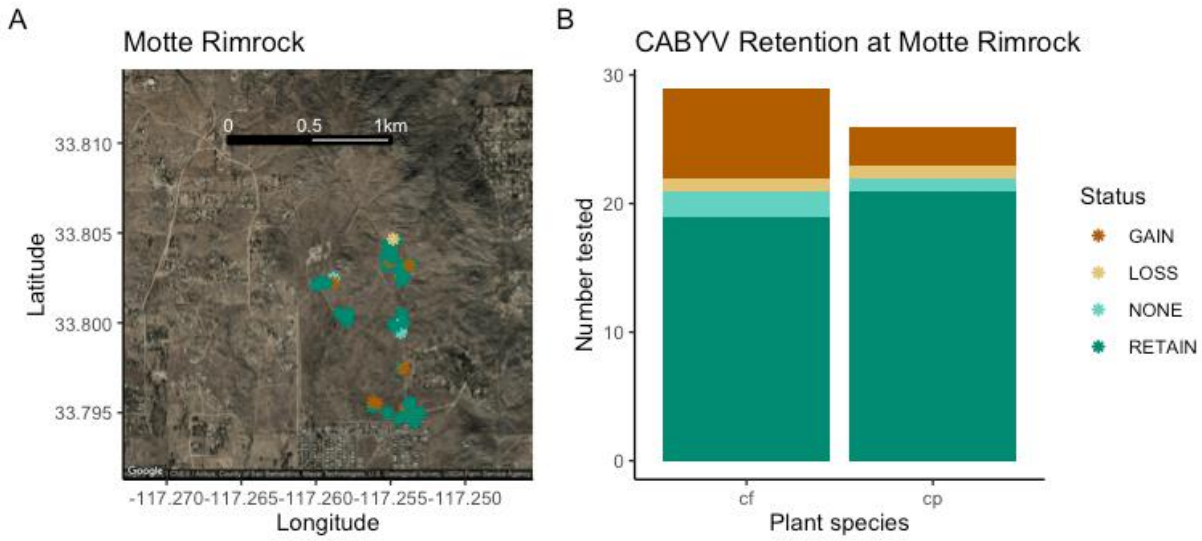

Figure 2.6. *A) Map of samples tested for CABYV retention in 2019 and 2020. B) The number of plants that gained, retained, or lost CABYV detectability. Those without detections both years are named "NONE" for no detection either year.*

Land cover relationship to CABYV and CYSDV infection prevalence

Land cover type, proportion of land of each cover type, and distance of land cover types from each population were not significant predictors of CABYV or CYSDV detections. The land covers included were open water, developed open space and developed low, medium, high intensity land, barren land, evergreen shrub, mixed forest, shrub scrub, grassland herbaceous, pasture/hay, cultivated crops, woody wetlands, and emergent herbaceous wetlands (NLCD 2019 Legend). Comparisons of land cover proportions between reserves are shown in Figures S2.6-S2.9.

For CABYV models using the 2019 dataset and 2018-2020 dataset (excluding Anza-Borrego because it was only sampled in 2019), all three plant species pair-

comparisons were significant predictors of CABYV. For the 2018-2020 dataset focusing on Motte Rimrock and Shipley Skinner Multispecies reserves, year sampled is also a significant predictor of CABYV infection. The model for CYSDV between 2019-2020, which excludes Anza-Borrego, does not have any significant differences between possible factors. For CABYV, Motte Rimrock Reserve is significantly different from Anza-Borrego and Shipley Skinner Multispecies Reserve, but the latter two reserves do not differ from each other. For CYSDV, Anza-Borrego and Motte Rimrock are significantly different. Shipley Skinner reserve likely did not show as significant because there were no CYSDV detections at that reserve. The list of significant contrasts are in Supplementary Table 2.1.

Discussion

Plant viromes are understudied as communities in wild systems. Most studies of virus infections in wild plants seek to identify species and assess some taxonomic (alpha and gamma) diversity. However, each assemblage of virus species within a host is a community itself and can be studied from this perspective. Here, we used tools from community ecology to determine drivers of virome composition in three perennial, drought-adapted plants endemic to the Southern California chaparral environment. We hypothesized that (i) target hosts would each support multiple infections by crop-associated and potentially novel viruses, (ii) virus communities would differ by reserve, with smaller, more fragmented reserves having more crop-associated virus diversity, (iii) land cover (surrounding habitat near the reserves) would influence the prevalence of key crop virus species, and (iv) hosts would retain infections across seasons.

Working with the viromes of 37 plants, belonging to three species, from three reserves in southern California (Fig. 2.1), we identified infections by 12 crop-associated viruses. We confirmed that target hosts support multiple infections in the same individual, consistent with previous studies (Shates et al., 2019). At the community level, the largest, most intact reserve was distinct from the two smaller, more fragmented reserves, and had more crop virus diversity (Table 3, Fig. 2.4). We found that CYSDV and CABYV are key virus species delineating sites. CYSDV was largely found in Anza-Borrego reserve, so it is supported as an indicator species (Table 2.4). In turn, the patterns of CYSDV and CABYV prevalence are structured by plant species and reserve, but not surrounding landscape cover metrics. The virus communities at Anza-Borrego were distinct from Shipley Skinner Multispecies and Motte Rimrock reserves. However, for differences in virus prevalence, Motte Rimrock Reserve was distinct from the other two preserves, likely because of the high prevalence of CABYV at this reserve. This contrasts with our third hypothesis and suggests that organismal-level mechanisms, such as cumulative exposure to pathogens over multiple seasons, may play a larger role than habitat in structuring virus communities of perennials. After resampling a subset of the same plants at Motte Rimrock across two years, we found that CABYV is retained, supporting the idea that viruses can be accumulated over time.

We found no obvious reduction in crop virus diversity in hosts in a larger, more intact habitat. In fact, we found the opposite trend. The land cover proportions, and distance from each plant population, did not have a significant effect on the prevalence of two of the most common crop viruses, CABYV and CYSDV. This contrasts with other

works reporting that distance from agriculture and managed landscapes influenced virus communities (Bernardo et al., 2018; Ingwell et al., 2017) and virus prevalence (Rodelo-Urrego et al., 2013). Our results may be influenced by the number of plants we sampled – 81 from Anza-Borrego in 2019, 249 from Motte Rimrock Reserve, and 172 from Shipley Skinner Multispecies Reserve across three years. There also may be climatic differences because of the larger distance between reserves; the distance-based Moran's Eigenvector Maps (dbMEM) found that space between reserves contributes to 20% of the community variation. There may be effects of land cover/use on virus communities and virus prevalence beyond the one-to-five-kilometer buffer zone we included. For example, Anza-Borrego may have long-distance insect migrants (~20-50km, or greater) from Imperial Valley agriculture and Coachella Valley agriculture. For Anza-Borrego, our analysis suggests CYSDV is an indicator species. CYSDV, which is transmitted by whiteflies, emerged in Imperial Valley in 2006 (Wintermantel et al., 2017) and has spread throughout this region in the ensuing years, infecting melon crops, alfalfa, and common weeds (Wintermantel et al., 2017). More recently, the whitefly-transmitted *Squash vein yellowing virus* (SqVYV) emerged in Imperial Valley (Batuman et al., 2015), also infecting melon crops. In our virome surveys, Anza-Borrego was the only site where SqVYV was detected. The appearance of both CYSDV and SqVYV in Anza-Borrego cucurbit populations, along with our observations of whiteflies in these populations, suggests this location could be experiencing incursions of crop-associated viruses from Imperial Valley agriculture despite it being a large, intact natural area. We also can't rule out the possibility that CYSDV established first in wild *Cucurbita* hosts in Anza-Borrego

before moving into cultivated cucurbits growing in the Imperial Valley (Wintermantel et al., 2009; Kuo et al., 2007). CYSDV was first detected and described in Europe (Spain) and the first appearance in the US was in the year 2000 in Southern Texas (Celix et al., 1996; Kao et al., 2007). By 2006 it was present in melon crops around Niland, California (just on the border with Arizona) and in Yuma, Arizona. In the ensuing years, CYSDV occurrence was documented throughout the Imperial Valley. Recently, it was documented in the melon-growing areas of the Central Valley agriculture production area in California (Mondal et al., 2021). Although these reports are not a complete picture, the timeline does suggest a westward expansion from Texas to Arizona and the western Imperial Valley, followed by proliferation and establishment in this area and a subsequent a northwestern expansion to Central California. This trajectory would include Anza-Borrego, as well as our other sites, with Anza-Borrego being closest in proximity to the initially affected areas of the Imperial Valley.

Host species identity also structured virus communities. The two cucurbit species were similar to each other but distinct from *D. wrightii* based on the PERMANOVA and RDA (Table 3, Fig. 2.4). *Cucurbita palmata* also tended to host more viruses (Fig. 2.2) and was more often infected by either of the two key viruses (CABYV/CYSDV) alone and as coinfections. The host species identity is also a significant factor for virus prevalence. Within Anza-Borrego, which has multiple populations separated by greater than three kilometers between samples (based on cucurbit outcrossing distance), there are differences in virus communities. These differences are likely driven by host plants present. Some are mixed with all three species and some only have *C. palmata*, which

seems to host co-infections more often. Virus prevalence between populations has been reported to vary widely (Prendeville et al., 2012), with host species identity and abundance being implicated as factors underlying virus prevalence in virus ecology studies focused on mixed annual/perennial grasslands (Ingwell et al., 2017; Rodríguez-Nevado et al., 2019; Rúa et al., 2014; Seabloom et al., 2010) and riparian forest communities (Ingwell et al., 2017; Rodríguez-Nevado et al., 2019; Rúa et al., 2014; Seabloom et al., 2010). Our study extends these findings to a new habitat type, desert chaparral, which is dominated by drought-tolerant perennials.

The importance of host species in our study, even above and beyond landscape features, is probably due, in part, to the perennial lifestyle of our target hosts. Ingwell et al. (2017) found that *Barley yellow dwarf virus* and *Cereal yellow dwarf virus* were more prevalent in perennial grasses than annual grasses. Perennial plants are exposed to vectors and the pathogens they transmit over many seasons, which may result in greater potential to accumulate viruses. We found that between 2019 and 2020, CABYV infections are retained across seasons in both *C. palmata* and *C. foetidissima* (Fig. 2.4). Also, roots from Motte Rimrock Reserve taken during the winter season and grown in the greenhouse were infected with CABYV that was subsequently transmissible to squash and melon plants. In one site (Motte Rimrock) we monitored vector activity over multiple seasons. These efforts revealed that vector dynamics are highly variable across years (Supplementary Table 2.1); some years we observed large infestations and other years certain insect groups were difficult to find at all. However, since viruses may persist between years with high and low vector pressure, secondary spread can still occur once a

virus is established in a perennial host. There is also a potential for facilitation or synergistic effects. For instance, infection by one virus may predispose a plant to being infected by additional viruses. This may be the case for the two most prevalent cucurbit infecting pathogens (CYSDV and CABYV). Even though they are transmitted by different insects, the co-infection rates were as high as 15% in *C. foetidissima* at Anza-Borrego (Table 2.5). CYSDV is transmitted by *B. tabaci* (Wintermantel et al., 2017) and CABYV by *A. gossypii* (Lecoq et al., 1992), and both cucurbits targeted in this study appear to be suitable hosts for these insects. Both whiteflies and aphids are attracted to yellowing symptoms (Chesnais et al., 2021; Kenney et al., 2020; Schröder et al., 2017), and we have observed yellowing in plants with CABYV detections (Fig. 2.7). If whiteflies were attracted to plants with CABYV infections, this may explain why co-infections between CYSDV and CABYV are common in the target plant populations. Future studies could explore this through behavioral experiments assessing whitefly attraction to CABYV-infected hosts.

Figure 2.7*: Photograph of C. foetidissima with signs of yellowing. The plant shown here (located in the Shipley Skinner Multispecies Reserve) tested positive for CABYV infection after RNA extraction, reverse transcription and PCR using targeted primers for the virus coat protein.*

Overall, we found that virus infections are common in the target hosts sampled. We found support for spatial structuring of virus communities by location, and by population within each reserve, but did not find significant connections to land type in the 5km perimeter around reserves. The most prevalent virus in target host populations, CABYV, was also an indicator species for Motte Rimrock Reserve; an indicator species is one whose presence, absence or abundance reflects a specific environmental condition. Host species identity is also a key driver of virus infections; Motte Rimrock Reserve and Anza-Borrego both have *C. palmata* and more infections, whereas Shipley Skinner

Multispecies Reserve does not have any observed *C. palmata* and is also the reserve with the least number of viruses recorded. Even though wild plants have been regarded as asymptomatic carriers of crop viruses (Hasiów-Jaroszewska et al., 2021), we found that symptoms are apparent in some individuals. And at least one virus (CABYV) is often retained over multiple years. This might be the case for CYSDV, as well; one of the subset plants for the retention study from Motte Rimrock had a positive CYSDV detection both years it was sampled. Infections by the viruses detected in our study may be affecting plant survival, fitness, or persistence (Chapter Four). Both viruses were introduced to the United States within the last few decades, and now infect native, key perennial plants in southern California reserves. This important finding demonstrates that pathogen introductions have consequences beyond negative effects on agriculture through spillover, accumulation, and spread within native plant communities. Establishing that these viruses have negative effects on wild hosts following spillover is outside the scope of this study, but in Chapter Four we use greenhouse experiments to determine effects of virus infection on hosts and the insect vector.

References

Alexander, H. M., Bruns, E., Schebor, H., & Malmstrom, C. M. (2017). Crop-associated virus infection in a native perennial grass: reduction in plant fitness and dynamic patterns of virus detection. *The Journal of Ecology*, *105*(4), 1021–1031.

Alexander, H. M., Mauck, K. E., Whitfield, A. E., Garrett, K. A., & Malmstrom, C. M. (2014). Plant-virus interactions and the agro-ecological interface. *European Journal of Plant Pathology*, *138*(3), 529–547.

Angelella, G. M., Holland, J. D., & Kaplan, I. (2016). Landscape composition is more important than local management for crop virus–insect vector interactions. *Agriculture, Ecosystems & Environment*, *233*, 253–261.

Bar-Massada, A., Radeloff, V. C., & Stewart, S. I. (2014). Biotic and abiotic effects of human settlements in the wildland–urban interface. *Bioscience*, *64*(5), 429–437.

Batuman, O., Natwick, E. T., Wintermantel, W. M., Tian, T., McCreight, J. D., Hladky, L. L., & Gilbertson, R. L. (2015). First report of an Ipomovirus infecting cucurbits in the Imperial Valley of California. *Plant Disease*, *99*(7), 1042–1042.

Bates D, Maechler M, Bolker B. 2013. lme4: Linear mixed-effects models using S4 classes. *in*. http://CRAN.R-project.org/package=lme4

Bernardo, P., Charles-Dominique, T., Barakat, M., Ortet, P., Fernandez, E., Filloux, D., Hartnady, P., Rebelo, T. A., Cousins, S. R., Mesleard, F., Cohez, D., Yavercovski, N., Varsani, A., Harkins, G. W., Peterschmitt, M., Malmstrom, C. M., Martin, D. P., & Roumagnac, P. (2018). Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *The ISME Journal*, *12*(1), 173–184.

Bolley, D. S., McCormack, R. H., & Curtis, L. C. (1950). The utilization of the seeds of the wild perennial gourds. *Journal of the American Oil Chemists' Society*, *27*(12), 571–574.

Bye, R. (2001). Solanaceae Potato Family. Part One: Datura L. *Journal of the Arizona-Nevada Academy of Science*, *33*(1), 58–64.

Celix, A., Lopez-Sese, A., Almarza, N., Gomez-Guillamon, M. L., Rodriguez-Cerezo, E., & Others. (1996). Characterization of cucurbit yellow stunting disorder virus, a Bemisia tabaci-transmitted Closterovirus. *Phytopathology*, *86*(12), 1370–1376.

Chesnais, Q., Sun, P., & Mauck, K. E. (2021). Advanced infections by cucurbit yellow stunting disorder virus encourage whitefly vector colonization while discouraging non-

vector aphid competitors. *Journal of Pest Science*. https://doi.org/10.1007/s10340-021-01394-z

De Cáceres, M., Jansen, F., Dell, N. (2020). indicspecies: Relationship Between Species and Groups of Sites. Version 1.7.9. https://cran.r-project.org/web/packages/indicspecies/index.html

Dray, S., Dufour, A., Thioulousse, J., Jombart, T., Pavoine, S., Lobry, J.R., Ollier, S., Borcard, D., Legendre, P., Bougeard, S., Siberchicot, A. (2021A). ade4: Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences. Version 1.18. http://CRAN.R-project.org/package=ade4

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., Wagner, H.H. (2021B). adespatial: Multivariate Multiscale Spatial Analysis. Version 0.14. https://cran.r-project.org/web/packages/adespatial/index.html

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652.

Hasiów-Jaroszewska, B., Boezen, D., & Zwart, M. P. (2021). Metagenomic studies of viruses in weeds and wild plants: a powerful approach to characterise variable virus communities. *Viruses*, *13*(10), 1939.

Dittmer H.J. and Talley B.T. (1964). Gross Morphology of Tap Roots of Desert Cucurbits. *Botanical Gazette*, *125*(2), 121–126.

Fox, J., and Weisberg, S. (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Goslee, S. and Urban D. (2020). ecodist: Dissimilarity-Based Functions for Ecological Analysis. Version 2.0.7. https://cran.r-project.org/web/packages/ecodist/index.html

Ingwell, L. L., Lacroix, C., Rhoades, P. R., Karasev, A. V., & Bosque-Pérez, N. A. (2017). Agroecological and environmental factors influence Barley yellow dwarf viruses in grasslands in the US Pacific Northwest. *Virus Research*, *241*, 185–195.

Jung, M. (2016). LecoS — A python plugin for automated landscape ecology analysis. *Ecological Informatics*, *31*, 18–21.

Kates, H. R. (2019). Pumpkins, Squashes, and Gourds (Cucurbita L.) of North America. In S. L. Greene, K. A. Williams, C. K. Khoury, M. B. Kantar, & L. F. Marek (Eds.), *North American Crop Wild Relatives, Volume 2: Important Species* (pp. 195–224). Springer International Publishing.

Kao, J., Jia, L., Tian, T., Rubio, L., & Falk, B. W. (2000). First Report of Cucurbit Yellow Stunting Disorder Virus (Genus Crinivirus) in North America. *Plant Disease, 84*(1), 101.

Kenney, J. R., Grandmont, M.-E., & Mauck, K. E. (2020). Priming melon fefenses with acibenzolar-*S*-methyl attenuates infections by phylogenetically distinct viruses and diminishes vector preferences for infected hosts. *Viruses*, *12*(3).

Kesanakurti, P., Belton, M., Saeed, H., Rast, H., Boyes, I., & Rott, M. (2016). Screening for plant viruses by next generation sequencing using a modified double strand RNA extraction protocol with an internal amplification control. *Journal of Virological Methods*, *236*, 35–40.

Kirkpatrick, K. J., & Wilson, H. D. (1988). Interspecific gene flow in *Cucurbita*: *C. texana* vs. *C. pepo*. *American Journal of Botany*, *75*(4), 519–527.

Kuo, Y.-W., Rojas, M. R., Gilbertson, R. L., & Wintermantel, W. M. (2007). First Report of Cucurbit yellow stunting disorder virus in California and Arizona, in Association with Cucurbit leaf crumple virus and Squash leaf curl virus. *Plant Disease*, *91*(3), 330.

Lacroix, C., Renner, K., Cole, E., Seabloom, E. W., Borer, E. T., & Malmstrom, C. M. (2016). Methodological guidelines for accurate detection of viruses in wild plant species. *Applied and Environmental Microbiology*, *82*(6), 1966–1975.

Lecoq, H., Bourdin, D., Wipf-Scheibel, C., Bon, M., Lot, H., Lemaire, O., & Herrbach, E. (1992). A new yellowing disease of cucurbits caused by a luteovirus, cucurbit aphid-borne yellows virus. *Plant Pathology*, *41*(6), 749–761.

Lenth R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.3., https://CRAN.R-project.org/package=emmeans

Maclot, F., Candresse, T., Filloux, D., Malmstrom, C. M., Roumagnac, P., van der Vlugt, R., & Massart, S. (2020). Illuminating an ecological blackbox: using high throughput sequencing to characterize the plant virome across scales. *Frontiers in Microbiology*, *11*, 2575.

Malmstrom, C. M., & Alexander, H. M. (2016). Effects of crop viruses on wild plants. *Current Opinion in Virology*, *19*, 30–36.

Malmstrom, C. M., Bigelow, P., Trębicki, P., Busch, A. K., Friel, C., Cole, E., Abdel-Azim, H., Phillippo, C., & Alexander, H. M. (2017). 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,* 172–184.

