Project title: Identifying novel viruses associated with the Saguaro cactus, *Carnegiea gigantea.*

Principal Investigator: Rafaela Salgado Fontenele, PhD student - email: rfontene@asu.edu

Grant: Student and Early Career Research grant

Affiliation: School of Life Sciences, Center of Fundamental and Applied Microbiomics, The Biodesign Institute, Arizona State University 1001 S. McAllister Ave Tempe, AZ 85287-5001

Disciplines: biological sciences, ecology

Introduction: The Saguaro cactus, *Carnegiea gigantea* (Engelm.) Britton & Rose*,* is an iconic symbol of the Sonoran desert. There has been extensive research undertaken on Saguaro cactus, investigating the physiology, reproduction, ecology and genomics [1]. In spite of this long history of research on saguaro, there is almost no information about the viral communities associated with Saguaro. To date, only one Saguaro infecting virus has been described, known as the Saguaro cactus virus (SCV). The SCV was originally identified in a flower in the early 1970s as part of a virus survey in the Saguaro National Monument in Arizona based in transmission electron microscopy [2, 3]. SCV is a small icosahedral virus with a single-stranded negative sense RNA genome of 3879 nucleotides classified into the genus *Alphacarmovirus,* in the family *Tombusviridae,* based on the capsid structure and genome size [4], which was further confirmed by sequencing and molecular analysis of the full genome [5]. SCV causes a systemic but asymptomatic infection in the Saguaro. However, it has been shown to cause mottling and vein chlorosis in mechanically inoculated *Chinopodium quinoa*, *C. amaranticolor* and *C. capitatum* in the plant family Amaranthaceae [3].Within other members of the family Cactaceae, a few viruses, which contain single-stranded RNA genomes, have previously been identified and those belong to the genus *Tobamovirus*, *Carlavirus*, *Potexvirus* and *Tospovirus* [6-12].

With the advent of high-throughput DNA sequencing (HTS) technologies, it is now possible to rapidly identify known and novel viruses within a sample. This approach has

led to a dramatic increase in the discovery of novel viruses across ecosystems and has broadly expanded what we know regarding plant-infecting viruses globally [13]. Gaining a greater understanding of plant-infecting viruses and their ecology is economically, agriculturally, culturally and ecologically important. Most efforts in plant virus research have been towards agriculturally important plants but very little research has focused on evaluating the impact of viruses on native plant biodiversity. In the case of those plant species that are rare and unique, like the Saguaro cactus, such information is essential for conservation, and informed management and cultivation practices.

Currently, as part of my PhD research, I am investigating the viral diversity of cacti from Arizona. This study has led to the discovery of known viruses as well as several novel ones. The viruses recovered to date are single-stranded DNA viruses belonging to the economically important plant-infecting virus family *Geminiviridae*. Two novel geminiviruses were isolated from *Opuntia* spp., *Cylindropuntia* spp. and *Lophocereus schottii*. Additionally, the known geminivirus, spinach curly top Arizona virus (SCTAV), which belongs to the genus *Bercurtovirus,* was identify in *Opuntia* spp. plants. SCTAV was first isolated in 2009 from a spinach plant displaying disease symptoms collected in a commercial field in Arizona [14]. Two geminiviruses from the genus *Begomovirus* were also identified in *Opuntia* spp. plants; those were the squash leaf curl virus and watermelon chlorotic stunt virus. (Fontenele et al., unpublished)

Spillover of viral pathogens from cultivated plants to uncultivated plants can have negative effects on native plants, highlighting the importance of viral surveys in these communities as well as in agro-ecological interface areas [15, 16]. The finding of known agricultural pathogens, such as the becurtovirus SCTAV and two species of begomovirus in cacti samples, validates the importance of the viral surveillance for the Cactaceae family, since they may have an impact on the cacti host biology. Identifying associated viruses can allow for a more informed framework for Saguaro conservation within an ecological framework.

Objective: The aim of this proposed research is to expand our current knowledge of the viral diversity associated with the iconic Saguaro cactus, *C. gigantea,* using a high throughput sequencing approach, thereby providing important baseline data on potential viral pathogens and gaining a better understanding of the viral ecology of cacti.