Martinez Arbizu, P. (2020). pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.4. https://github.com/pmartinezarbizu/pairwiseAdonis

Ma, Y., Marais, A., Lefebvre, M., Theil, S., Svanella-Dumas, L., Faure, C., & Candresse, T. (2019). Phytovirome analysis of wild plant populations: comparison of double-stranded RNA and virion-associated nucleic acid metagenomic approaches. *Journal of Virology, 94*(1).

Mondal, S., Jenkins Hladky, L. L., Fashing, P. L., McCreight, J. D., Turini, T. A., & Wintermantel, W. M. (2021). First report of cucurbit yellow stunting disorder virus and cucurbit chlorotic yellows virus in melon in the Central Valley of California. *Plant Disease*. https://doi.org/10.1094/PDIS-01-21-0184-PDN

Murdoch, D. and Adler, D. (2021). rgl: 3D Visualization Using OpenGL. Version 0.108.3. https://cran.r-project.org/web/packages/rgl/index.html

Oksanen, J.F., Guillaume, B., Friendly, M., Kindt, R., Legendre, P., McGlinn D., Minchin, P.R, O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H. (2020). Vegan: Community Ecology Package. Version 2.7. http://CRAN.R-project.org/package=vegan

Ooms, J., (2021). magick: Advanced Graphics and Image-Processing in R. Version. 2.7.3. https://cran.r-project.org/web/packages/magick/index.html

Prendeville, H. R., Ye, X., Morris, T. J., & Pilson, D. (2012). Virus infections in wild plant populations are both frequent and often unapparent. *American Journal of Botany, 99*(6), 1033–1042.

Ripley, B. (2021).MASS: Support Functions and Datasets for Venables and Ripley's MASS. Version. 7.53. https://cran.r-project.org/web/packages/MASS/index.html

Rodelo-Urrego, M., Pagán, I., González-Jara, P., Betancourt, M., Moreno-Letelier, A., Ayllón, M. A., Fraile, A., Piñero, D., & García-Arenal, F. (2013). Landscape heterogeneity shapes host-parasite interactions and results in apparent plant-virus codivergence. *Molecular Ecology, 22*(8), 2325–2340.

Rodríguez-Nevado, C., G Gavilán, R., & Pagán, I. (2019). Host abundance and identity determine the epidemiology and evolution of a generalist plant virus in a wild ecosystem. *Phytopathology, 110,* 94-105.

Roossinck, M. J. (2010). Lifestyles of plant viruses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1548), 1899–1905.

Rúa, M. A., McCulley, R. L., & Mitchell, C. E. (2014). Climate drivers, host identity and fungal endophyte infection determine virus prevalence in a grassland ecosystem. *The Journal of Ecology*, *102*(3), 690–699.

Schröder, M. L., Glinwood, R., Ignell, R., & Krüger, K. (2017). The role of visual and olfactory plant cues in aphid behaviour and the development of non-persistent virus management strategies. *Arthropod-Plant Interactions*, *11*(1), 1–13.

Seabloom, E. W., Borer, E. T., Mitchell, C. E., & Power, A. G. (2010). Viral diversity and prevalence gradients in North American Pacific Coast grasslands. *Ecology*, *91*(3), 721–732.

Shates, T. M., Sun, P., Malmstrom, C. M., Dominguez, C., & Mauck, K. E. (2019). Addressing research needs in the field of plant virus ecology by defining knowledge gaps and developing wild dicot study systems. *Frontiers in Microbiology*, *9*, 3305.

Simpson, G. (2021). ggvegan: 'ggplot2' Plots for the 'vegan' Package. R package version 0.1. https://rdrr.io/github/gavinsimpson/ggvegan/

Stobbe, A. H., & Roossinck, M. J. (2014). Plant virus metagenomics: what we know and why we need to know more. *Frontiers in Plant Science*, *5*, 150.

Susi, H., Filloux, D., Frilander, M. J., Roumagnac, P., & Laine, A.-L. (2019). Diverse and variable virus communities in wild plant populations revealed by metagenomic tools. *PeerJ*, *7*, e6140.

Van Dam, N. M., & Hare, J. D. (1998). Biological Activity of Datura wrightii Glandular Trichome Exudate Against Manduca Sexta Larvae. *Journal of Chemical Ecology*, *24*(9), 1529–1549.

Wickham, H., Chang, W., Henry, L., Pederson, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D. (2021A). ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. Version 3.3.5. https://cran.r-project.org/web/packages/ggplot2/index.html

Wickham, H., Francois R., Henry, L., Müller, K., (2021B). dplyr: A Grammar of Data Manipulation. Version 1.0.7. https://cran.r-project.org/web/packages/dplyr/index.html

Wintermantel, W. M., Hladky, L. L., & Cortez, A. A. (2009). A New Expanded Host Range of Cucurbit yellow stunting disorder virus Includes Three Agricultural Crops. *Plant Disease*, *93*(7), 685–690.

Wintermantel, W. M., Gilbertson, R. L., Natwick, E. T., & McCreight, J. D. (2017). Emergence and epidemiology of *Cucurbit yellow stunting disorder virus* in the American Desert Southwest, and development of host plant resistance in melon. *Virus Research*, *241*, 213–219.

National Land Cover Database Legend. https://www.mrlc.gov/data/legends/national-land-cover-database-2019-nlcd2019-legend (accessed on 05 November 2021).

QGIS Development Team. QGIS Geographic Information System: Open Source Geospatial Foundation. Available online: http://qgis.osgeo.org (accessed on 01 November 2021).

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Chapter 3. Geographic origins and genetic analyses of Cucurbit aphid-borne yellows virus

Abstract

Cucurbit aphid-borne yellows virus (CABYV, genus *Polerovirus*, family *Luteoviridae*) is a globally distributed pathogen of cucurbits that can cause yellowing symptoms on foliage and significant yield losses. CABYV was first reported in California in 1993, but surprisingly, has not become a major pathogen in agricultural regions in this state . Despite low prevalence in agriculture, we found that CABYV frequently infects perennial, native cucurbits and a co-occurring nightshade in desert and Mediterranean-type ecosystems in Southern California. Here, we used near complete genome sequences generated during virome sequencing of the native plant species to explore CABYV diversity in California, origins and relationships to other poleroviruses globally. We used phylogenetic methods to curate CABYV sequences and clarify relationships among regional groupings and found that the California isolates from wild plants grouped with Asian and Mediterranean isolates, suggesting at least two introductions. Using our sequences and those generated around the globe, we then evaluated evidence for putative novel polerovirus species and recombinants. Our results suggest that 1) intraspecific and interspecific recombination between CABYV and other poleroviruses are common, 2) interspecific recombinant and nonrecombinant poleroviruses are commonly mis-labeled as CABYV in public databases, 3) isolate identification as CABYV requires sequencing at least two open reading frames (ORFs) from either side of an intergenic spacer region

(non-coding region in between ORF2 [5’ side of the genome] and ORF3 [3’ side of the genome]); 4) there are up to two novel polerovirus species previously identified as CABYV, as well as geographic structure in selection on CABYV genes and genomes, based on estimates of genetic diversity within and among CABYV, recombinants, and mislabeled poleroviruses, and 6) the previous hypothesis of Mediterranean origin as well as a new region of origin (northeastern India) is supported by discrete and continuous phylogeographic analyses respectively. Overall, this study provides evidence for multiple introductions of CABYV to the US, an alternative origin for CABYV, and the existence of a CABYV species complex that has long complicated accurate identifications. Our results underscore the benefits of sequencing viromes from wild perennial hosts and indicate that future work on CABYV should focus on more complete sampling of putative origin locations and sequencing strategies that explicitly consider recombination points.

Introduction

Cucurbit aphid-borne yellows virus (CABYV, genus *Polerovirus,* family *Sobemoviridae*) is a globally distributed pathogen of cucurbits that causes yellowing of foliage and economically damaging yield losses. The virus is transmitted in a persistent, circulative, non-propagative manner by the aphids *Aphis gossypii* Glover and *Myzus persicae* (Sulzer) (Lecoq et al. 1992a; Carmo-Sousa et al. 2016). CABYV was first described in 1992 from melon and cucurbit extracts taken from leaves collected from agricultural fields in France (isolate N) (Lecoq et al. 1992b). Additionally, CABYV was detected in preserved leaves from surveys of cultivated cucurbit plots performed in 1982-1984 using DAS-ELISA, which suggests that CABYV was present in France at least 10 years prior to its first description (Lecoq et al. 1992a). In 1993, the second report of CABYV infections in cultivated Cucurbitaceae hosts was published, this time following detections in the most intensive melon production area of California (Central Valley region including the San Joaquin, Sacramento and Salinas valleys) in the United States (U.S) (Lemaire et al. 1993). Since these initial reports, CABYV has emerged as a pathogen causing economically relevant losses to Cucurbitaceae crops across the world (Mnari Hattab et al. 2005; Khanal and Ali 2017; Shates et al. 2019; Minicka et al. 2020). One very notable exception to this is the United States. After the first report of CABYV in California in 1993, there were no subsequent reports of outbreaks until 2017, when CABYV was detected in cucurbits in Oklahoma (Khanal and Ali 2017) and in wild perennial cucurbits and Solanaceae in California (Shates et al. 2019).

California is the most intensive production area of *Cucurbitaceae* crops in the U.S., especially muskmelons, which are the most economically important hosts for CABYV in other countries (Kassem et al. 2013; Vafaei and Mahmoodi 2017; Kwak et al. 2018). Therefore, it is surprising that the 2017 CABYV report in California was from wild cucurbit species in sage scrub communities (*Cucurbita palmata* Wats. and *Cucurbita foetidissima* Kunth) and a previously unreported solanaceous host that frequently co-occurs with perennial cucurbits (*Datura wrightii* Regel) (Shates et al. 2019). The fact that CABYV is not being detected in crops, but is being detected in wild hosts across two families, suggests that CABYV isolates circulating in wild plant communities may be unique relative to variants present in crops outside of the US. To determine whether this is the case, and to shed light on the origins of CABYV in the U.S., we sampled *C. palmata, C. foetidissima* and *D. wrightii* throughout southern California and used next-generation sequencing to obtain nearly complete genomes of CABYV from each host and location. The assembly of 17 new genomes substantially increases the number of previously available full genomes for CABYV from around the world and provides all but one of the sequences available from the US. We used this newly acquired genomic information to clarify the origins and genetic features of CABYV in the US while also advancing our understanding of CABYV diversity, origins, and relationships to other poleroviruses at a global scale.

Evidence to date suggests that the Mediterranean region is the probable origin for CABYV. This hypothesis is based on the earliest report of CABYV being from southern France (Lecoq et al. 1992b), and molecular clock analyses of the CABYV P0 gene

suggesting Spain as a possible location of origin (Lecoq et al. 1992a; Costa et al. 2019). Phylogenetic hypotheses for CABYV published over the past two decades further suggest that CABYV isolates can be grouped into three broad regions: Mediterranean (including Southern Europe and North Africa), Asian (including China, South Korea, and Japan), and Taiwanese (Xiang et al. 2008; Omar and Bagdady 2012; Kwak et al. 2018). However, to our knowledge, only one formal phylogeographic study has attempted to pinpoint CABYV origins (Costa et al. 2019). This study focused on the first open reading frame (ORF0, P0), and included 37 isolates from six countries. Results suggest that CABYV likely originated within Europe, specifically in Spain, with a most recent common ancestor 125.5 years ago, consistent with the broader diversification of *Luteoviridae* over the last 500 years. ORF0 encodes for the P0 protein, which is a viral suppressor of RNA silencing involved in countering antiviral defenses of the host. Because of this, P0 may evolve at a faster rate than more conserved elements, such as the coat protein genes which are under purifying selection due to complex interactions with the vector (Chare and Holmes 2004). ORF0 is also located on the 5' end of the genome, upstream of an intergenic spacer region (between ORF2 and ORF3) that is a key recombination point for poleroviruses. Thus, in the absence of additional sequences from the same isolate spanning the intergenic region, it is not possible to tell if an ORF0 sequence represents a P0 gene from a CABYV isolate *or* a P0 gene associated with isolates of a viable recombinant virus. Given the role of host environment in shaping P0 evolution, and different host ranges of recombinant vs. parental viruses, it is reasonable to conclude that robust phylogeographic studies on CABYV origins should instead focus on

genomic regions that permit concrete identification of an isolate as truly CABYV or recombinant.

To address this, we leveraged the additional CABYV genome sequences generated from our collections of wild squash populations in California, as well as recently published CABYV genomes to perform updated phylogeographic analyses, ancestral trait reconstruction, and analysis of molecular variance on validated CABYV sequences (ORF1-2 and ORF3 from 101 isolates and near complete genomes of 54 isolates). We also compared a traditional, discrete phylogeography approach to a new continuous phylogeography approach (Dellicour et al. 2021). Discrete phylogeography is limited by potentially artificial groupings of isolates from different locations. However, with continuous phylogeography, the model tests a correlational basis of genetic distances, groupings, and different origin points. This provides a realistic alternative for inferred dispersal history and, importantly, includes inferred unsampled internal points that are likely to be important for reconstructing dispersal history (Dellicour et al. 2021). Ultimately, this can inform future sampling, or efforts to obtain sequences from previously collected material (e.g., herbarium specimens).

In addition to producing a more comprehensive and robust phylogeographic analysis of CABYV that considers validation of virus identity, we also sought to clarify relationships to closely related poleroviruses and address the poor curation of sequence data for the cucurbit polerovirus complex. Since the first description (Lecoq et al. 1992b) and publication of genetic sequence data for CABYV (Guilley et al. 1994), three closely-related, cucurbit-infecting *Polerovirus* species have been described: *Melon aphid-borne*

yellows virus (MABYV) first documented from China (Xiang et al. 2008), *Suakwa aphid-borne yellows virus* (SABYV) first documented from China (Shang et al. 2009), and *Pepo aphid-borne yellows virus* (PABYV) first documented in Mali, with further detections in other African countries and in Greece (Knierim et al. 2014; Ibaba et al. 2017). Recombination is one mechanism by which new virus species arise and is particularly common in the family *Luteoviridae* (Pagán and Holmes 2010). MABYV may be a recombinant of CABYV and SABYV or corresponding ancestral viruses (Shang et al. 2009), and PABYV is likely a recombinant between CABYV and SABYV (Ibaba et al. 2017). CABYV recombination can even lead to drastic changes in virus-vector relationships. For example, one recombinant between CABYV and an unknown virus, isolated in Brazil, is the first whitefly-transmitted virus described for the family Luteoviridae (Costa et al. 2019, 2020). Additionally, recombination between co-infecting CABYV and MABYV has produced at least one stable recombinant (isolates designated as the "R group") that is currently recognized as CABYV in the literature (Knierim et al. 2010; Vafaei and Mahmoodi 2017; Kwak et al. 2018). The CABYV-MABYV recombinant encodes a coat protein with amino acids that are >90% similar to CABYV, and infections by this recombinant virus provide cross-protection against CABYV, but apparently not against MABYV (Knierim et al. 2010). Because CABYV and MABYV overlap geographically, and are genetically similar, they can be easily misidentified when sequence-based diagnostics use only partial genomic characteristics. Misidentification of sequences from isolates of stable, infectious recombinants of CABYV plus another parental virus can further complicate virus identifications based on partial sequences. The

result is a quagmire of misidentified (and poorly curated) accessions in the public NCBI GenBank sequence repository. When such poorly curated sequences are allowed to persist in public databases without correction, their impact is perpetuated, as researchers will base identifications and phylogenetic inferences on analyses that include incorrect information. Therefore, in addition to the phylogenetic analyses discussed above, we collected meta-data on publicly available CABYV sequences and used phylogenetics to parse misidentified sequences and provide a curated database for researchers.

Methods

Description of CABYV

Like other members of the genus *Polerovirus*, CABYV has a single-stranded, positive sense RNA genome of approximately 5.7 kb nucleotides, and six open reading frames (ORFs) (Fomitcheva et al. 2004). The six ORFs are separated by an intergenic, non-coding region of ~200 nucleotides, with ORFs 0-2 on the 5' side and ORFs 3-5 on the 3' side. Each ORF encodes a protein for which putative functions have been proposed. The P0 protein encoded by ORF0 is hypothesized to be a suppressor of post-transcriptional gene silencing and host specific responses (Pfeffer et al. 2002; Prüfer et al. 2006). ORF1 and 2 encode the ribosomal frameshift protein P1-2, which contains characteristic motifs of an RNA-dependent RNA polymerase (RdRp). Viral RdRps perform multiple functions, including genome replication and facilitating interactions between virus and host cell components (Guilley et al. 1994; Mayo and Miller 1999; Shen et al. 2020). ORF 3 encodes the coat protein (P3) which is involved in aphid transmission (Brault et al.

2005), ORF 4 encodes the movement protein (P4) for cell-to-cell movement (Schmitz et al. 1997); and ORF3-5 encodes P3-5, a read-through protein involved in aphid transmission and vector specificity (Mayo and Miller 1999; Brault et al. 2005).

Tissue collection from target hosts

We collected plant tissues from three perennial hosts that are known to support CABYV infections (*C. foetidissima*, *C. palmata*, and *D. wrightii*) and that are key species in California sage scrub habitat (Shates et al. 2019). Tissues were sampled from individuals growing in three reserves in southern California: Motte Rimrock Reserve (DOI 10.21973/N31T0W) (August 2017), Shipley Skinner Multispecies Reserve (August 2017), and Anza-Borrego Desert State Park (Steele/Burnand Anza-Borrego Desert Research Center DOI 10.21973/N3Q94F) (June 2019) (Fig. 3.1). Target plants were selected randomly and without *a priori* knowledge of virus infections or regard to symptoms. We collected both leaf and stem tissue from each plant by inverting a clean plastic Ziplock bag over the tissue, pulling with force, then sealing the bag. Approximately 10 grams of leaf tissue were collected from each individual and stored on dry ice until return to the laboratory. Four grams of leaf tissue from each sample were carefully partitioned into 50-mL RNase-free Falcon tubes and stored at −80°C in preparation for dsRNA extraction. Remaining tissue (~6g) was archived at −80°C.

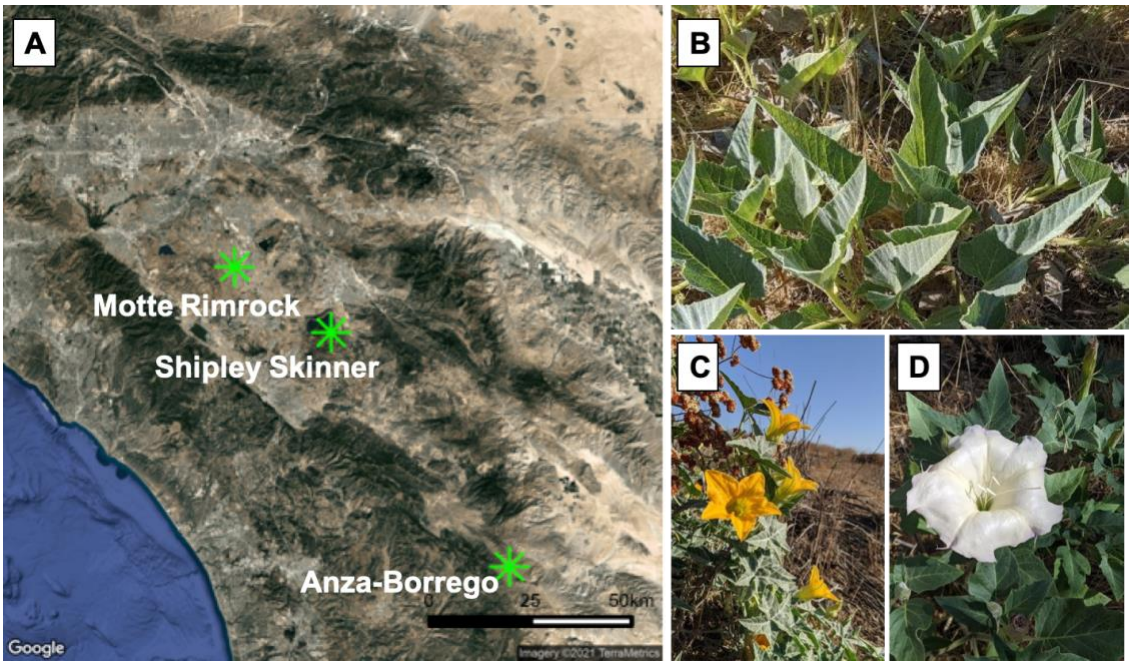

Figure 3.1*: A) Map of field sites in Riverside County and surrounding areas in Southern California made with R and the ggmap package (Kahle and Wickham 2013). B) Wild host Cucurbita foetidissima from which we have ten sequences of CABYV. C) Wild host Cucurbita palmata from which we have six sequences. D) Wild host Datura wrightii from which we have one sequence of CABYV.*

Nucleic acid extraction and next-generation sequencing

We extracted double-stranded RNA (dsRNA) for non-targeted sequencing as described in (Shates et al. 2019), which was originally adapted from (Kesanakurti et al. 2016; Ma et al. 2019). To assess the quality of extractions and sequencing, each sample received a uniformly sized leaf punch of bell pepper (Capsicum annum L. cv. California Wonder, which contains the vertically-transmitted, dsRNA virus, *Bell pepper endornavirus* (BPEV, *Endornaviridae*). Before library preparation, we performed reverse transcription-PCR for BPEV to evaluate dsRNA suitability for sequencing. For samples passing this quality control check, we then denatured an aliquot of 5 µl of dsRNA solution from each

sample by incubating at 99°C for 5 min, then placed the solution on ice and prepared libraries using the NEBNextR UltraTM II Directional RNA Library Prep Kit for Illumina, following the manufacturer protocol (average insert size of 250–300 bp). Library checks with an Agilent 2100 Bioanalyzer and quantitative PCR were performed by staff at the UCR Genomics Core. Sequencing was performed at the University of California Riverside (UCR) Genomics Core on the Illumina NextSeq platform using a Mid Output flow cell (v2) with a paired-end 75 bp read-length, which produces approximately 260M paired-end reads. Adaptors were trimmed by Core staff, then we processed the output using a workflow we have previously employed for wild plant virome analysis (Shates et al. 2019). Briefly, this workflow uses the Galaxy Platform to filter out host genomes, perform de novo assembly using Trinity, and use nucleotide BLAST to determine putative virus identities (Shates et al. 2019; Afgan et al. 2018). Positive detections of CABYV in hosts from Motte Rimrock (eight) and Shipley Skinner reserves (five) were published previously along with partial genomic information for these isolates (Shates et al. 2019; sequences in read archive of Genbank accession SRP149013). CABYV isolates from Anza-Borrego Desert State Park (four) are reported here for the first time. For the present study, CABYV genomes from all California isolates were assembled by aligning contigs from each host against a reference genome using Clustal Omega (Goujon et al. 2010; Sievers et al. 2011), then manually removing regions of overlap.