Material and methods: Professor Martin Wojciechowski of Arizona State University provided DNA samples of Saguaro from his lab collections, (collected with the appropriate permits, from the Desert Laboratory on Tumamoc Hill, Tucson, Arizona). Total DNA was isolated from 27 Saguaro samples using the CTAB protocol and circular nucleic acid was amplified preferentially by rolling circle amplification (RCA) using Phi29 polymerase. The RCA products of seven samples were pooled and sequenced on an Illumina HiSeq 4000 platform (paired-end 2x100 bp) at Macrogen Inc. (South Korea). The raw reads were *de novo* assembled using the programs metaSPAdes [17] and viral-like signatures were analyzed by BLASTx [18] against a local viral Refseq protein database compiled from the public database Genbank. Two contigs containing the full-length genome of viruses from the family *Genomoviridae* were identified. Analysis of the two genomoviruses genomes and encoded protein sequence pairwise identities were determined using SDT v1.2 [19] by comparison to a dataset of all genomoviruses present in GenBank. Maximum-likelihood phylogenetic analysis of the amino acid Rep gene sequence from the two identified genomoviruses and with those encoded by genomoviruses in GenBank was also carried out.

The remaining samples will be sent for high through sequencing in the near future. Nonetheless, the RCA of all the samples have been screened for viral sequences by PCR using KAPA HiFi HotStart DNA polymerase (Roche, USA)) and the following amplification protocol: 95°C [1 min], (94°C [15 sec], 60°C [15 sec], 72°C [3 min]) × 25 cycles, 72°C [3 min], 4°C [10 min] with specific primers developed for the novel and known geminiviruses we have previously identified in other cacti.

Results:

Based on our preliminary data (Fontenele, unpublished data), we expected to identify both novel and known viruses associated with Saguaro, especially since we had previously identified a novel virus in *Lophocereus schottii,* a species closely related to Saguaro [20]. Based on the BLASTx results from analyses of the assembled contigs from the initial seven sequenced Saguaro samples we did not identify any geminivirus-derived contigs. However, we recovered the full-length genome of two novel genomoviruses. Genomoviruses are a recently identified group of circular single-stranded DNA viruses that compose the family *Genomoviridae* [21, 22]. Currently the family is divided into nine genera: *Gemycircularvirus*, *Gemykibivirus*, *Gemygorvirus*, *Gemykolovirus*, *Gemykrogvirus*, *Gemyvongviru*s, *Gemytondvirus*, *Gemykroznavirus* and *Gemyduguivirus* [21]. Their genome are approximately 2 kb long and encode the capsid protein (CP) on the virion sense strand and the replication associated protein (Rep) on the complementary sense strand [21]. The majority of genomoviruses has been discovered through HTS approaches only and thus there is very little information associated with their biology. The first isolated genomovirus was *Sclerotinia sclerotiorum* hypovirulence DNA virus (SsHADV-1), a member of the species *Sclerotinia gemycircularvirus* 1. SsHADV-1 was shown to infect and cause hypovirulence in *Sclerotinia sclerotiorum*, a plant fungal pathogen [23]. It is likely that all genomoviruses infect fungi, however, this viruses have been recovered from a variety of different samples (i.e., fungi, plants, insects, birds, and mammals, and others were isolated from environmental samples collected in sediments, sewage, and wastewater) [21, 22] which may indicate they can infect other eukaryotes or fungi associate with those animals.

The genomoviruses recovered in this project were named “Plant-Associated Genomovirus 28 isolate Sag_SP218” - PGm28 (GenBank accession number MK947375) and “Plant-Associated genomovirus 29 isolate Sag_SP219” - PGm29 (MK947376), and have been deposited in GenBank. PGm28 and PGm29 genomes are 2181 and 2199 nucleotides in length, respectively, and both encode the expected open reading frames Rep and CP and the origin of replication characteristic of the viruses in the family *Genomoviridae*. PGm28 shares 70% genome-wide pairwise identity with a tortoise genomovirus 6 isolate Tor12_165 (MK570206), and 74% Rep and 68% CP amino acid

identity with a finch-associated genomovirus 3 isolate S30P_D (MK249305) and capybara genomovirus 12 isolate cap1_1725 (MK483084), respectively. PGm29 shares 63% nucleotide pairwise identity with plant-associated genomovirus 1 isolate 206_BA744 (MH939362), and 63.5% Rep and 47% CP amino acid identity with the giant panda-associated gemycircularvirus strain gpge014 (MF327571) and cattle blood-associated gemycircularvirus strain BGmv001 (MF669480), respectively.