Acquisition and curation of polerovirus sequences for phylogenetic analysis

All full genomes, coat protein, and RNA-dependent RNA polymerase (RdRp) sequences for CABYV (detailed below) were downloaded from NCBI GenBank as fasta files. Sequences smaller than 500 base pairs were not included in subsequent analyses because this is a small fragment compared to the whole genome, although an acceptable size for a single gene. If necessary, coat proteins and RdRp sequences were trimmed from full length genomes to be included in phylogenetic analyses focusing on these specific regions (detailed below). This curation effort is phylogeny-based, with methods described in the next section. The final list of accessions used for this study is assembled in Table S1, which includes the following meta-data: host information, collection year (if given), country, region, isolate name, given identity (to denote the identification given on GenBank), and true identity (to denote identification errors). Host information was recorded to document the range of curation quality but was not used for analyses because of ambiguities in host identity (e.g., use of common names covering multiple species). These ambiguities are documented in Supplementary Table 3.1, and their impact on inferences described in the results and discussion. For accessions without a clear collection date, the earliest year associated with the accession sequence was recorded. Sequences for outgroup taxa also were downloaded from GenBank: *Melon aphid-borne yellows virus* (MABYV), *Pepo aphid-borne yellows virus* (PABYV), and *Suakwa aphid-borne yellows virus* (SABYV). The original accessions submitted for MABYV in 2009 (EU091150, EU091151, and EU091149 (Xiang et al. 2008) were labelled as CABYV on GenBank. In Table S1 under the "true identity" column we designated these sequences as

MABYV. Sequences from a recombinant CABYV from Brazil (Costa et al. 2019, 2020) later proposed to be a new virus, cucurbit whitefly-borne yellows virus (Costa et al. 2019, 2020), were also assigned the "given identity" of CABYV (accessions LC217993, LC516688, LC217994) and were not included in analyses detailed below based on putative new species status. Sequences recognized by contributors as CABYV "R groups" (a recombinant between CABYV and MABYV) were removed from phylogeography (discrete and continuous) analyses and analysis of molecular variance (AMOVA). Further curation of the CABYV sequence database (MABYV or SABYV sequences that were incorrectly labeled as CABYV) was completed using phylogenetic methods described below.

Phylogenetic analysis and species assignment

We used phylogenetics to verify species identities and explore regional groupings of isolates. Using Qiagen's CLC Main Workbench (https://digitalinsights.qiagen.com) we aligned nucleotide sequences, translated California isolate ORFs to corresponding amino acid sequences, and aligned amino sequences. To perform model testing to prepare datasets for analyses, we used ModelFinder Plus (Kalyaanamoorthy et al. 2017) and 1000 Ultrafast Bootstrap replicates to generate the maximum likelihood phylogeny on the IQTree2 executable (Hoang et al. 2018). The nucleotide sequences of the full genome, coat protein, RdRp and concatenated coat protein + RdRp were all assigned TIM2 model variants by the Bayesian information criterion (BIC) for model selection, but we used General time reversible with topography and rate variation (GTR+G+T) for phylogeography analyses as the assigned model is not available on BEAST software. The

amino acid sequences of the coat protein and RdRp were assigned the models HIVw+G4 and JTTDCMut+R3. Based on the results of each tree, accessions that did not group with CABYV but with outgroups (documented in Table S1) were removed from phylogeographic analyses and AMOVA. For all phylogenies, we included three outgroups: MABYV, SABYV, and PABYV.

We found that some isolates of CABYV, not classified as RGroup, consistently grouped outside of most CABYV sequences with high bootstrap (>95) support. These included the isolates from Taiwan and island nations of the Pacific Ocean. We therefore removed these from phylogeographic analyses. Accessions removed from whole genome analyses included JQ700305 [Taiwan], KY617826 [EastTimor], LC472499 [Indonesia], and MG780352 [Papua New Guinea]. Accessions removed from CP, RdRp, and concatenated CP+RdRp analyses included the previous accessions as well as five other isolates from Taiwan (GU324100, GU324095, GU324096, GU324098, and GU324103). To illustrate overall patterns while maintaining clarity, we created subset phylogenetic trees; regions with more than five representatives were reduced to five randomly selected representative sequences, except for the California isolates that are described for the first time in this manuscript. Subsetted phylogenies are presented in the main text with full phylogenies included in the Supplemental materials.

Population genomics

To explore the genetic diversity of the cucurbit polerovirus complex, we calculated key population metrics for each amino acid sequence from CABYV, RGroup (recombinant CABYV-MABYV), the "CABYV-like" group encompassing sequences from Taiwan

and other Asia-Pacific locations, and MABYV. We used MegaX (Kumar et al. 2018) to calculate *p*-distances between and within virus species at the amino acid level, as for poleroviruses and luteoviruses, species are delimited by a 10% demarcation cutoff between amino acids of a gene product (Lefkowitz et al. 2018). To ensure all P3 sequences were truly from CABYV, we only selected sequences that had corresponding whole genome or RdRp data for the same isolate. For the same reason, we only selected sequences of ORFs 0, 1-2, 4, and 3-5 from among isolates with fully or nearly fully sequenced genomes. For ORF 1-2, we only used isolates that had the entire gene product available and did not include isolates with only partial sequences.

We used R version 4.0.3 (R Core Team, 2021) to calculate nucleotide and haplotype diversity, as well as Tajima's D and Fu's F for each open reading frame (Tajima, 1989; Fu and Li, 1993). For calculation of nucleotide and haplotype diversity, we used the *pegas* R package (Paradis 2010), and for calculation of Tajima's D and Fu's F we used the *PopGenome* R package (Pfeifer et al. 2014). To infer the type of natural selection acting on each gene, for each ORF we calculated pairwise nonsynonymous (dN) and synonymous (dS) substitutions per site (dN-dS) and a global dN/dS estimate using the Pamilo-Bianchi-Li method (Li, 1993; Pamilo and Bianchi, 1993), respectively, with MEGAX. Here, we only analyzed CABYV, CABYV-like, and RGroup datasets.

Phylogeography

To determine the probable origin location of CABYV, we performed both discrete and continuous phylogeographic analyses. For discrete phylogeographic analyses we used the BEAST package V2.6.3 (Bouckaert et al. 2019) according to the Ancestral

Reconstruction/Discrete Phylogeography with BEAST 2.3x tutorial by Bouckaert and Xie (https://www.beast2.org/tutorials/) with a few modifications: We changed the site model to GTR, the frequencies to "Empirical" and set the prior for the additional location trait ("nonZeroRates.s:location") to exponential. We set the Markov chain Monte Carlo (MCMC) chain length (essentially generation time) to 100 million. Posterior values associated to each location represent support as the most likely origin areas. We performed these tests on full genomes of CABYV, on P1-2 (RdRp), on P3 (CP) that had RdRp associated with it (so we could exclude RGroups), and on a concatenated P1-2, P3 (RdRp+CP) dataset.

To test for the origin and distribution history of CABYV, we performed continuous phylogeographic analyses, using the BEAST package V1.10.4 (Suchard et al. 2018). We followed the methods in Dellicour et al. (2021) to run the analysis separately on CP, CP + RdRp, RdRp, and the full genome sequences. In order to parameterize the SkyGrid analysis for our datasets (specifically - finding the Time since last transition and inferring Number of parameters on the Tree panel in Beauti), we used TempEst (Rambaut et al. 2016) to identify outliers and find the putative root age (x-intercept) to be used as a starting point for the BEAST parameter Time since last transition (Hill and Baele 2019). Based on TempEst residuals, we removed JF939813 (Spain). This sequence may have more or fewer mutations within the sampling time (Hill and Baele 2019). Based on these values and multiple preliminary runs of BEAST to assess the effective sample size (ESS) values, we used 250 as the Number of parameters and 300 as Time since the last transition. The only other change we made to the methods in (Dellicour et al. 2021) was

to use the GTR model instead of HKY. The time since the last transition was roughly corroborated by the likely time to most recent common ancestry calculated in previous work for *Luteoviridae* (Pagán and Holmes 2010). For the location trait partition, we used precise coordinates for our California isolates, but we used the default Google-searched coordinates for the country, county or province, and state as provided on GenBank (unfortunately coordinates are rarely submitted as metadata) (Table S1). For the molecular clock panel, we set the generation time at 200,000,000, with Echo state to screen and Log parameters set to 20,000. The program was run through the High-Performance Computing Center at UC Riverside.

Analysis of molecular variance

To perform the AMOVA on the nucleotide sequences of full genomes, ORF3, ORF1-2, and a ORF 1-2 + intergenic region + ORF3, we used R version 4.0.3 and the package *poppr* (Kamvar et al. 2014, 2015). Population hierarchies were assigned by "broader" or "smaller" regions as reported on GenBank (i.e., USA or western USA). Regions with fewer than three isolate sequences were removed because at least three sequences are required to test for significance of fixation indices. Based on this cutoff, AMOVA on full genomes included sequences from South Korea, western U.S. (our contributed sequences), and Mediterranean. AMOVA analyses on the concatenated coat protein+RdRp and RdRp included 10 regions, and AMOVA on the coat protein included 23 regions. For all analyses we report percent variation, proportion of variation within populations (F_{ST}), proportion of variation between broad groups (F_{CT}), and proportion of variation among populations within groups (F_{SC}).

Recombination analysis

To detect recombination in full genome alignments and concatenated coat protein+RdRp alignments (including PABYV and SABYV) we used the program RDP version 5.5 (Martin et al. 2015). First, we performed a PHI test for overall evidence of recombination, followed by a full exploratory recombination scan using seven methods (RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan and 3Seq) and a Bonferroni-corrected *P*-value cutoff of 0.01. To be conservative and avoid false detection of recombination, only signals with 4/7 supportive tests were included in the results (Rabadán et al. 2021). We excluded recombination events with signals that could have been caused by processes other than recombination. To visualize these recombinants and their relationships with other isolates, we used the program SplitsTree4 (Huson and Bryant 2006) to create split decomposition networks, with 1000 bootstrap replicates. Recombination events are visualized by reticulations in the network.

Results

Sequencing

From the Illumina sequencing reads, we assembled 17 CABYV genomes from the three target hosts (Supplementary Table 3.1). Two complete genomes were recovered as individual contigs from a single host without manually assembling: CCP2A and CCP2B. The other genomes were all nearly complete, with only one isolate (MRCF2C) missing an ORF3. Genome sizes and alignments are consistent with previously published isolates as far as composition and length (Supplementary Table 3.1).

Phylogenetics, species assignment, and filtering of mislabeled sequences

The California sequences clustered with two of the previously recognized clades - both the Asian group and the Mediterranean group. Overall, the genome phylogeny did not give high bootstrap support for there being three main regions of CABYV diversity (Kwak et al. 2018); the subset full genome phylogeny (Fig. 3.2) and complete full genome phylogeny (Figure S3.6) both have maximum bootstrap support (100) for splitting Taiwan, East Timor, Indonesia and Papua New Guinea sequences from the other CABYV sequences, but no other regional support. This separation was also supported by individual RdRp (ORF1-2) and CP (ORF3) analyses: full and subset CP analyses (CP Fig. 3.3, Figure S3.7) and full and subset RdRp analyses (Figure 3.4 and Figure S3.8) show strong bootstrap support (>95) for separating Taiwan sequences from other CABYV sequences. The RdRp phylogeny also suggests sequences from Thailand and Pakistan are also distinct from those in the Mediterranean and Asia (Fig. 3.4). In contrast, the full and subset concatenated ORF1+ORF2 phylogenies (Fig 3.5 and Figure S3.9) showed high support for the previously proposed regional groups (bootstrap values >95). All phylogenetic analyses suggested that recombinant sequences (designated here as "R group") clustered neither with CABYV, nor with MABYV. Sequences that were not labeled as recombinant isolates, but clustered within the recombinant clade (EU091148, EU244326, HQ439023, MN688219, MN688220, MN843966, MN843967, MN862007), were treated as recombinant isolates in later analyses. Additionally, several sequences identified as CABYV on GenBank clustered instead with MABYV (EF063708, EU262628, DQ973123, and EF063704) or SABYV (EU259784 and EU259785). Correcting these misidentifications expands upon the original accessions for MABYV

and SABYV by adding seven MABYV accessions mislabeled as CABYV and two SABYV accessions mislabeled as CABYV (Supplementary Data Sheet 3.1). We also identified nine recombinants previously labeled as CABYV (Supplementary Data Sheet 3.1).

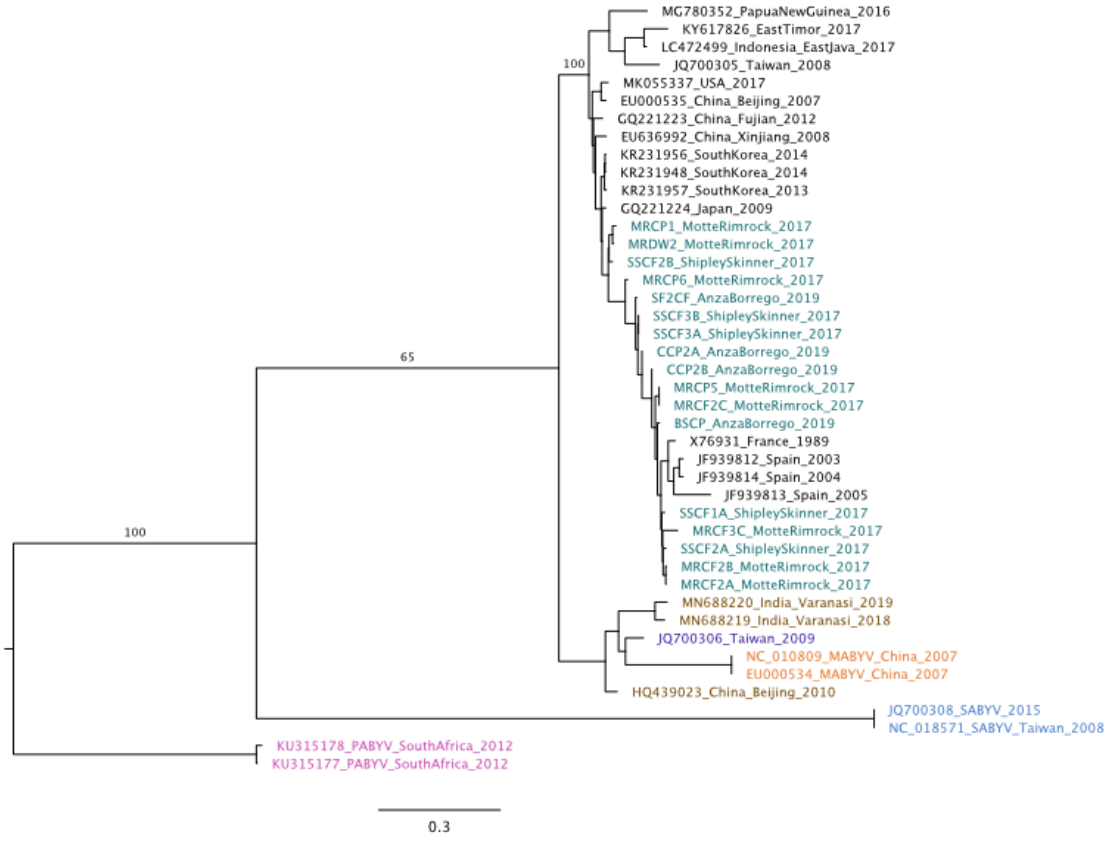

Figure 3.2. *Phylogeny of a subset of nucleotide sequences based on the full genome. For regions with greater than 5 representatives, we randomly kept only 5, except for our own 17 sequences. The color coding is as follows: California sequences are in teal, known RGroup in purple, "mislabeled" in brown, MABYV in orange, PABYV in pink, and SABYV in blue.*

Figure 3.3. *Phylogeny of a subset of nucleotide sequences based on ORF3 (CP). For regions with greater than 5 representatives, we randomly kept only 5, except for our own 17 sequences. The color coding is as follows: California sequences are in teal, known RGroup in purple, "mislabeled" in brown, MABYV in orange, PABYV in pink, and SABYV in blue.*

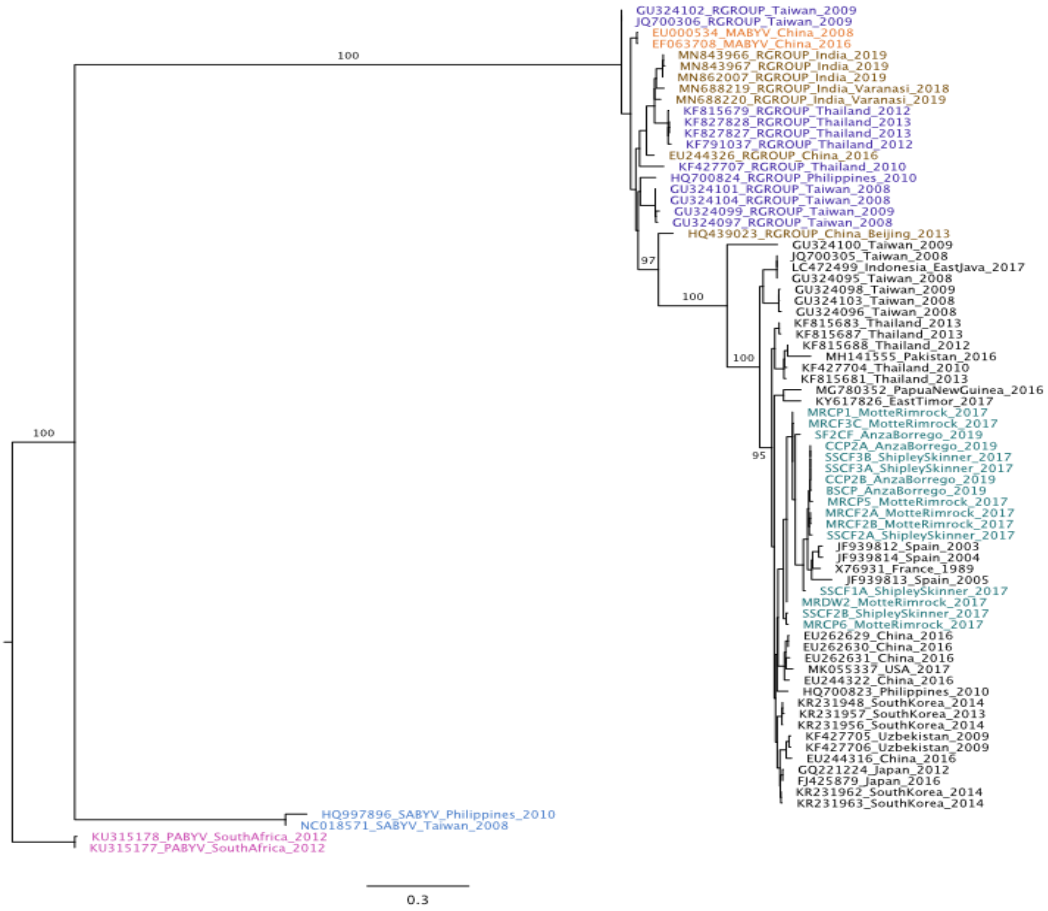

Figure 3.4. *Phylogeny of a subset of nucleotide sequences based on ORF1-2 (RdRp). For regions with greater than 5 representatives, we randomly kept only 5, except for our own 17 sequences. The color coding is as follows: California sequences are in teal, known RGroup in purple, "mislabeled" in brown, MABYV in orange, PABYV in pink, and SABYV in blue.*

Figure 3.5. *Phylogeny of a subset of nucleotide sequences based on ORF1-2+ORF3 (RdRp+CP concatenated). For regions with greater than 5 representatives, we randomly kept only 5, except for our own 17 sequences. The color coding is as follows: California sequences are in teal, known RGroup in purple, "mislabeled" in brown, MABYV in orange, PABYV in pink, and SABYV in blue.*

Population Genomics

Percent variation of amino acid sequences of each protein of each virus species are listed in Table 3.1, and percent differentiation between virus species at the level of individual proteins are reported in Table 3.2. A high intraspecific variation was detected for the CABYV-like group identified in phylogenetic analyses (Taiwan + Asia Pacific

sequences) and for the RGroup based on P0 (both), P4 (CABYV-like group) and P3-5 (Rgroup), suggesting that these groups might include more than one species (Table 3.1). The CABYV-like group differed by more than 10% from confirmed CABYV sequences for P0, P4, and P3-5 proteins (Table 3.2). The CABYV-like group also differed by more than 10% from all RGroup sequences except P3, and all MABYV sequences (Table 3.2). RGroup sequences also differed by more than 10% from CABYV (P0, P1-2, P3-5) and MABYV (all except P1-2) (Table 3.2). This analysis suggests that the CABYV-like group includes isolates of a novel polerovirus distinct from both CABYV and MABYV and that the recombinant virus formed from CABYV and MABYV parents should also be considered a distinct species.

Virus	P0: PTGS	P1-2: RdRp	P3: CP	P4: MP	P3-5: RTP
CABYV	8.61%	4.70%	4.70%	6.22%	5.80%
CABYV-like	19.73%	8.87%	6.23%	11.30%	9.40%
MABYV	4.35%	1.11%	0.65%	0.57%	0.56%
RGROUP	22.87%	9.28%	5.52%	8.81%	11.25%

Table 3.1. *Percent pairwise variation within virus species. (>10% amino acid differences between species).*

Virus pair	P0: PTGS	P1-2: RdRp	P3: CP	P4: MP	P3-5: RTP
CABYV - CABYV-like	20.40%	9.53%	6.49%	11.08%	11.20%
CABYV - RGroup	21.80%	25.71%	7.09%	9.10%	10.50%
CABYV-like - RGroup	24.80%	25.51%	6.49%	11.31%	12.00%
CABYV - MABYV	24.90%	25.98%	23.07%	34.75%	40.40%
CABYV-like - MABYV	26.20%	25.46%	22.39%	31.75%	41.40%
MABYV - RGroup	26.80%	8.84%	21.27%	34.05%	40.20%

Table 3.2. *Percent pairwise variation between species for each amino acid sequence. (>10% amino acid differences between species).*

The RGroup and CABYV-like groups have the highest nucleotide diversity for both P0 and P3-5 ORFs (greater than or equal 0.10) (Table 3.3). Haplotype diversity is also greater in general for CABYV-like and RGroup (within-group diversity is higher in these groups relative to other groups). For CABYV, all ORFs have negative values for Tajima's D and Fu's F (Fu 1997) (Table 3.3). This indicates a population expansion such as a selective sweep or bottleneck. For CABYV-like, Tajima's D is negative for half of the ORFs, and all ORFs are positive with Fu's F test. RGroup's ORF1-2 and ORF3 have negative Tajima's D, but only ORF1-2 has negative Fu's F. Fu's F is a more robust test than Tajima's D for recent population expansion (Ramos-Onsins and Rozas 2002), and the results in this column show that there may be some false positives (negative values) using Tajima's D.

Virus	ORF	Nucleotide diversity	Haplotype diversity	Tajima's D	Fu's F
CABYV	P0	0.059	0.998	-0.647	-20.855
	P1-2	0.060	0.996	-1.027	-71.257e
	P3	0.038	0.990	-0.100	-2.359
	P4	0.049	0.988	-0.572	-20.137
	P3-5	0.078	0.999	-0.165	-7.639
CABYV-like	P0	0.147	1.0	-0.350	2.118
	P1-2	0.088	0.917	-0.507	1.765
	P3	0.048	0.889	0.189	0.334
	P4	0.032	1.0	-0.025	1.029
	P3-5	0.111	1.0	0.118	2.423
RGroup	P0	0.159	1.0	0.335	2.186
	P1-2	0.077	0.985	-0.658	-0.583
	P3	0.038	0.982	-0.834	1.506
	P4	0.044	1.0	0.565	0.989
	P3-5	0.109	1.0	0.169	2.452

***Table 3.3**. Genomic diversity and selection pressure metrics for nucleotide sequences of each ORF.*

Our results for average dN/dS (dN-dS) show that most ORFs 0, 1-2, 3, and 5 are under negative or purifying selection (Table 3.4) except for ORF0 of the RGroup. ORF 4 is not under negative or purifying selection. The P-values in the table represent the probability of rejecting a null hypothesis of neutrality (dN=dS). For most ORFs, we can reject the

null hypothesis. We cannot reject the null hypothesis for for CABYV ORF0, RGroup ORF3, and ORF4 for all virus groups.

Virus	ORF0	ORF0	ORF1-2	ORF1-2	ORF3	ORF3	ORF4	ORF4	ORF3-5	ORF3-5
	dN - dS	*P*-value	dN - dS	*P*-value	dN - dS	*P*-value	dN - dS	*P*-value	dN - dS	*P*-value
CABYV	-1.32	0.19	-8.53	0	-4.2	0	0.07	0.95	-10.57	0
CABYV-like	-3.42	0	-7.77	0	-4.04	0	1.56	0.12	-13.26	0
RGroup	2.42	0.02	-7.84	0	-1.41	0.16	0.88	0.38	-11.41	0

Table 3.4. *Codon-based Test of Neutrality for analysis averaging over all sequence pairs. The P-value is the probability of rejecting the null hypothesis of strict neutrality (dN = dS). The variance of the difference was computed using the bootstrap method (1000 replicates), and analyses were conducted using Pamilo-Bianchi-Li method.*

Phylogeography

Table 3.5 displays the posterior values of potential geographic origin regions as computed through a discrete phylogeography analysis using BEAST (Suchard et al. 2018). For all datasets (genome, CP, RdRp, CP+RdRp), the Mediterranean (Spain and France) was the most likely origin with posterior values of 0.655 (genome), 0.768 (CP), 0.756 (RdRp) and 0.714 (CP+RdRp).

Location	Genome	n Genome	CP	Rdrp	Concat	n Genes
China	0.062	3	0.023	0.044	0.046	34
Japan	0.056	1	0.0197	0.016	0.014	2
Mediterranean	**0.655**	4	**0.768**	**0.756**	**0.714**	4
Pakistan	NA	0	0.05	0.025	0.042	1
Philippines	NA	0	0.012	0.027	0.021	1
South Korea	0.021	28	0.006	0.009	0.009	28
Thailand	NA	0	0.018	0.019	0.024	12
USA central	0.037	1	0.018	0.017	0.018	1
USA west	0.169	17	0.063	0.063	0.087	16
Uzbekistan	NA	0	0.015	0.023	0.023	2

Table 3.5. *The posterior value results from discrete phylogeography analysis. The number of sequences used for each region are reported in the column to the right of the posterior values. Mediterranean sequences are in orange text. California isolates are in teal text.*

The visual (animated) results from the continuous phylogeography analysis include inferred regions of origins and spread (Genome in Supplementary Folder 3.1, CP in Supplementary Folder 3.2, RdRp in Supplementary Folder 3.3, and concatenated genes in Supplementary Folder 3.4). In contrast to the discrete phylogeography results, these inferences place the likely origins in Northern India/Western China (CP), Northeastern China (RdRp), and Northeast India or Nepal (CP+RdRp). The tree representation of the results (Figures S3.2 - S3.5) show Mediterranean origins for CP, RdRp, and CP+RdRp. The analysis from the full genome inferred an ocean origin, which suggests that whole

genome sequences available are not sufficient to draw robust inferences from this method.

Analysis of molecular variance (AMOVA)

AMOVA results are displayed in Table 3.6. All datasets had >90% variation within samples. For the CP, there was evidence of geographic structure, with 2.26% and 3.78% variation between broad regions and within broad region populations, respectively. For all genomic regions, AMOVA revealed small but significant variation between samples within each region, and small but significant variation between broad regions for CP. Overall there's evidence of gene flow between regions, which is expected for this virus.

Variation source	df	Sum of squares	Percent variation	Fixation indices
Genome				
Between broad regions	2	2	0	FCT=0
Between samples within broad regions	11	11	0	FSC= 0
Within samples	25	35	100	FST= 0
Coat protein				
Between broad regions	9	11.11	2.26	FCT= 0.0226*
Between samples within broad regions	28	28.74	3.78	FSC= 0.0386*
Within samples	55	51.61	93.98	FST= 0.0603*
RdRp				
Between broad regions	9	10.12	0.56	FCT= 0.0056
Between samples within broad regions	28	29.64	4.93	FSC= 0.0496*
Within samples	55	51.9	94.51	FST= 0.0549
CP & RdRp Concatenated				
Between broad regions	9	9.26	-0.18	FCT= -0.0018
Between samples within broad regions	28	28.98	2.52	FSC= 0.0251*
Within samples	55	53.7	97.67	FST= 0.0233*

Table 3.6. *Results from Analysis of Molecular Variance. Fixation indices with a significant associated P-value ($p<0.05$) noted with a *.*

Recombination analysis

For the full genome, the overall recombination test detected 40 events, after removing events that were non-significant and could be caused by processes other than

recombination (Supplementary Data Sheet 3.1). The concatenated gene dataset resulted in four events with intraspecific recombination and four with interspecific recombination (Table 3.7). One of the interspecific recombination events includes a mislabeled SABYV isolate, the others are RGroup and MABYV. For the full genome dataset, we found 32 intraspecific and eight interspecific events. One of the interspecific events includes the PABYV outgroup. For both the whole genome and concatenated datasets, there were multiple minor and major parents options for some events, but we reported the top choice given by RDP after cross-checking with Shimodaira–Hasegawa and approximately unbiased phylogenetic tree comparison tests (Shimodaira and Hasegawa 2001; Shimodaira 2002).The split decomposition networks (concatenated genes: Figure 3.6; full genome Figure S3.1) showed recombination networks between PABYV, SABYV, MABYV, and RGroups. The full genome network also shows a reticulation of CABYV-like isolates separate from the true CABYV.

Event number	Begin	End	Recombinant Sequence(s)	Minor Parental Sequence(s)	Major Parental Sequence(s)	Test	Highest P-value
1	1462	648	^GU324102__R__Taiwan__2009 EU24432K6_China_2016 GU324097_Taiwan_R_2008 GU324099_Taiwan_R_2009[P] GU324101_Taiwan_R_2008[P] GU324104_Taiwan_R_2008[P] HQ700824_Philippines_2010 JQ700306_Taiwan_2009 KF791037_Thailand_2012[P] KF815679_Thailand_20212[P] KF827827_Thailand_2013[P] KF827828_Tailand_2013[P] MN688219_India_2018 MNG688220_India_2019[P] MN843966_India_2019[P] MN843967_India_2019[P] MN862007_India_2019[P]	FJ460215__MABYV__China__2016	GU324098__Taiwan__2009	Bootscan	3.44E-18
2	1460	418	HQ439023__China__Beijing__2013	FJ460215__MABYV__China__2016	EF488996__China__2016	Chimaera	1.63E-13
3	108	578	^KF427707__Thailand__2010	FJ460217__MABYV__China__2016	KF815686__Thailand__2013	SiSscan	8.59E-13
4	648	1442	^FJ460213__MABYV__China__2016 DQ973123_China_2016[P] EF063704_China_2016[P] Ef063708_China_2016 EUO00534_MABYV_China_2008 EUO91149_MABYV_China_2008 EUO91150_MABYV_China_2008 EUO91151_MABYV_China_2008 EU262628_MABYV_China_2016 FJ460215_MABYV_China_2016 FJ460217_MABYV_China_2016	EU259785__China__2016	MH141555__Pakistan__2016	Bootscan	1.96E-05
5	628	1435	MRCP6__MotteRimrock__2017	MRCP5__MotteRimrock__2017	SSCF2B__ShipleySkinner__2017	GENECONV	1.87E-03
6	144	628	MRCF3C__MotteRimrock__2017	MRCP1__MotteRimrock__2017	X76931__France__1989	GENECONV	4.12E-03
7	1450	409	^MRCP1__MotteRimrock__2017	MRCP5__MotteRimrock__2017	EU000535__China__Beijing	3Seq	2.93E-02
9	1460	832	LC472499__Indonesia__EastJava__2017	GU324095__Taiwan__2008	KY617826__EastTimor__2017	GENECONV	7.60E-03

Table 3.7. *Concatenated gene region recombinants. Gray font shows those sequences that have evidence for the same recombination event, sequences with a [P] at the end mean that they are partially supported to have the same recombination event. Breakpoints with * show points where the breakpoints could not be determined, so uncertainty is denoted. When there is ambiguity between recombinant and parent sequences, this is annotated with a ^ in front of the accession.*

The full genome network (Figure S3.1) shows multiple reticulation events, which are evidence of recombination, among the outgroups and putative RGroups. Also visible is a reticulation among sequences from Taiwan, East Timor and Indonesia. And, within the "true" CABYV group, there is an apparent reticulation event likely representing the multiple intra-species recombination events detected by RDP. The concatenated gene split decomposition network (Figure 6) shows one large reticulation between CABYV,

MABYV, PABYV, and SABYV. All of the known and putative R groups are on one side of the split (with the other outgroups) in the network, and all of the "true" CABYV are on the other side of the split.

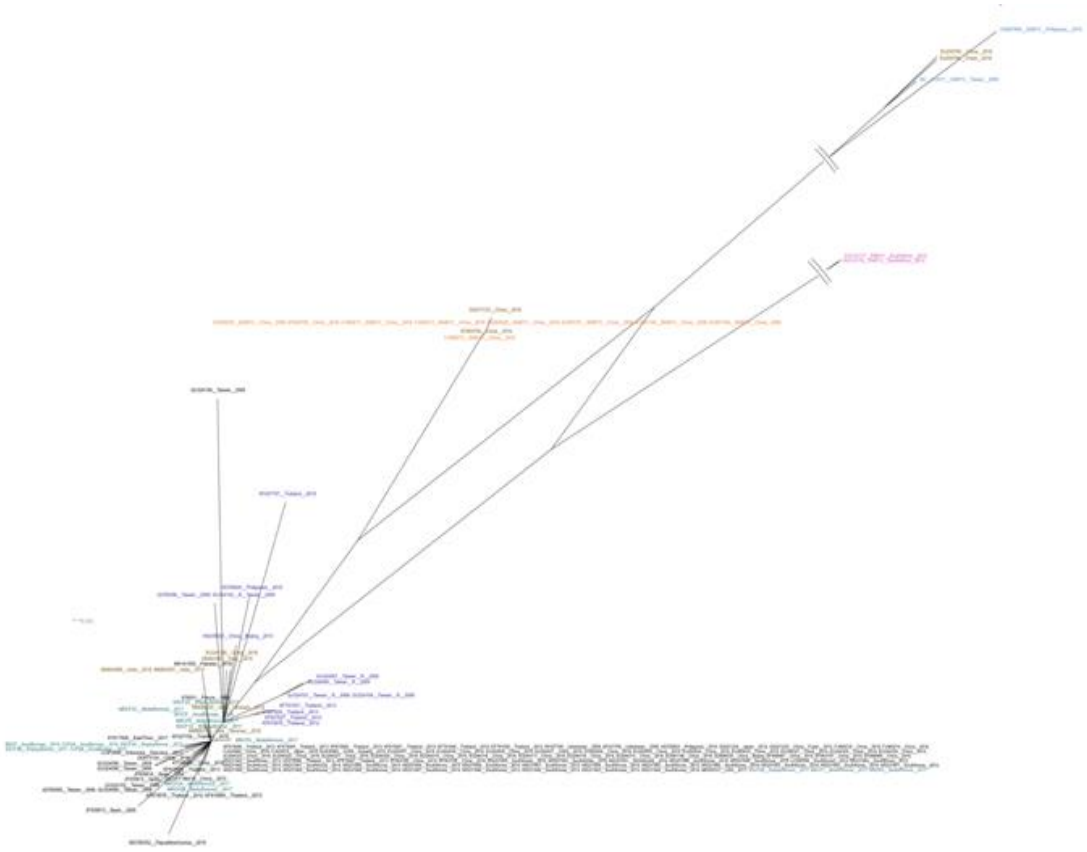

Figure 3.6. *Split decomposition network for the concatenated dataset for CABYV and all outgroups. The network is color-coded to match phylogenetic tree: California sequences are in teal, known RGroup in purple, "mislabeled" in brown, MABYV in orange, PABYV in pink, and SABYV in blue. Reticulation events (connected lines) indicate recombination events. Network created by the SplitsTree software (Huson and Bryant 2006), and edited in Affinity Designer software (RRID:SCR_016952).*

Discussion

Our study clarifies the origins and genetic features of CABYV found in California wild plants while advancing our understanding of CABYV diversity, origins and

relationships to other poleroviruses globally. We clarified relationships among regional groupings and found evidence for putative novel *Polerovirus* species and recombinants. We also found that CABYV sequences generated to date include mislabeled recombinants and possibly new *Polerovirus* species from the Taiwan/Asia Pacific region. Additionally, there is scant evidence for regional groupings as previously proposed; only phylogenies from ORF1-2 supported these groupings, but ORFs1-2+3, ORF3, and the full genomes did not. Our discrete phylogeography analysis, which is limited by user-based regional boundaries, identified the Mediterranean region as a probable origin of CABYV. However, continuous phylogeography provided support for an alternative hypothesis of Northern India as a possible origin, even without Indian isolates in the analysis. Based on all CABYV ORFs, there is evidence for population bottlenecks or selective sweeps, while the analysis of CABYV-like and RGroup datasets did not show consistent evidence for these events, except for the RGroup ORF1-2 (P1-2). Nucleotide diversity within ORFs is higher in the CABYV-like and RGroup datasets than CABYV datasets, showing high amounts of diversity, especially for ORFs coding P0 and P3-5 proteins.

The California isolates group with sequences from both Asia (3) and the Mediterranean (14). This is evidence that there have been at least two introductions of CABYV into southern California. However, we did not find consistent evidence for regional groupings that have been reported before for Asian, Mediterranean, and Taiwanese clades (Kassem et al. 2013; Kwak et al. 2018). Using only ORF3, (Minicka et al. 2020) also found no evidence for regional groups. We did find support for

Taiwan/Asia-Pacific isolates forming their own clade with high bootstrap values; however these isolates may represent a new virus species according to our subsequent analyses. To delimit species within the genus *Polerovirus*, a cut-off of 10% amino acid sequence difference between any gene product has been proposed (King et al. 2011). Our amino acid pairwise comparisons show that RGroup (the stable CABYV-MABYV recombinant) has a greater than 10% difference from CABYV for two out of five proteins, and 10.5% difference for a third. This difference is even greater for sequences we designated as CABYV-like, with four out five proteins greater than 10% difference. We also found that CABYV-like isolates, which are distinct from RGroup, but consistently form their own clade among CABYV-like sequences, have greater than 10% difference from CABYV for three out five proteins. Thus, our CABYV-like group may contain up to two new *Polerovirus* species that are closely related to CABYV. Collectively, our analyses suggest that it may be more useful, biologically, to consider CABYV as a species complex rather than a single species. Depending on the protein analyzed, an isolate may or may not be identified as CABYV. For some of these isolates, it will be essential to do transmission and co-infection tests to conclusively determine species status.

Recombination-driven evolution is common in the family Luteoviridae. In this CABYV polerovirus complex, we found that inter- and intra-specific recombination are both common. With just the concatenated gene dataset, we identified four events with intraspecific recombination and four with interspecific recombination. For the full genome dataset, we found 32 intraspecific and eight interspecific events. These

recombinants, and the similarity between the viruses in this group, which are themselves likely to be stable recombinants, have resulted in mislabeled publicly available genetic information. In total, we found that 15 isolates were mislabeled as CABYV on GenBank. Two isolates were labelled as CABYV but are in fact SABYV. Seven were labelled as CABYV, but are actually MABYV, including three sequences from the original MABYV description (EU091151.1, EU091149.1, EU091150.1 in (Xiang et al. 2008). Six isolates are grouped with known recombinants (RGroup) but are not labelled as recombinants on GenBank. And there are 12 recombinant-labelled isolates on GenBank that we can determine based on publication or an "R" in the isolate name. However, even with this R designation, they are still CABYV first, which leads to confusion: additional CABYV RGroup sequences are added under CABYV without any indication that they are recombinants. Misidentification errors such as those identified here will persist if not corrected (e.g., sequences for MABYV are still identified as CABYV). To understand the ecology, distribution, and evolution of CABYV it is important to perform phylogenetic analyses on sequences to determine species identities - and to include multiple representatives from CABYV that include recombinants.

Besides resolving misidentifications, phylogenetic analyses also provided insights into the minimum sequence information needed to identify virus species. For CABYV, Knierim et al. (2010) proposed that the 3' section of ORF2, the intergenic region, and ORF3/ORF4 are the minimum sequence information required for CABYV identification. This partially includes the RdRp (ORF1-2) and the ORF for coat protein and movement proteins, which occupy the same region, but are in a different reading frame (Xiang et al.

2008). Our work shows that the complete RdRp or partial RdRp (ORF1-2 or ORF2), plus the intergenic region and coat protein (ORF3), are sufficient to determine species identity and identify recombination events. Including the intergenic region is important because recombination often occurs at this point. The coat protein region alone is not sufficient for identification. However, the majority of available CABYV sequences consist of only this region, which greatly reduces the available sequence information for this species. Our analysis indicates that recombinant isolates of different species are present in China, Taiwan and Thailand. There are apparent recombination events between isolates of the same species in California, South Korea, and the Mediterranean. And, the CABYV-like group forms a reticulation event, which is also evidence of recombination. More sequences from underrepresented regions would be especially helpful for understanding the movement/dispersal and evolution of CABYV, recombinants, and related polerovirus species.

Both discrete and continuous phylogeography can be used to determine the geographic origins using sequence data. The Mediterranean region is the hypothesized origin for CABYV based on the original isolation of this species from regions in France and Spain. Our discrete phylogeography analyses using the full genome, coat protein (ORF3), RNA-dependent RNA polymerase (ORF1-2), and both genes concatenated, all estimated the origin of CABYV in the Mediterranean. However, discrete phylogeography can be influenced by artificially grouping isolates into geographic regions, as well as sampling biases. Here, we see that even when there are only four isolates representing a region (one from France, and three from Spain), as opposed to 17 from California and 28

from South Korea, the model did not identify a region based on sampling bias - all tests supported the Mediterranean as the origin. To further address the issue of artificial grouping, we compared these results to those obtained by using a continuous phylogeographic analysis. The benefit of using continuous phylogeography is that the location trait is less arbitrary (not on artificial grouping, but instead latitude longitude coordinates), and the model can infer origin and other internal nodes where the virus may be located. However, there are no reported latitude and longitude points for most, if any, GenBank submitted sequences, and at best the isolates had county, state, or province associated with them. We therefore had to use generic coordinates for the country, county, state, or province in our analyses. The continuous phylogeography approach suggested in fact a different origin, Northeastern India, based on ORF3 (CP) and ORF1-2+3, and Northeastern China based on ORF1-2 (RdRp). The full genome resulted in a less clear origin (center of Atlantic ocean), followed by a split towards California and South Korea, likely because of the much patchier sampling data available. It appears that this method is sensitive to sampling biases; there were many more distributed locations in the ORF datasets compared to the full genome, for which isolates from the US and South Korea provided the most sequences.

Even though there are isolates of CABYV from India available on GenBank (Kumar et al. 2021), our phylogenetic work and recombination detection tests showed that they are RGroup isolates (stable recombinants with MABYV) instead. Therefore, there are no sequences from India included in this analysis. This is the first time this region has been associated with CABYV origin, but the suggestion is not without

biological backing. There is indeed high genetic diversity of wild *Cucumis melo* L. in South and East Asia (Sebastian et al. 2010), and India is a primary center of melon diversity and domestication (Gonzalo et al. 2019). Melons and cucumbers are of Asian origin, with multiple domestication events in Asia (and independently in Africa, and cucurbits in the Americas) (Sebastian et al. 2010; Endl et al. 2018; Chomicki et al. 2020). CABYV has primarily been reported infecting plants in the family Cucurbitaceae, and particularly *Cucumis*, but it can infect hosts in other families (cabi.org CABYV entry). Our analysis suggests that CABYV could have evolved in tandem with *Cucumis* hosts during the domestication process. Exploring this will require more extensive sampling in the putative origin location or isolation of sequences from historical samples (Malmstrom et al. 2007).

Overall, our study shows that CABYV has high intraspecific diversity, even within geographic regions, and likely is a species complex. Using phylogenetic methods, recombination detection tests, and pairwise comparisons between groups, we found that there is evidence that the recombinant CABYV is its own unique species, and that isolates in the Taiwan regional grouping are also different species from CABYV. For those with access to these virus populations, it will be important to test for co-infection and other biological differences. Within decades of discovery, this virus has spread to many countries and continents, likely assisted by global trade and the widespread distribution of the aphid vectors *A. gossypii* and *M. persicae* (cabi.org *A.s gossypii* entry). CABYV was first described in samples from France, which is the putative origin of this virus. However, using new continuous phylogeography methods we found that the origin

of CABYV may be in Asia (northeastern India by analyzing open reading frames for coat protein and coat protein concatenated with RNA-dependent RNA-polymerase, or China by RNA-dependent RNA-polymerase alone). This study demonstrates the challenges of understanding a virus that frequently undergoes recombination with other poleroviruses. Future work on CABYV should include sampling of additional putative origin points as well as sequencing of genomic regions that enable researchers to easily discern recombinants from parental viruses.

References

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. https://doi.org/10.1093/nar/gky379

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *15*(4), e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Brault, V., Périgon, S., Reinbold, C., Erdinger, M., Scheidecker, D., Herrbach, E., Richards, K., & Ziegler-Graff, V. (2005). The polerovirus minor capsid protein determines vector specificity and intestinal tropism in the aphid. *Journal of Virology*, *79*(15), 9685–9693. https://doi.org/10.1128/JVI.79.15.9685-9693.2005

Carmo-Sousa, M., Moreno, A., Plaza, M., Garzo, E., & Fereres, A. (2016). Cucurbit aphid-borne yellows virus (CABYV) modifies the alighting, settling and probing behaviour of its vector *Aphis gossypii* favouring its own spread. *The Annals of Applied Biology*, *169*(2), 284–297. https://doi.org/10.1111/aab.12300

Chare, E. R., & Holmes, E. C. (2004). Selection pressures in the capsid genes of plant RNA viruses reflect mode of transmission. *The Journal of General Virology*, *85*(Pt 10), 3149–3157. https://doi.org/10.1099/vir.0.80134-0

Chomicki, G., Schaefer, H., & Renner, S. S. (2020). Origin and domestication of Cucurbitaceae crops: insights from phylogenies, genomics and archaeology. *The New Phytologist*, *226*(5), 1240–1255. https://doi.org/10.1111/nph.16015

Costa, T. M., Blawid, R., Aranda, M. A., Freitas, D. M. S., Andrade, G. P., Inoue-Nagata, A. K., & Nagata, T. (2019). Cucurbit aphid-borne yellows virus from melon plants in Brazil is an interspecific recombinant. *Archives of Virology*, *164*(1), 249–254. https://doi.org/10.1007/s00705-018-4024-2

Costa, T. M., Inoue-Nagata, A. K., Vidal, A. H., Ribeiro, S. G., & Nagata, T. (2020). The recombinant isolate of cucurbit aphid-borne yellows virus from Brazil is a polerovirus transmitted by whiteflies. *Plant Pathology*. https://doi.org/10.1111/ppa.13186

Dellicour, S., Gill, M. S., Faria, N. R., Rambaut, A., Pybus, O. G., Suchard, M. A., & Lemey, P. (2021). Relax, keep walking-a practical guide to continuous phylogeographic inference with BEAST. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msab031

Endl, J., Achigan-Dako, E. G., Pandey, A. K., Monforte, A. J., Pico, B., & Schaefer, H. (2018). Repeated domestication of melon (*Cucumis melo*) in Africa and Asia and a new close relative from India. *American Journal of Botany*, *105*(10), 1662–1671. https://doi.org/10.1002/ajb2.1172

Fomitcheva, V. W., Sukhacheva, E. A., & Schubert, J. (2004). Detection of Turnip yellows virus-encoded RNA-dependent RNA polymerase using monoclonal antibodies. *Archives of Phytopathology and Plant Protection*, *37*(1), 9–17. https://doi.org/10.1080/03235400410001645958

Fu, Y. X., & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics*, *133*(3), 693–709.

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, *147*(2), 915–925. https://www.ncbi.nlm.nih.gov/pubmed/9335623

Gonzalo, M. J., Díaz, A., Dhillon, N. P. S., Reddy, U. K., Picó, B., & Monforte, A. J. (2019). Re-evaluation of the role of Indian germplasm as center of melon diversification based on genotyping-by-sequencing analysis. *BMC Genomics*, *20*(1), 448. https://doi.org/10.1186/s12864-019-5784-0

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*, *38*(Web Server issue), W695–W699. https://doi.org/10.1093/nar/gkq313

Guilley, H., Wipf-Scheibel, C., Richards, K., Lecoq, H., & Jonard, G. (1994). Nucleotide sequence of cucurbit aphid-borne yellows luteovirus. *Virology*, *202*(2), 1012–1017. https://doi.org/10.1006/viro.1994.1429

Hill, V., & Baele, G. (2019). Bayesian estimation of past population dynamics in BEAST 1.10 using the Skygrid coalescent model. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msz172

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267. https://doi.org/10.1093/molbev/msj030

Ibaba, J. D., Laing, M. D., & Gubba, A. (2017). Pepo aphid-borne yellows virus: a new species in the genus Polerovirus. *Virus Genes*, *53*(1), 134–136. https://doi.org/10.1007/s11262-016-1390-2

Kahle, D., & Wickham, H. (2013). Ggmap: Spatial visualization with ggplot2. *The R Journal*, *5*(1), 144. https://doi.org/10.32614/rj-2013-014

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, *6*, 208. https://doi.org/10.3389/fgene.2015.00208

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281

Kassem, M. A., Juarez, M., Gómez, P., Mengual, C. M., Sempere, R. N., Plaza, M., Elena, S. F., Moreno, A., Fereres, A., & Aranda, M. A. (2013). Genetic diversity and potential vectors and reservoirs of *Cucurbit aphid-borne yellows virus* in southeastern Spain. *Phytopathology*, *103*(11), 1188–1197. https://doi.org/10.1094/PHYTO-11-12-0280-R

Kesanakurti, P., Belton, M., Saeed, H., Rast, H., Boyes, I., & Rott, M. (2016). Screening for plant viruses by next generation sequencing using a modified double strand RNA extraction protocol with an internal amplification control. *Journal of Virological Methods*, *236*, 35–40. https://doi.org/10.1016/j.jviromet.2016.07.001

Khanal, V., & Ali, A. (2017). First report of *Cucurbit aphid-borne yellows virus* infecting *Cucurbita pepo* in Oklahoma. *Plant Disease*. https://doi.org/10.1094/PDIS-10-17-1675-PDN

King, A. M. Q., Lefkowitz, E., Adams, M. J., & Carstens, E. B. (2011). *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier. https://play.google.com/store/books/details?id=aFYaE9KXEXUC

Knierim, D., Deng, T. C., Tsai, W. S., Green, S. K., & Kenyon, L. (2010). Molecular identification of three distinct *Polerovirus* species and a recombinant *Cucurbit aphid-*

borne yellows virus strain infecting cucurbit crops in Taiwan. *Plant Pathology*, *59*(5), 991–1002. https://doi.org/10.1111/j.1365-3059.2010.02327.x

Knierim, D., Tsai, W. S., Maiss, E., & Kenyon, L. (2014). Molecular diversity of poleroviruses infecting cucurbit crops in four countries reveals the presence of members of six distinct species. *Archives of Virology*, *159*(6), 1459–1465. https://doi.org/10.1007/s00705-013-1939-5

Kumar, A., Bhattarai, A., Rathore, A. S., Jat, G. S., Behera, T. K., Renukadevi, P., Baranwal, V. K., Jain, R. K., & Basavaraj, Y. B. (2021). Association of cucurbit aphid-borne yellows virus with cucumber plants in India. *Virusdisease*, *32*(1), 183–185. https://doi.org/10.1007/s13337-020-00645-4

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kwak, H.-R., Lee, H. J., Kim, E.-A., Seo, J.-K., Kim, C.-S., Lee, S. G., Kim, J.-S., Choi, H.-S., & Kim, M. (2018). Complete genome sequences and evolutionary analysis of *Cucurbit aphid-borne yellows virus* isolates from melon in Korea. *The Plant Pathology Journal*, *34*(6), 532–543. https://doi.org/10.5423/PPJ.OA.03.2018.0049

Lecoq, H., Bourdin, D., Wipf-Scheibel, C., Bon, M., Lot, H., Lemaire, O., & Herrbach, E. (1992a). A new yellowing disease of cucurbits caused by a luteovirus, cucurbit aphid-borne yellows virus. *Plant Pathology*, *41*(6), 749–761. https://doi.org/10.1111/j.1365-3059.1992.tb02559.x

Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., & Smith, D. B. (2018). Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Research*, *46*(D1), D708–D717. https://doi.org/10.1093/nar/gkx932

Lemaire, O. J., Gubler, W. D., Valencia, J., Lecoq, H., Falk, B. W., & Others. (1993). First report of cucurbit aphid-borne yellows luteovirus in the United States. *Plant Disease*, *77*(11). https://www.cabdirect.org/cabdirect/abstract/19942300629

Malmstrom, C. M., Shu, R., Linton, E. W., Newton, L. A., & Cook, M. A. (2007). Barley yellow dwarf viruses (BYDVs) preserved in herbarium specimens illuminate historical disease ecology of invasive and native grasses. *The Journal of Ecology*, *95*(6), 1153–1166. https://doi.org/10.1111/j.1365-2745.2007.01307.x

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*(1), vev003. https://doi.org/10.1093/ve/vev003

Ma, Y., Marais, A., Lefebvre, M., Theil, S., Svanella-Dumas, L., Faure, C., & Candresse, T. (2019). Phytovirome analysis of wild plant populations: Comparison of double-stranded RNA and virion-associated nucleic acid metagenomic approaches. *Journal of Virology, 94*(1). https://doi.org/10.1128/JVI.01462-19

Mayo, M. A., & Miller, W. A. (1999). The structure and expression of luteovirus genomes. *The Luteoviridae,* 23–42.

rMinicka, J., Zarzyńska-Nowak, A., Budzyńska, D., Borodynko-Filas, N., & Hasiów-Jaroszewska, B. (2020). High-throughput sequencing facilitates discovery of new plant viruses in Poland. *Plants*, *9*(7). https://doi.org/10.3390/plants9070820

Mnari Hattab, M., Kummert, J., Roussel, S., Ezzaier, K., Zouba, A., & Jijakli, M. H. (2005). First report of *Cucurbit aphid-borne yellows virus* in Tunisia causing yellows on five cucurbitaceous species. *Plant Disease*, *89*(7), 776. https://doi.org/10.1094/PD-89-0776B

Omar, A. F., & Bagdady, N. A. (2012). *Cucurbit aphid-borne yellows virus* in Egypt. *Phytoparasitica; Israel Journal of Plant Protection Sciences*. https://doi.org/10.1007/s12600--0212

Pagán, I., & Holmes, E. C. (2010). Long-term evolution of the Luteoviridae: time scale and mode of virus speciation. *Journal of Virology*, *84*(12), 6177–6187. https://doi.org/10.1128/JVI.02160-09

Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* , *26*(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

Pfeffer, S., Dunoyer, P., Heim, F., Richards, K. E., Jonard, G., & Ziegler-Graff, V. (2002). P0 of beet *Western yellows virus* is a suppressor of posttranscriptional gene silencing. *Journal of Virology*, *76*(13), 6815–6824. https://doi.org/10.1128/jvi.76.13.6815-6824.2002

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, *31*(7), 1929–1936. https://doi.org/10.1093/molbev/msu136

Prüfer, D., Kawchuk, L. M., & Rohde, W. (2006). Polerovirus ORF0 genes induce a host-specific response resembling viral infection. *Canadian Journal of Plant Pathology. Revue Canadienne de Phytopathologie*, *28*(2), 302–309. https://doi.org/10.1080/07060660609507299

Rabadán, M. P., Juárez, M., De Moya-Ruiz, C., & Gómez, P. (2021). Aphid-borne viruses infecting cultivated watermelon and squash in Spain: characterization of a variant of cucurbit aphid-borne yellows virus (CABYV). *Plant Pathology*, *n/a*(n/a). https://doi.org/10.1111/ppa.13390

Rambaut, A., Lam, T. T., Max Carvalho, L., & Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution*, *2*(1), vew007. https://doi.org/10.1093/ve/vew007

Ramos-Onsins, S. E., & Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, *19*(12), 2092–2100. https://doi.org/10.1093/oxfordjournals.molbev.a004034

Schmitz, J., Stussi-Garaud, C., Tacke, E., Prüfer, D., Rohde, W., & Rohfritsch, O. (1997). In situ localization of the putative movement protein (pr17) from potato leafroll luteovirus (PLRV) in infected and transgenic potato plants. *Virology*, *235*(2), 311–322. https://doi.org/10.1006/viro.1997.8679

Sebastian, P., Schaefer, H., Telford, I. R. H., & Renner, S. S. (2010). Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(32), 14269–14273. https://doi.org/10.1073/pnas.1005338107

Shang, Q.-X., Xiang, H.-Y., Han, C.-G., Li, D.-W., & Yu, J.-L. (2009). Distribution and molecular diversity of three cucurbit-infecting poleroviruses in China. *Virus Research*, *145*(2), 341–346. https://doi.org/10.1016/j.virusres.2009.07.017

Shates, T. M., Sun, P., Malmstrom, C. M., Dominguez, C., & Mauck, K. E. (2019). Addressing research needs in the field of plant virus ecology by defining knowledge gaps and developing wild dicot study systems. Frontiers in Microbiology, 9, 3305. https://doi.org/10.3389/fmicb.2018.03305

Shen, W., Shi, Y., Dai, Z., & Wang, A. (2020). The RNA-dependent RNA polymerase NIb of potyviruses plays multifunctional, contrasting roles during viral infection. *Viruses*, *12*(1). https://doi.org/10.3390/v12010077

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*, 539. https://doi.org/10.1038/msb.2011.75

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1), vey016. https://doi.org/10.1093/ve/vey016

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595.

Vafaei, S. H., & Mahmoodi, M. (2017). Presence of recombinant strain of cucurbit aphid borne yellows virus in Iran. *Iranian Journal of Biotechnology*, *15*(4), 289–295. https://doi.org/10.15171/ijb.1541

Xiang, H.-Y., Shang, Q.-X., Han, C.-G., Li, D.-W., & Yu, J.-L. (2008). Complete sequence analysis reveals two distinct poleroviruses infecting cucurbits in China. *Archives of Virology*, *153*(6), 1155–1160. https://doi.org/10.1007/s00705-008-0083-0

Chapter 4. Cucurbit aphid-borne yellows virus has species-specific effects on wild squash establishment, disease resistance, and interactions with virus vectors

Abstract

The effects of crop-associated viruses on wild plants are understudied but may be important drivers of species interactions in plant communities. Previously, we found that one crop-associated virus, *Cucurbit aphid-borne yellows virus*, is highly prevalent in wild growing populations of *Cucurbita foetidissima* and *Cucurbita palmata* in chaparral communities of the southwestern U.S.. Here, we used controlled greenhouse experiments, paired with observations of disease symptoms in wild-growing plants and insect feeding behavior assays, to test the hypothesis that CABYV infection negatively affects wild squash performance and alters interactions with vectors. We found that over the course of eight weeks, virus-infected *C. palmata* had more severe symptoms than *C. foetidissima*, but that both species had reduced aboveground biomass in response to infection. Wild growing plants show symptoms of disease, and the cumulative effects of those symptoms result in changes in growth. We also found that virus infection altered the feeding behavior of the main aphid vector, *Aphis gossypii*, in ways that are conducive to transmission, but only *C. palmata* which is more susceptible to severe CABYV infections *and* to aphid feeding. Overall, our results show that wild plants are susceptible to the negative impacts of the pathogens circulating and evolving within their cultivated counterparts, with significant consequences for plant recruitment and survival.

Introduction

There is an increase of agriculture around the globe, including intensification on existing land and expansion into new, previously uncultivated areas (Zabel et al., 2019). Both increases hurt efforts to conserve biodiversity. For example, the conversion of land from unmanaged communities to agriculture reduces endemic species habitat. And intensification on existing lands introduces more agro-chemicals (Zabel et al., 2019). Another important side effect of agricultural expansion/intensification is the increase in endemic pathogens infecting newly introduced crops (Jones, 2020). Historically, the movement of pathogens from wild to cultivated systems has received the majority of attention. However, we are now shifting our focus to virus movement into wild plants and realizing that one this movement of pests and pathogens has been overlooked. As agricultural practices intensify and expand into previously uncultivated areas, there is an increased chance of encounters between crop pathogens and wild plants.

Due to their reliance on insect vectors capable of flight and long-distance migrations on air currents, insect-transmitted pathogens, especially viruses, are likely to spill over to wild plants (Alexander et al., 2014). Despite this, much of the research on crop-associated pathogen effects in wild plants focus on bacteria and fungi. This is due to the fact that the former two pathogens are easier to identify in wild plants using symptoms, or in the absence of these, molecular techniques using universal coding sequences that amplify conserved regions. Viruses lack universal coding sequences and virus infections are frequently unapparent in wild plant populations. A handful of studies

completed over the last several decades document negative effects of crop-associated viruses on plant fitness (Malmstrom and Alexander, 2016). However, these studies frequently focus on annual plants and are biased toward grasses (family Poaceae)(Shates et al., 2019). To understand the roles and impacts of viruses in wild plant communities, especially those spilling over from crops, manipulative studies are needed. Additionally, these studies should focus on perennials, which dominate wild plant communities, and expand beyond Poaceae species to encompass more dicot hosts (Alexander et al., 2014).

We address this knowledge gap through a combination of manipulative and observational experiments using two perennial wild cucurbit species endemic to southern California: *Cucurbita foetidissima* Kunth and *C. palmata* Wats. (Figure 4.1). We chose this system because these plants are both key species common in chaparral/grassland habitats and used by other organisms and; they are congeneric and confamilial with crops grown in this region (squash and melons) and may therefore share the pathogen complexes found in these crops (Tyack et al., 2020). *Cucurbita foetidissima* is associated with semi-arid lands and *C. palmata* is a true xerophytic species, and both have adaptations to survive hot, dry environments (Bemis & Whitaker, 1969). Previously, we found that these plants host diverse crop-associated virus communities (Shates et al., 2019); Chapter 1 & Chapter 2 of this dissertation). We identified one virus that is common across many populations: *Cucurbit aphid-borne yellows virus* (CABYV). CABYV (originally classified as family Luteoviridae and now reassigned to family Solemoviridae, genus *Polerovirus*) was first described in 1992 in France and is transmitted by two aphid species - *Aphis gossypii* Glover and *Myzus persicae* (Sulzer)

(Lecoq et al., 1992). In crops, CABYV causes symptoms that are especially severe in hotter weather: mosaic/mottling early in disease progression, yellowing and thickening of basal leaves, necrosis of basal leaves, interveinal chlorosis, reduced plant growth, reduced fruit set and quality, and increased flower abortion (Kassem et al., 2007). As a result, CABYV causes economic losses to cucurbit-producing regions where it is present in crops (Bananej et al., 2009; Desbiez et al., 2019; Kassem et al., 2007; Knierim et al., 2010; Kwak et al., 2018).

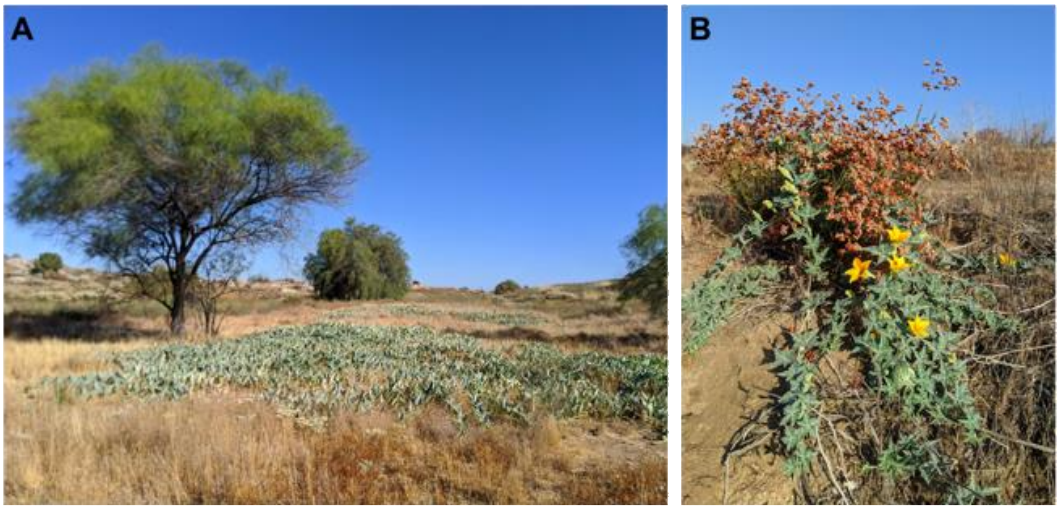

Figure 4.1. *Target hosts in Motte Rimrock Reserve. A) Mature Cucurbita foetidissima, which can grow across a large area with long vines extending from a central taproot, often overlapping with other individuals. B) Mature Cucurbita palmata plant, showing the difference in size and growth habit relative to C. foetidissima. Individuals are more easily distinguished and do not grow in dense, physically overlapping populations.*

We hypothesized that CABYV infection would negatively affect wild squash performance in ways that are relevant for survival or fitness. To test this hypothesis, we carried out greenhouse experiments under controlled conditions to quantify disease symptoms and performance metrics of young CAYBV-infected and sham-inoculated *C. foetidissima* and *C. palmata*. To connect these findings to outcomes for mature plants in

the field, we also tracked performance and fitness metrics of naturally occurring wild cucurbits with positive CABYV detections and determined correlations with a disease index quantifying the severity of CABYV symptoms. Lastly, to explore how CABYV infection affects host interactions with the aphid vector (a key herbivore in this system), we performed preliminary experiments on aphid feeding behavior. We evaluated effects of cucurbit species, virus infection, and their interaction on feeding behaviors of *Aphis gossypii* that are important for virus acquisition and host plant exploitation.

Methods

Organisms

Aphis gossypii used in our experiments were from a laboratory line originally collected from squash about a decade ago near Reedley, CA, USA. Aphid colonies were maintained on melons (*Cucumis melo* L. cv. "Iroquois"). We collected the CABYV isolate in the summer of 2019 from wild-growing *C. palmata* plants at Motte Rimrock Reserve (DOI: 10.21973/N31T0W) and maintained the isolate in cultivated squash (*Cucurbita pepo* cv. Dixie) through transmission by *Aphis gossypii*. Both cultures were maintained in the laboratory at room conditions in mesh cages near windows with natural light and supplemental lighting (16L:8D photoperiods).

Cucurbita foetidissima and *C. palmata* seeds originated from four individuals (two of each species) growing wild at Motte Rimrock and opportunistically around residential areas in Riverside County. Field observations were conducted at Motte Rimrock Reserve (DOI: 10.21973/N31T0W) on established plant populations. Motte

Rimrock Reserve (289 ha) is primarily an inland coastal sage scrub community with some riparian and grassland communities and is adjacent to varying levels of land development (Motte Rimrock Reserve Statistics, 2010). In the summer months (when *Cucurbita* species are leafed out), temperature highs are on average 30-33°C, and lows 13-17°C, with little to no rainfall.

Plant culture and virus inoculations

We followed seed germination protocols tested previously on these species in another laboratory (personal communication with M. Uribe-López). For each species, we filled a mason jar with water and added seeds. *C. palmata* seeds were left to soak in the mason jar unperturbed for 72 hours at room temperature. Soft plastic tubing attached to the building's airline and affixed to the mason jar was used to aerate *C. foetidissima* seeds for 72 hours at room temperature. After pre-treatments, seeds were placed in petri dishes with damp filter paper for 48 hours, then planted in 2 x 2 x 2 inch pots later kept in a greenhouse with natural and artificial light at 16L:8D photoperiod, between 25-28 °C, 20-30% RH. Approximately two weeks post germination, when most plants had two true leaves, seedlings were moved to a temperature controlled walk-in growth chamber at 23±1 °C, 60±5% RH, and 16L:8D photoperiod. To generate CABYV-infected and sham-inoculated treatments for the greenhouse plant health experiments, 20 aphids from a CABYV culture, and 20 aphids from a clean culture, were placed on each plant. CABYV is a persistently-transmitted virus that is acquired and inoculated from the phloem, generally requiring at least 24 hours for acquisition/inoculation to be successful (Carmo-

Sousa et al., 2016). We allowed aphids to feed for 72 hours, then treated plants with Dinotefuran at the label rate for cucurbits to remove aphids. Two weeks later, plants were repotted into 6" x 5-3/4" inch large round pots and returned to the original greenhouse for the rest of the experiment. Every other row of plants was a different treatment group (CAYBV-inoculated or sham), with *C. palmata* and *C. foetidissima* randomly distributed through each row. Plants were watered every day and sprayed with miticide and fungicide as needed.

The plants used for aphid feeding experiments (using electrical penetration graph technique) followed the same planting protocol up to the petri dish step. From there, seed *groups* were planted into 6" x 5-3/4" inch round pots with a mixture of soil, perlite, and vermiculite. Then, individual seedlings that successfully germinated were transplanted to 4 ¼" L x 4 ¼" W x 4 7/8" H individual pots individual pots. Twenty aphids from CABYV infected plants and 20 aphids from the non-virus colony were placed on the plants for 72 hours when Safer® Soap was applied to the leaves. One week before the aphid feeding experiment, Safer® Soap was reapplied to ensure there are as few aphids as possible on the experimental plants. Safer® Soap was used instead of Dinotefuran, because as a systemic it would interfere with aphid feeding. Aphids were allowed to feed on leaves that were not exposed to Safer® Soap.

Plant health measurements in the greenhouse

We recorded the total number of leaves and flowers of each plant every two weeks, starting at 2 weeks post inoculation (WPI) and running to 8 WPI. At 8 WPI, we

collected leaf punches and cut root tissue from sham-inoculated and virus-inoculated plants. These tissues were stored in -80°C before RNA extraction and virus detection with RT-PCR to confirm infections. Aboveground (foliar tissue) was cut at the base of the stem, then weighed fresh and after drying for 72 hours. Belowground (root tissue) was removed from pots and gently washed to clean as much soil as possible while retaining fine root tissue attached to the main tap root, then dried for 72 hours. We recorded masses of fresh and dried tissue, but only used dried tissue for subsequent analyses. In total, the experiment included 109 plants: 35 CABYV-inoculated and 26 sham-inoculated *C. foetidissima*, and 26 CABYV-inoculated and 22 sham-inoculated *C. palmata*.

Evaluation of symptom severity

In the greenhouse experiments, we scored each fully expanded leaf with a scale developed by Takeshita et al., (2013) and subsequently refined for use in our laboratory (Kenney et al., 2020). All leaf ratings were combined to calculate whole-plant symptom severity with this formula: D = [(di)/n/4], where D is symptom severity, di is the symptom rating, n is the total number of leaves, and 4 represents the 4 levels of rating. For the mottling and yellowing symptoms, we used the following scale: 0 = no symptoms, 1 = slight mottling (tiny light spots visible), 2 ≤ 20% leaf area yellow (but bigger, distinctly yellow spots, rather than tiny light spots), 3 = 21%–50% leaf area yellow, and 4 ≥ 50% leaf area yellow. For leaf distortion, we used this scale: 0 = no symptoms, 1 = slight distortion (leaf shape is irregular, or surface slightly bumpy), 2 ≤

20% leaf area is curled, has large bumps, or twisted, 3 = 21%–50% leaf area is curled, has large bumps, or twisted, and 4 ≥ 50% is curled, twisted, and is not typically "leaf-shaped." At 8 WPI, we also recorded the intensity of powdery mildew infection, which established on plants in the latter two weeks of growth. For powdery mildew, we used the following scale: 0 = no symptoms, 1 = fewer than 4 developing powdery mildew (PM) spots, 2 ≤ 20% leaf area covered in PM, 3 = 21%–50% leaf area is covered in PM, and 4 ≥ 50% leaf area is covered in PM.

Field observations of infected plant symptoms and performance metrics

We visited previously identified CABYV-infected plants of *C. foetidissima* and *C. palmata* at Motte Rimrock Reserve at three time points: June 12, 2020, July 3, 2020, and July 23, 2020. We attempted to also track non-infected plants, but CABYV prevalence was too high to identify enough individuals. We tagged two vines per plant by tying flagging tape at the base and tip of the vine. For these two vines, we collected data on growth metrics, fitness proxies (flowers, fruits), and symptom severity. For growth and fitness metrics, we added another tag at the tip of the vine as it grew and measured the distance using a measuring stick between the new and previous tag to collect the metric "recent growth." We counted the total number of leaves at the start, and every three weeks we counted the number of new leaves produced on the recent growth portion of the vine. We also counted flowers and fruits on the two vines. For quantifying symptom and damage severity, we used the same rating scale as for the greenhouse. We rated discoloring (mottling and yellowing), distortion, and senescence severity. The scale: 0 =

no symptoms, 1 = slight damage/signs of symptoms, 2 ≤ 20% leaf area with the symptom, 3 = 21%–50% leaf area covered, and 4 ≥ 50% leaf area covered. For analysis, the severities of all the symptoms were summed into one metric: severity index.

The leaves were sampled mid-season to confirm infections with RT-PCR. Sampled tissue was stored at -80°C until processing. We tracked performance metrics for 15 *C. foetidissima*, with 14 confirmed infections for 2020 and one infection confirmed in 2019 (the extraction failed for 2020 but infections were retained (Chapter 2]). We tracked performance metrics for 19 *C. palmata*, with 15 2020 confirmed infections and three confirmed in 2019. Only one plant of this group did not have a CABYV detection.

Effects of CABYV on host suitability for aphid vectors

We used the electrical penetration graphing (EPG) technique to test for the effects of CABYV infection and plant species on aphid vector feeding behavior. EPG is an established technique that uses 12.5-micron sized gold wire attached to an aphid with silver glue, which is further connected to a DC current probe to visualize feeding waveforms at any given moment (Tjallingii & Esch, 1993). Aphids were tethered under a microscope while feeding and were gently pulled from the leaf and held for 10-30 minutes before being used in the experiment. Experiments were conducted inside a Faraday cage located in a climate-controlled room at 24 ± 1 °C. We recorded feeding behavior of *A. gossypii* apterous adults on infected and sham-inoculated individuals of both plant species at four weeks post inoculation (*C. foetidissima*: N=10 CABYV, 10 SHAM; *C. palmata*: N=10 CABYV, 7 SHAM). Waveforms corresponding to stylet

activities were recorded for eight hours using the PROBE 3.5 software (EPG Systems, www. epgsystems.eu) .Waveforms were manually assigned to distinct being analyzed by EPG-Calc (Giordanengo, 2014).

Virus detection

Greenhouse- and field-collected leaves

We used ~50-100mg of leaf tissue for virus detection. Tissues were distributed to 2mL Eppendorf tubes with two stainless steel grinding balls (4mm from Spex SamplePrep), positioned in an aluminum block cooled in liquid nitrogen, then homogenized in the Geno/GrinderR (SPEX SamplePrep) for one minute at 1100 RPM. RNA was extracted from each sample using TRI Reagent (Sigma-Aldrich), following the manufacturer's protocol for tissue and using a LegendMicro21R microcentrifuge (ThermoScientific) at 4ºC for all steps. At the RNA precipitation step, we added 250µL isopropanol and 250µL of 0.8M sodium citrate and 1.2M sodium chloride solution. Following extractions, RNA was purified to remove excess polysaccharides and other inhibitors that are frequent contaminants accompanying nucleic acid extractions from wild plants (Lacroix et al., 2016). For purifications, we added 500µL water to each sample, followed by 50µL 3M sodium acetate (pH 5.2) and 500µL of room temperature isopropanol. Samples were mixed well and incubated at room temperature for 20 minutes. Next, we pelleted the RNA by centrifuging at 12,000 x g and 4°C, then washed the pellet with 500µL ice cold 70% ethanol and centrifuged each sample for 30 seconds at 7,500 x g, twice. After carefully removing the ethanol, the pellets were air-dried for 15

minutes, then resuspended in 50µL Ultrapure water. In order to facilitate dissolving the pellet, we incubated the samples for 10 minutes at 55°C before storing in -80°C or proceeding with reverse-transcription PCR. Greenhouse tissues were not purified because overall RNA yields were low and further cleanup might have resulted in sample losses. RT-PCR was performed on 51 extractions from infected plants (all with sufficient quality for RT-PCR) and a randomly sampled subset of sham-inoculated plants (13).

For reverse transcription, reaction components included Ultrapure water, 10mM dNTP mix, random hexamers, 100mM DTT, and 5x SSIV Buffer. We used an enzyme (Superscript IV, Invitrogen) that is robust against most plant-derived inhibitors, and added an RNase inhibitor (Ribolock, Thermo Fisher) to prevent RNA degradation. We used random hexamers in the reverse transcription instead of gene-specific primers because this method had the greatest sensitivity in trial reactions. cDNA products can also be used to detect any RNA virus present in the sample. The subsequent PCR reactions (20 µl reaction volume) consisted of 1 µl of template cDNA, 4 µl of 5X HF buffer (Thermo Fisher Scientific or NEB), 1 µl of each 10 µM primer, 2 µl of dNTP mix (2 mM each) and 0.2 µl of Phusion DNA polymerase (Thermo Fisher Scientific & NEB). One PCR protocol was programmed as 98°C for 3 min for initial denaturation, followed by 40 cycles of 98°C for 10 s, 55°C for 30 s, 72°C for 1 min, then 72°C for 10 min. CABYV was detected using coat protein specific primers to amplify a 600 base pair amplicon. The forward primer (5'-3') is ATGAATACGGCCGCGGCTAGAAATC, and reverse (5'-3') is CTATTTCGGGTTCTGGACCTGGCA (Choi et al., 2015). Virus

detection was confirmed by visualizing bands using gel electrophoresis in a 1.5% agar gel with SYBR Safe DNA gel stain (ThermoFisher Scientific).

EPG experiment virus detection

To determine treatments for EPG experiments, we performed an enzyme-linked immunosorbent assays using a double antibody sandwich ELISA CABYV kit from Nano Diagnostics (http://www.nanodiaincs.com/A-CABYV.htm). The ~100mg from each leaf was distributed to 2mL Eppendorf tubes with two stainless steel grinding balls (4mm from Spex SamplePrep), positioned in an aluminum block cooled in liquid nitrogen, then homogenized in the Geno/GrinderR (SPEX SamplePrep) for two minutes at 1100 RPM. We followed the manufacturer's protocol for the kit before testing optical density values.

Statistics

The greenhouse plant fitness and symptom progression statistical analyses included two different methods. For data collected only at the end point (root and shoot masses, powdery mildew infection severity), we used the Wilcoxon Rank Sum Test for nonparametric data to compare CABYV and sham-inoculated treatments within each host species. For metrics repeatedly recorded throughout the experiment, we used a repeated measures generalized linear mixed effects model with the symptom or metric measured as a response, treatment x weeks post inoculation (WPI) as a factor, and plant individual as the repeated factor, using a Poisson distribution with Log specified as the link function in the model. We used the R library *lme4* (Bates et al., 2013) to run the model and used the R library *car* (Fox and Weisberg, 2019) to test the model for significance with an

analysis of deviance table. Post hoc tests were performed with the R library *emmeans* (Lenth, 2019) with a false discovery rate *P*-value adjustment.

For field-recorded data on disease symptoms and performance metrics, we used a generalized linear mixed effects model to test for differences in performance variables based on the factor severity index ("Sum_D" or sum of D symptom severity). The response variables include growth in centimeters since last time point, new leaves since last time point, fruit at the time, and flowers at the time. The random effect in this model is the plant individual because each plant was sampled twice since two vines were recorded from each plant. We used the R library *lme4* to run the model and used the R library *car* to test the model for significance with an analysis of deviance table, and R library *MuMIn* (Bartón, 2020) to obtain model R^2 values. We chose to use a Poisson distribution because the data were skewed with many zeros.

For preliminary EPG data, we used two different models depending on the variable type. All variables had non-normal distributions. The continuous variables we analyzed include duration in pathway phase, salivation into the sieve element, ingestion from phloem, and sustained ingestion from phloem (ingestion events greater than ten minutes). For these variables, we used a Gamma (link = "inverse") distribution in our model. The number of probes an aphid performed are count data for which we used a Poisson (link="log") distribution. Data were analyzed with a generalized linear model (R library *lme4* [Bates et al., 2013]), with the feeding behavior as the response variable and plant species, treatment, and plant species x treatment as factors, and distributions as

described for each variable type. When a significant effect was detected, we used a pairwise comparison using estimated marginal means with R library *emmeans* (Lenth, 2019) and a *P*-value adjustment with false discovery rate to test for differences between pairs . One outlier was removed from the dataset that had three times more probes than any other aphid. In total, we compared ten infected and nine sham *C. foetidissima*, and ten infected and seven sham *C. palmata*. Ideally, EPG experiments have 20 of each treatment, but these analyses are performed to demonstrate the method and show preliminary results.

Results

CABYV effects on plant performance under greenhouse conditions

For the greenhouse experiments, RT-PCR CABYV detections matched respective treatment groups. A summary of the *P* values determined by statistical analyses of greenhouse experiments is reported in Table 4.1. Shoot mass for both species was significantly decreased for CABYV-infected plants, and root mass was significantly decreased for CABYV-infected *C. foetidissima* (Figure 4.2I, 4.2J). *Cucurbita palmata* root mass also was reduced by infection, but this was not statistically significant, *P*= 0.059 (Figure 1J). There was no difference in powdery mildew severity due to infection status for *C. foetidissima*, but CABYV-infected *C. palmata* plants had significantly less severe powdery mildew compared to sham-inoculated plants (p=0.0071) (Figure 4.2K). There were no statistically significant differences between total leaves of infected vs. sham-inoculated plants for either species (Figure 4.2G, 4.2H). For mottle symptom

severity, *C. foetidissima* plants infected with CABYV had significantly greater mottling than sham-inoculated plants at two weeks post inoculation, whereas infected *C. palmata* were more mottled at all time points (Table 4.1, Figure 1C,1D). Yellowing severity for *C. foetidissima* was significantly higher for CABYV infection only at eight WPI, but was significantly higher at all time points for *C. palmata* (Table 4.1, Figure 4.2E, 4.2F). Leaf distortion was significantly higher for CABYV-infected *C. foetidissima* at two, four, and eight WPI and at two and four WPI for *C. palmata* (Table 4.1, Figure 4.2A, 4.2B).

Plant species	Metric	P-value at different time points			
		2 WPI	4 WPI	6 WPI	8 WPI
C. foetidissima	Mottle	**0.006**	0.99	0.105	0.99
	Yellows	0.052	0.005	0.222	**0.001**
	Distortion	**0.0001**	**0.031**	0.089	**0.002**
	Total leaves	0.957	0.5172	0.599	0.214
	Powdery mildew	NA	NA	NA	0.934
	Shoot mass	NA	NA	NA	**0.001**
	Root mass	NA	NA	NA	**0.015**
C. palmata	Mottle	**0.001**	**<.0001**	**<.0001**	**<.0001**
	Yellows	**0.0001**	**<.0001**	**<.0001**	**<.0001**
	Distortion	**<.0001**	**0.0001**	0.457	0.184
	Total leaves	0.451	0.118	0.077	0.065
	Powdery mildew	NA	NA	NA	**0.007**
	Shoot mass	NA	NA	NA	**0.025**
	Root mass	NA	NA	NA	0.06

Table 4.1. P-*values of metrics tested for this experiment comparing sham- and virus-inoculated plants. Symptom progression metrics include mottle, yellow, and distortion severity. Shoot and root mass serve as proxies for plant fitness. Powdery mildew is a disease severity metric, but only measured at one time point. Bold numbers indicate significant differences between sham and virus-inoculated plants.*

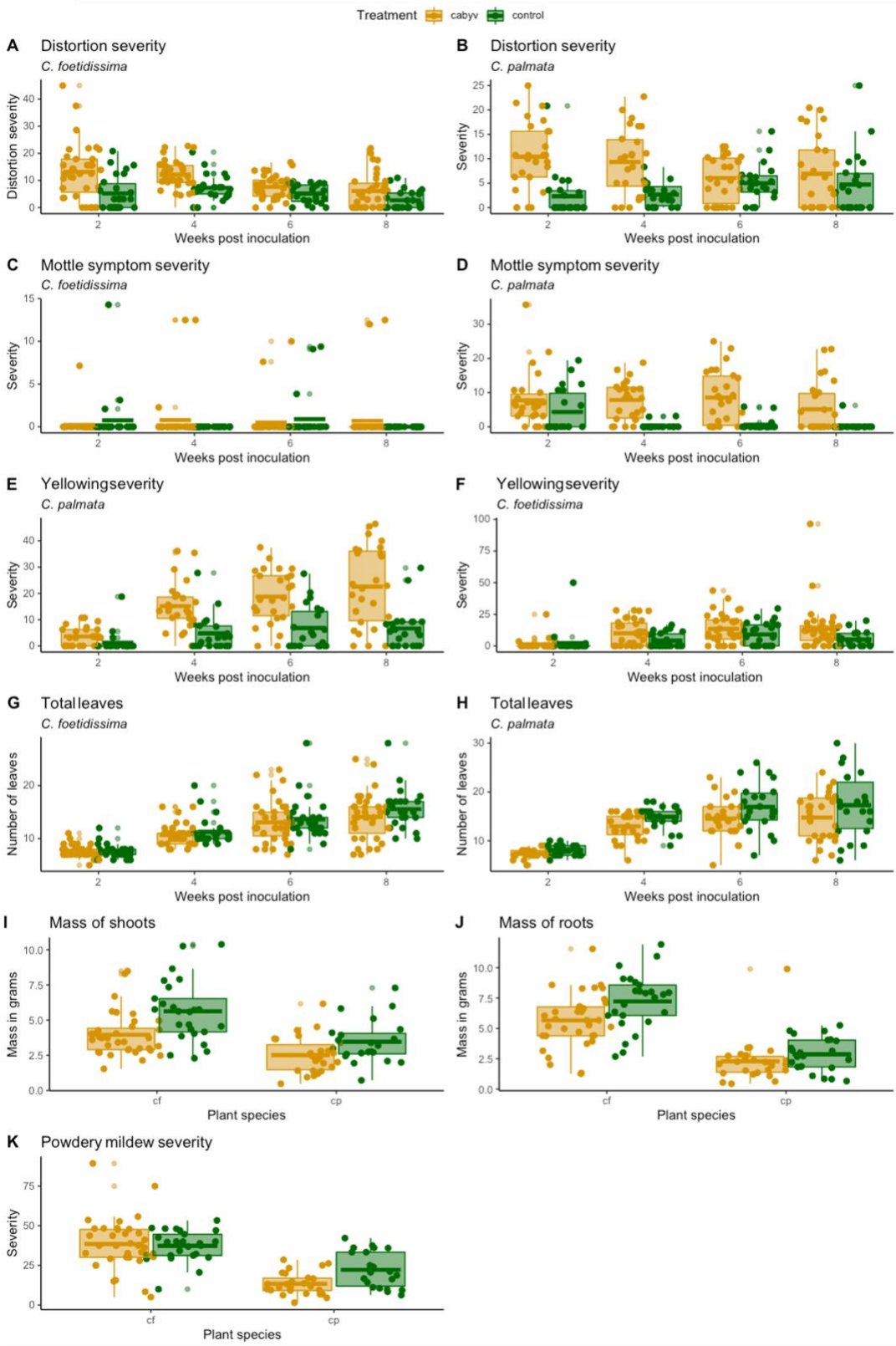
Treatment
cabyv
control
A
Distortion severity
C. foetidissima
Distortion severity
Weeks post inoculation
B
Distortion severity
C. palmata
Severity
Weeks post inoculation
C
Mottle symptom severity
C. foetidissima
Severity
Weeks post inoculation
D
Mottle symptom severity
C. palmata
Severity
Weeks post inoculation
E
Yellowing severity
C. palmata
Severity
Weeks post inoculation
F
Yellowing severity
C. foetidissima
Severity
Weeks post inoculation
G
Total leaves
C. foetidissima
Number of leaves
Weeks post inoculation
H
Total leaves
C. palmata
Number of leaves
Weeks post inoculation
I
Mass of shoots
Mass in grams
cf
cp
Plant species
J
Mass of roots
Mass in grams
cf
cp
Plant species
K
Powdery mildew severity
Severity
cf
cp
Plant species

Figure 4.2. *Plots for all metrics measured for greenhouse experiment. 4.2A-4.2H are plots for continuously recorded metrics. Figures 4.2I, 4.2J, and 4.2K are for endpoint metrics. Plots created using R libraries* ggplot2 *(Wickham et al., 2021A) and* ggpubr *(Kassambra, 2020).*

Disease severity effects on performance metrics in the field

For June 12 data, we found that the symptom severity index was a significant predictor of new leaf numbers since the last time point for *C. palmata* (*P*=0.0344) (Table 4.2). Symptom severity index was not a significant predictor of any other plant health variables on the June 12 date. For the July 3 dataset, the symptom severity index was a significant predictor for centimeters of vine growth since the last time point for *C. palmata* (*P*=0.0026)(Table 4.2). For the July 23 dataset, the symptom severity index was a significant predictor of new leaves for *C. foetidissima* (*P*=0.0142) (Table 4.2) and centimeters of growth since last time point for both *C. foetidissima* (*P*=0.004) (Table 4.2) and *C. palmata* (*P*=0.0142) (Table 4.2). Symptom severity index was not a significant predictor of fruit or flower number at any time point (*P*>0.05). R^2M values show the variance of the model explained by the fixed effects (severity index). The R^2C values show the variance of the entire model including the fixed effects (severity index) and random effects (plant individual).

Date	Metric	*C. foetidissima*			*C. palmata*		
		P-value	R^2M	R^2C	*P*-value	R^2M	R^2C
June 12, 2021	New growth (cm)	0.5501	0.0005	0.9879	0.6657	0.0009	0.9837
	New leaves	0.937	0.0002	0.4618	**0.0344**	**0.1587971**	**0.562**
	Flower count	0.2085	0.4271	0.5373	0.9182	0.0002	0.1782
	Fruit count	0.5458	0.06755	0.1006	0.1821	0.1607	0.2065
July 3, 2021	New growth (cm)	0.5371	0.0067	0.9783	**0.0026**	**0.3254**	**0.9645**
	New leaves	0.677	0.00855	0.7982	0.386	0.0351	0.8167
	Flower count	0.5359	0.0158	0.4172	0.1629	0.0194	0.1141
	Fruit count	0.8186	0.0033	0.4212	0.8405	0.0002	0.0002
July 23, 2021	New growth (cm)	**0.0042**	**0.061**	**0.9566**	**0.0182**	**0.1148**	**0.905**
	New leaves	**0.0142**	**0.1746**	**0.627**	0.587	0.0012	0.6774
	Flower count	0.51	0.5889	1	0.323	0.0509	0.4562
	Fruit count	0.4611	0.0146	0.5387	0.1488	0.0986	0.1629

Table 4.2. P-*values and* R^2 *values of symptom severity index as a factor for plant fitness proxies: centimeters of growth and new leaves since last time point, and fruit and flowers at each time point.* R^2M *is the marginal* $GLMM^2$ *value that represents the variance explained by the fixed effects.* R^2C *is the conditional* $GLMM^2$ *value that represents the variance explained by the entire model, including both fixed and random effects. Significant* P-*values and corresponding* R^2 *values are bolded.*

Effects of CABYV on host suitability for aphid vectors

We did not detect an effect of CABYV infection or plant species on the length of time aphids spent salivating (E1) or ingesting sap (E2)(Table 4.3). There also were no differences between infection treatments or plant species for the duration of sustained ingestion (ingestion events that lasted longer than ten minutes) (Table 4.3). Regardless of infection treatment, aphids spent significantly less time in pathway (“C”) phase when feeding on *C. palmata* than aphids feeding on *C. foetidissima* (*P*=0.0002775). Species

and treatment also were significant factors for the number of probes (n_Pr) aphids performed. Aphids on *C. foetidissima* performed significantly more probes than those on *C. palmata* (P=0.0196). On *C. palmata*, aphids on CABYV-infected plants made fewer probes than those on sham-inoculated plants (P=0.0045). There was no statistically significant effect of infection on the number of probes for aphids feeding on *C. foetidissima*. The differences between groups are visualized in Figure 4.3.

EPG Classes	Behavior	*C. foetidissima*		*C. palmata*	
		CABYV	Sham	CABYV	Sham
		N=10	N=9	N=10	N=7
General probing behavior Pr	No. of probes	12.7 ± 2.68	21.7 ± 7.89	8.7 ± 1.23	13.29 ± 3.4
	Duration of probing	449.53 ± 9.49	395.96 ± 43.24	416.23 ± 1 42.52	423.19 ± 26.19
Pathway phase C Phase	No. of pathway phases	13.4 ± 2.78	23.1 ± 7.9	9.6 ± 2.27	15.28 ± 4.17
	Duration of pathway (minutes)	197.76 ± 52.33	185.02 ± 44.8	56.93 ± 13.19	87.07 ± 23.45
Phloem phases E1 – Salivation E2– Ingestion	No. of phloem salivation events	1.3 ± 0.3	1.8 ± 0.36	1.9 ± 0.41	2.29 ± 0.75
	Duration of phloem salivation (minutes)	10.8 ± 6.29	7.23 ± 3.32	16.11 ± 8.78	4.12 ± 1.94
	No. aphids perform salivation	8	9	10	7
	No. of phloem ingestion events	1.2 ± 0.33	1.6 ± 0.27	1.6 ± 0.34	2 ± 0.65
	Duration of phloem ingestion (minutes)	321.19 +/- 30.52	257.47 +/- 50.79	366.52 +/- 23.57	282.45 +/- 50.2
	No. aphids perform ingestion	7	9	10	7
	No. of phloem ingestion events, sustained	1 ± 0.26	1.4 ± 0.27	1.4 ± 0.27	1.86 ± 0.55
	Duration of phloem ingestion sustained(minutes)	329 ± 30.5	256.47 ± 51.22	365.97 ± 23.64	281.06 ± 50.83
	No. aphids perform ingestion, sustained	7	9	10	7
Xylem and stylet incompatibility	No. of aphids with xylem ingestion	1	3	1	2
	No. of aphids with stylet incompatiblity	2	4	2	3

Table 4.3. *Selected feeding behavior variables (average across aphids ± SEM). Feeding behaviors selected for analysis are associated with host acceptance and successful feeding, which are directly related to CABYV transmission. Behaviors are separated into the number of times events occurred, and the duration of those events. For salivation, ingestion, sustained ingestion, xylem ingestion, and stylet incompatibility, the number of aphids that performed those events are included.*

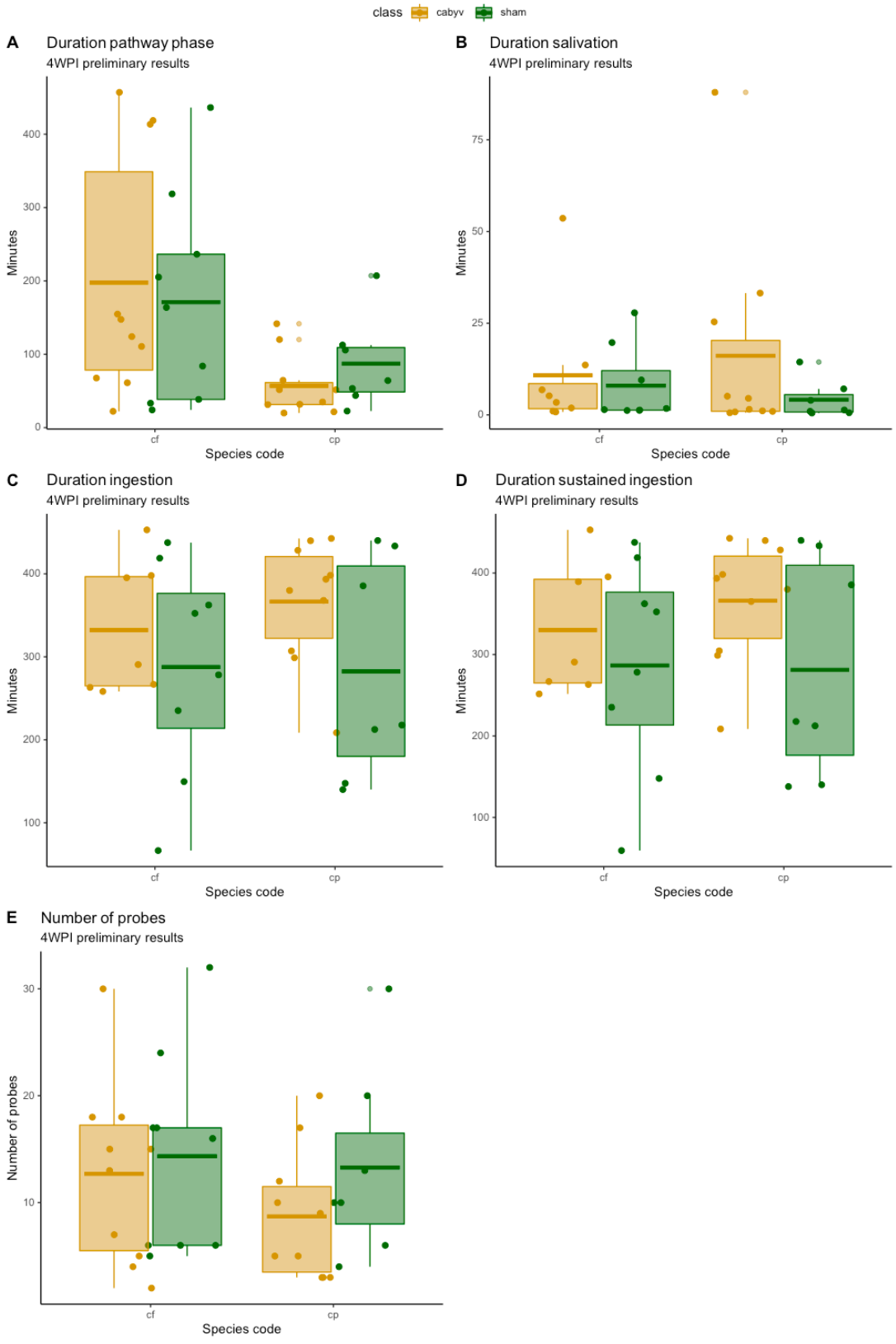
class
cabyv
sham
A
Duration pathway phase
4WPI preliminary results
Minutes
Species code
cf
cp
B
Duration salivation
4WPI preliminary results
Minutes
Species code
C
Duration ingestion
4WPI preliminary results
Minutes
Species code
D
Duration sustained ingestion
4WPI preliminary results
Minutes
Species code
E
Number of probes
4WPI preliminary results
Number of probes
Species code

Figure 4.3. *Feeding behaviors by treatment (sham and virus) and plant species. The two treatments ("class") are CABYV-infected and sham-inoculated. The continuous variables are presented in minutes. The duration of the recording is eight hours*.

Discussion

Crop-associated viruses are widespread in wild plant systems, but there have been few attempts to understand the effects of these viruses on their uncultivated host plants. Most studies to date have focused on annual plants, even though perennials dominate wild systems. Additionally, most studies on virus effects in wild hosts were done on just one pathogen complex (barley and cereal yellow dwarf viruses). Here, we address these knowledge gaps through a combination of field, greenhouse, and laboratory studies documenting the impacts of a crop-associated virus (CABYV) on growth, reproduction metrics, and host-vector interactions for two key dicot perennials common in drought prone chaparral habitat. We hypothesized that i. there would be negative effects of infection on metrics relevant for plant fitness, ii. there would be changes over time in disease symptoms, iii. virus infection would result in vector feeding behaviors conducive to virus transmission. Our data suggest that infection negatively affects the above ground and below ground growth of both cucurbit species, but that infection symptoms are more severe, and progress more rapidly in *C. palmata.* Additionally, vector feeding experiments revealed that *C. palmata* is more palatable to aphid vectors and that CABYV infection modifies some vector feeding behaviors in ways that are conducive to transmission. We also found an unexpected benefit of infection for *C. palmata* in the form of protection from a fungal pathogen.

In controlled greenhouse experiments, CABYV infection negatively affected growth of the two cucurbits, but this effect varied by species. Both species showed symptoms throughout the course of the experiment (Fig. 4.6). However, *C. palmata* symptoms were apparent sooner and progressed more rapidly in severity (Table 4.1). Both *C. palmata* and *C. foetidissima* experienced significant reductions in aboveground tissue, but *C. foetidissima* also had significantly reduced root tissue (Table 4.1). *Cucurbita palmata* root mass also was reduced, but this was not statistically significant (P= 0.059). Both plants senesce in winter, regrowing from large tap roots in the spring. These tap roots are modifications that allow survival in semiarid and xerophytic lands, and are important for water collection and storage (Bemis & Whitaker, 1969). It is therefore reasonable to speculate that CABYV infection in first or second year plants may reduce the likelihood of perennial cucurbit seedlings surviving to a stage where they can persist for multiple years. There is no prior research on *C. palmata* disease resistance or responses to pathogen infection. Previously, *C. foetidissima* had been tested for virus resistance and effects on root starch content and seed yields (Rosemeyer, 1982). A significant relationship between symptoms and seed yield reduction was found, but sample sizes in this study were low and infections included multiple pathogens of uncertain identities (Rosemeyer, 1982). Efforts to assess *C. foetidissima* as a species for cultivation (for seed oils) included some assessments of resistance to common crop-associated viruses, including *Cucumber mosaic virus*, *Tobacco ringspot virus*, *Watermelon mosaic virus*, and *Bean yellow mosaic virus*. Surprisingly, *C. foetidissima* had some resistance to these pathogens, making it a good candidate for cultivation based

on disease resistance profile (DeVeaux & Shultz, 1985). Even though this plant is considered disease resistant, our results show that infection with CABYV produces smaller plants.

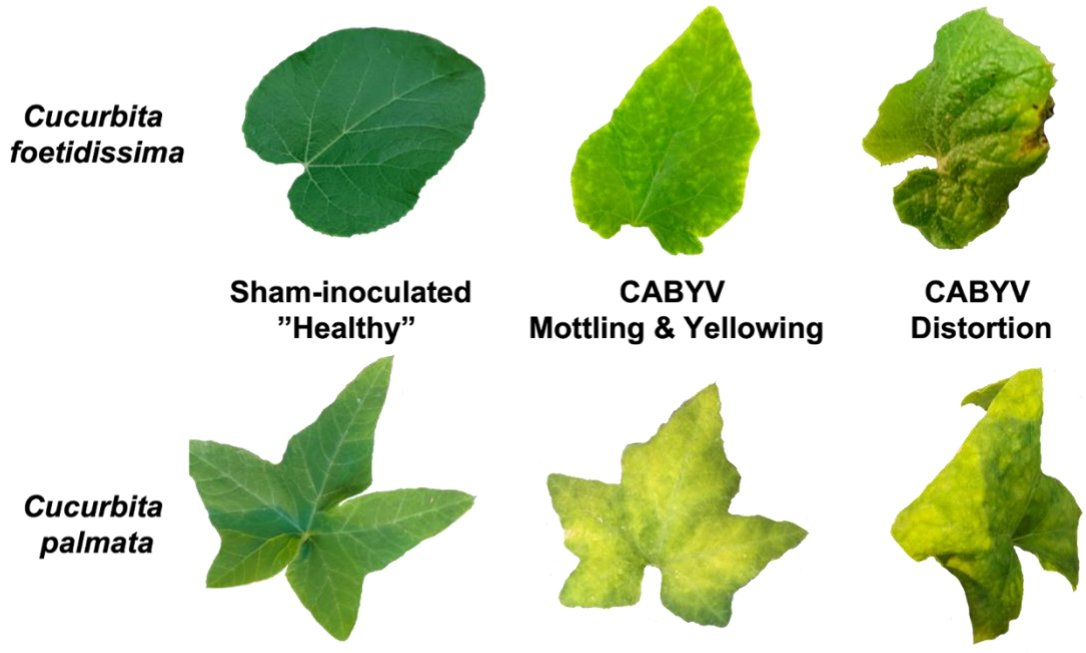

Figure 4.4. *Symptoms of CABYV in* C. foetidissima *and* C. palmata*. Photographs were taken at various time points within two to eight weeks post inoculation and reflect the most obvious examples of the stated symptoms.*

Late in the experiment, plants became naturally infected with powdery mildew (a common issue in cucurbit greenhouse cultivation). We took advantage of this invading pathogen to evaluate the interactions between powdery mildew and infected and sham-inoculated hosts of each species. Symptom severity of powdery mildew, measured as leaf area covered, differed between plant species. There was no difference in severity of powdery mildew between *C. foetidissima* virus-infected and sham-inoculated treatments.

However, CABYV-inoculated *C. palmata* had less severe powdery mildew infection. Previous research on *Zucchini yellow mosaic virus* (ZYMV) shows that viral infection limits powdery mildew infection in wild *Cucurbita pepo,* likely through virus effects on production of salicylic acid in the host (Harth et al., 2018). Powdery mildew is also prevalent in wild cucurbits in California (personal observation). Resistance to this common pathogen may be a benefit of virus infection, similar to delayed progression of bacterial wilt symptoms conferred by ZYMV infection in *C. pepo* (Shapiro et al., 2013). Thus, even though CABYV infection appears to have negative effects, there may be conditions (such as presence of competing pathogens) that virus infection confers benefits to hosts (Roossinck, 2011).

To complement greenhouse studies, we performed field observations of mature plants to explore relationships between expression of CABYV symptoms and other metrics of plant health. Even though plants had previously been identified as having CABYV infection, symptoms were not apparent in the early stages of regrowth (observations in May 2020, not analyzed here). As the season progressed, more typical (and severe) symptoms became apparent. Most of the plants (save for one) we monitored tested positive for CABYV in the season in which we performed observations. However, other virus infections can be unapparent and co-infections are common - especially in this plant system (Prendeville et al., 2012; Shates et al., 2019). It is likely that there are other disease pressures on these plants, as well as environmental factors, and plant genetic factors, that we were not able to account for in this study or our statistical models. We found that the severity index was a significant predictor for vegetative growth (leaves,

vine length) but not reproductive outputs (fruits, flowers). A relationship between CABYV symptom severity and growth was especially apparent for *C. palmata*; the severity index was negatively correlated with new leaves on June 12 and negatively correlated with vine growth on July 3, and July 23 (Table 4.2). For *C. foetidissima*, most time points showed no significant correlation between symptom severity and growth with the exception of July 23, for which the severity index was *positively correlated* with new leaves and vine growth (Table 4.2). From these models, we can see that the two cucurbit species perform differently across the season, and respond differently to the cumulative symptoms they experience. Consistent with greenhouse experiments, our data also show that at the earliest time point, relative to data for *C. foetidissima*, more *C. palmata* individuals have measurable symptoms and fall into the moderate to high symptom expression categories (Table 4.2). Collectively, these results corroborate our greenhouse observations of negative effects of CABYV infection on the growth of first year plants of each species, with more severe effects occurring for *C. palmata*. The model including severity index did not explain all of the variation, with differences between plant individuals contributing more to our models (Table 4.2). It is surprising that we did not observe effects on fitness, but we only measured two proxies of this (fruit and flower number). Only a few individuals at each time point were producing fruits or flowers to count on the vines chosen for observation. Additional data, perhaps at the whole plant level and over multiple seasons, would clarify whether effects on growth scale up to influence fitness.

Another way that virus infection can influence plant performance is through effects on interactions with other species, including the insect herbivores that transmit plant viruses (Mauck et al., 2018). Studies in crop systems provide evidence that persistently transmitted viruses (including CABYV) can improve palatability and quality of cucurbit crop hosts for aphid vectors in ways that enhance the probability of virus acquisition (Bosque-Pérez & Eigenbrode, 2011; Carmo-Sousa et al., 2016). We have observed large infestations of the cotton-melon aphid (*Aphis gossypii*), and rarely the green peach aphid (*Myzus persicae*), on wild growing *Cucurbita foetidissima* and *C. palmata*. These aphids are the two known vectors of CABYV (Lecoq et al., 1992). CABYV transmission is favored by longer phloem sap ingestion for acquisition and longer sieve element salivation periods for transmission (Martín et al., 1997; Prado & Tjallingii, 1994) and an EPG study using cucumber (*Cucumis sativus*) found that *A. gossypii* phloem access and sap uptake are improved by CABYV infection (Carmo-Sousa et al., 2016). In our EPG experiments, we found that CABYV infection did not alter pathway phase duration (time to reach the phloem), duration of salivation into the phloem, phloem sap ingestion, or sustained phloem sap ingestion duration. However, aphids feeding on CABYV-infected *C. palmata* plants performed fewer surface-level probes, which indicates more rapid host acceptance. Additionally, compared to *C. foetidissima*, aphids also performed fewer probes on *C. palmata* regardless of infection status and spent less time in the pathway phase.

These data suggest aphids can feed more easily and reach the phloem more rapidly on *C. palmata* and that CABYV infection may further improve *C. palmata*

palatability and suitability. It is interesting that this occurs only for the species that, according to our other experiments, appears to be more susceptible to CABYV infection. In crops, virus effects on host suitability for vectors are also more pronounced in varieties that are more susceptible to infection overall (Legarrea et al., 2015; Mauck et al., 2014; Rajabaskar et al., 2013). The vector feeding behavior on this species and greater ease of feeding on infected hosts may explain the proportionally higher prevalence of CABYV infecting *C. palmata* at Motte Rimrock (Chapter 3). And we observed that when aphid infestations reach high levels, plants prematurely senesce.

Overall, by combining manipulative experiments with field observations, this study improves our understanding how crop-associated viruses impact perennial host performance and interactions with vector insects. Contrary to conventional perceptions of virus infections in wild hosts, we found that CABYV-infected wild *Cucurbita* do exhibit symptoms of infection, including mottling, yellowing, and leaf distortion. These symptoms are apparent in both greenhouse and field environments, on young and mature hosts. In the field, symptoms of infection (from CABYV and likely other pathogens or environmental factors), are correlated with increased growth in *C. foetidissima* and decreased growth in *C. palmata*. Our preliminary insect behavior data shows that *Aphis gossypii* feeding on infected *C. palmata* spends less time searching for phloem elements (pathway phase). Future work will focus on gathering more feeding behavior data to understand how virus infection may predispose plants to infestation with aphid vectors and facilitate transmission conducive feeding behaviors. These results contribute to the

expanding field of plant virus ecology, especially our knowledge of interactions within perennial wild plant systems.

References

Alexander, H. M., Mauck, K. E., Whitfield, A. E., Garrett, K. A., & Malmstrom, C. M. (2014). Plant-virus interactions and the agro-ecological interface. *European Journal of Plant Pathology, 138*(3), 529–547.

Bananej, K., Vahdat, A., Predajna, L., & Glasa, M. (2009). Molecular characterization of geographically different cucurbit aphid-borne yellows virus isolates. *Acta Virologica*, *53*(1), 61–64.

Bartón, K., 2020. MuMIn: Multi-Model Inference. Version 1.43.17. https://cran.r-project.org/web/packages/MuMIn/index.html

Bates D, Maechler M, Bolker B. 2013. lme4: Linear mixed-effects models using S4 classes. *in*. http://CRAN.R-project.org/package=lme4

Bemis, W. P., & Whitaker, T. W. (1969). The xerophytic Cucurbita of northwestern Mexico and the southwestern United States. *Madroño*, *20*(2), 33–41.

Bosque-Pérez, N. A., & Eigenbrode, S. D. (2011/8). The influence of virus-induced changes in plants on aphid vectors: Insights from luteovirus pathosystems. *Virus Research*, *159*(2), 201–205.

Carmo-Sousa, M., Moreno, A., Plaza, M., Garzo, E., & Fereres, A. (2016). Cucurbit aphid-borne yellows virus (CABYV) modifies the alighting, settling and probing behaviour of its vector *Aphis gossypii* favouring its own spread. *The Annals of Applied Biology*, *169*(2), 284–297.

Choi, S.-K., Yoon, J.-Y., & Choi, G.-S. (2015). Biological and molecular characterization of a korean isolate of *Cucurbit aphid-borne yellows virus* infecting *Cucumis* species in Korea. *The Plant Pathology Journal / the Korean Society of Plant Pathology*, *31*(4), 371–378.

Desbiez, C., Verdin, E., Moury, B., Lecoq, H., Millot, P., Wipf-Scheibel, C., Mirzayeva, S., Sultanova, N., Balakishiyeva, G., Mammadov, A., Kheyr-Pour, A., & Huseynova, I. (2019). Prevalence and molecular diversity of the main viruses infecting cucurbit and solanaceous crops in Azerbaijan. *European Journal of Plant Pathology*, *153*(2), 359–369.

DeVeaux, J. S., & Shultz, E. B. (1985). Development of buffalo gourd (*Cucurbita foetidissima*) as a semiaridland starch and oil crop. *Economic Botany*, *39*(4), 454–472.

Fox, J., and Weisberg, S. (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Giordanengo, P. (2014). EPG-Calc: a PHP-based script to calculate electrical penetration graph (EPG) parameters. *Arthropod-Plant Interactions*, *8*(2), 163–169.

Harth, J. E., Ferrari, M. J., Tooker, J. F., & Stephenson, A. G. (2018). *Zucchini yellow mosaic virus* infection limits establishment and severity of powdery mildew in wild populations of *Cucurbita pepo*. *Frontiers in Plant Science*, *9*, 792.

Jones, R. A. C. (2020). Disease pandemics and major epidemics arising from new encounters between indigenous viruses and introduced crops. *Viruses*, *12*(12).

Kassambra, A., 2020. **ggpubr: 'ggplot2' Based Publication Ready Plots.** Version 0.4.0. https://cran.r-project.org/web/packages/ggpubr/index.html

Kassem, M. A., Sempere, R. N., Juárez, M., Aranda, M. A., & Truniger, V. (2007). *Cucurbit aphid-borne yellows virus* is prevalent in field-grown cucurbit crops of southeastern Spain. *Plant Disease*, *91*(3), 232–238.

Kenney, J. R., Grandmont, M.-E., & Mauck, K. E. (2020). Priming Melon Defenses with Acibenzolar-S-methyl attenuates infections by phylogenetically distinct viruses and diminishes vector preferences for infected hosts. *Viruses*, *12*(3).

Knierim, D., Deng, T. C., Tsai, W. S., Green, S. K., & Kenyon, L. (2010). Molecular identification of three distinct *Polerovirus* species and a recombinant *Cucurbit aphid-borne yellows virus* strain infecting cucurbit crops in Taiwan. *Plant Pathology*, *59*(5), 991–1002.

Kwak, H.-R., Lee, H. J., Kim, E.-A., Seo, J.-K., Kim, C.-S., Lee, S. G., Kim, J.-S., Choi, H.-S., & Kim, M. (2018). Complete genome sequences and evolutionary analysis of *Cucurbit aphid-borne yellows virus* isolates from melon in Korea. *The Plant Pathology Journal*, *34*(6), 532–543.

Lacroix, C., Renner, K., Cole, E., Seabloom, E. W., Borer, E. T., & Malmstrom, C. M. (2016). Methodological guidelines for accurate detection of viruses in wild plant species. *Applied and Environmental Microbiology*, *82*(6), 1966–1975.

Lecoq, H., Bourdin, D., Wipf-Scheibel, C., Bon, M., Lot, H., Lemaire, O., & Herrbach, E. (1992). A new yellowing disease of cucurbits caused by a luteovirus, cucurbit aphid-borne yellows virus. *Plant Pathology*, *41*(6), 749–761.

Legarrea, S., Barman, A., Marchant, W., Diffie, S., & Srinivasan, R. (2015). Temporal Effects of a begomovirus infection and host plant resistance on the preference and development of an insect vector, *Bemisia tabaci*, and implications for epidemics. *PloS One*, *10*(11), e0142114.

Lenth R. 2019. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.3., https://CRAN.R-project.org/package=emmeans

Malmstrom, C. M., & Alexander, H. M. (2016). Effects of crop viruses on wild plants. *Current Opinion in Virology*, *19*, 30–36.

MartÍn, B., Fereres, A., Tjallingii, W. F., & Collar, J. L. (1997). Intracellular ingestion and salivation by aphids may cause the acquisition and inoculation of non-persistently transmitted plant viruses. In *Journal of General Virology* (Vol. 78, Issue 10, pp. 2701–2705).

Mauck, K. E., Chesnais, Q., & Shapiro, L. R. (2018). Evolutionary determinants of host and vector manipulation by plant viruses. *Advances in Virus Research*, *101*, 189–250.

Mauck, K. E., De Moraes, C. M., & Mescher, M. C. (2014). Evidence of local adaptation in plant virus effects on host-vector interactions. *Integrative and Comparative Biology*, *54*(2), 193–209.

Prado, E., & Tjallingii, W. F. (1994). Aphid activities during sieve element punctures. *Entomologia Experimentalis et Applicata*.

Prendeville, H. R., Ye, X., Morris, T. J., & Pilson, D. (2012). Virus infections in wild plant populations are both frequent and often unapparent. *American Journal of Botany*, *99*(6), 1033–1042.

Rajabaskar, D., Ding, H., Wu, Y., & Eigenbrode, S. D. (2013). Different reactions of potato varieties to infection by *Potato leafroll virus*, and associated responses by its vector, *Myzus persicae* (Sulzer). *Journal of Chemical Ecology*, *39*(7), 1027–1035.

Roossinck, M. J. (2011). The good viruses: viral mutualistic symbioses. *Nature Reviews. Microbiology*, *9*(2), 99–108.

Rosemeyer, A. E. M. (1982). *Isolation and characterization of two viruses from Cucurbita foetidissima H.B.K., buffalo gourd*. (W. P. Bemis (ed.)) [Masters, University of Arizona].

Shapiro, L. R., Salvaudon, L., Mauck, K. E., Pulido, H., De Moraes, C. M., Stephenson, A. G., & Mescher, M. C. (2013). Disease interactions in a shared host plant: effects of pre-existing viral infection on cucurbit plant defense responses and resistance to bacterial wilt disease. *PloS One*, *8*(10), e77393.

Shates, T. M., Sun, P., Malmstrom, C. M., Dominguez, C., & Mauck, K. E. (2019). Addressing research needs in the field of plant virus ecology by defining knowledge gaps and developing wild dicot study systems. *Frontiers in Microbiology*, *9*, 3305.

Takeshita, M., Okuda, M., Okuda, S., Hyodo, A., Hamano, K., Furuya, N., & Tsuchiya, K. (2013). Induction of antiviral responses by acibenzolar-*S*-methyl against cucurbit chlorotic yellows virus in melon. *Phytopathology*, *103*(9), 960–965.

Tjallingii, W. F., & Esch, T. H. H. (1993). Fine structure of aphid stylet routes in plant tissues in correlation with EPG signals. *Physiological Entomology*, *18*(3), 317–328.

Tyack, N., Dempewolf, H., & Khoury, C. K. (2020). The potential of payment for ecosystem services for crop wild relative conservation. *Plants*, *9*(10). https://doi.org/10.3390/plants9101305

Wickham, H., Chang, W., Henry, L., Pederson, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D. (2021A). ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. Version 3.3.5. https://cran.r-project.org/web/packages/ggplot2/index.html

Zabel, F., Delzeit, R., Schneider, J. M., Seppelt, R., Mauser, W., & Václavík, T. (2019). Global impacts of future cropland expansion and intensification on agricultural markets and biodiversity. *Nature Communications*, *10*(1), 2844.

Conclusion

Wild, perennial plants have been historically overlooked in research on plant-virus-vector interactions. In this dissertation, I report that wild crop-wild relatives in the family Cucurbitaceae (*Cucurbita palmata* and *Cucurbita foetidissima*), and a co-occurring nightshade (*Datura wrightii*), are frequently infected with crop-associated viruses in three southern California reserves in Riverside/ San Diego counties. The virus communities are structured by space between reserves, differences between reserves themselves, and biological differences between hosts. During characterization of virus communities, I discovered that two crop-associated cucurbit viruses are particularly prevalent in these systems: *Cucurbit yellow stunting disorder virus* (CYSDV, whitefly transmitted) (Family *Closteroviridae*, genus *Crinivirus*) and *Cucurbit aphid-borne yellows virus* (CABYV, aphid-transmitted) (Family *Solemoviridae* [Previously *Luteoviridae*], genus *Polerovirus*). In multi-year sampling studies of the same individuals, I showed that one of these viruses, CABYV, is retained across multiple seasons. CABYV is also the most common virus detected in cucurbit hosts across all three reserves. This virus is common in cucurbit production areas in Europe, but not in the United States. Despite this, it is highly abundant in wild cucurbit populations. Using phylogenetic methods, I found that the wild-associated CABYV in the plants I sampled was likely introduced to California twice - once from Europe and once from Asia. In a broader phylogeographic analysis incorporating genes on either side of a recombination point, I also found evidence for a new putative CABYV origin in northern India. Recombination runs rampant in this virus genus, and I found evidence that there is

intraspecific recombination among the isolates I sampled, but not between different *Polerovirus* species, unlike in other regions that have CABYV co-occurring with other cucurbit poleroviruses. Lastly, I found that CABYV has negative effects on the two wild cucurbits I developed for my study system. This work shows the importance of virus discovery in new systems, as well as following up virome profiling with manipulative experiments to determine the impacts of crop-associated, insect-transmitted viruses on the ecosystems we strive to preserve.