Based on the current species demarcation for the family *Genomoviridae*, PGm28 and PGmV29 are novel species since they share <78% genome-wide identity with other genomoviruses [21]. [here I think you need to put this in some perspective – if they aren't that closely related to 'other genomoviruses' what are they related to?] Based on phylogenetic analysis of the Rep amino acid sequences, PGm28 and PGm29 cluster within the genus *Gemykibirus*. Although no known plant-infecting virus was identified in this study, the results are encouraging as the two novel genomoviruses are the first ssDNA viruses described associated with Saguaro plants. This project will continue to analyze more Saguaro samples in this collaborative project with Professor Martin Wojciechowski of Arizona State University. The number of Saguaro samples analyzed is still very low when compared with the number of *Opuntia* sp. samples previously analyzed in which we identify viruses of the family *Geminiviridae* at low incidence.

The results obtained in this project will be included in a manuscript for publication that is currently under development. Our lab has recovered several other genomoviruses from plant samples in different areas of the world that represent a broad diversity of plant families. This manuscript aim to provide a broader perspective of the correlation between plants and their associated genomoviruses that will hopefully provide more information regarding their interactions in vivo. N-Gen will be acknowledge in this publication and any other future research results with Saguaro-associated viruses.

References

1. Drezner, T.D., *The keystone saguaro (Carnegiea gigantea, Cactaceae): a review of its ecology, associations, reproduction, limits, and demographics.* Plant Ecology, 2014. **215**(6): p. 581-595.

2. Milbrath, G.M., M.R. Nelson, and R.E. Wheeler, *Distribution and Electron-Microscopy of Viruses of Cacti in Southern Arizona.* Phytopathology, 1973. **63**(9): p. 1133-1138.
3. Milbrath, G.M. and M.R. Nelson, *Isolation and Characterization of a Virus from Saguaro Cactus.* Phytopathology, 1972. **62**(7): p. 739-742.
4. Murphy, F.A., *Virus Taxonomy.Sixth Report of the International Committee on Taxonomy of Viruses.* Arch.Virol.(Suppl.), 1995. **10**: p. 412-415.
5. Weng, Z. and Z. Xiong, *Genome organization and gene expression of saguaro cactus carmovirus.* J Gen Virol, 1997. **78 (Pt 3)**(3): p. 525-34.
6. Blockley, A. and R. Mumford, *Identification and isolation of Impatiens necrotic spot virus from prickly pear cactus (Opuntia microdasys) NEW DISEASE REPORT.* Plant Pathology, 2001. **50**(6): p. 805-805.
7. Barrera, B.A., et al., *Association of a Potexvirus as a Causal Agent of Chlorotic Spots on Opuntia ficus-indica.* Revista Mexicana de Fitopatología, 2015. **33**(1).
8. Peng, L., et al., *Molecular characterization and detection of two carlaviruses infecting cactus.* Archives of virology, 2019. **164**(7): p. 1873-1876.
9. Park, C.H., E.G. Song, and K.H. Ryu, *Detection of Co-Infection of Notocactus leninghausii f. cristatus with Six Virus Species in South Korea.* Plant Pathol J, 2018. **34**(1): p. 65-70.
10. Min, B., Y. Song, and K. Ryu, *Complete sequence and genome structure of cactus mild mottle virus.* Archives of virology, 2009. **154**(8): p. 1371-1374.
11. Kim, N.R., et al., *The complete genome sequence of a member of a new species of tobamovirus (rattail cactus necrosis-associated virus) isolated from Aporcactus flagelliformis.* Archives of Virology, 2012. **157**(1): p. 185-187.
12. Duarte, L., et al., *Potexvirus diversity in Cactaceae from Sao Paulo state in Brazil.* Journal of Plant Pathology, 2008: p. 545-551.
13. Roossinck, M.J., D.P. Martin, and P. Roumagnac, *Plant Virus Metagenomics: Advances in Virus Discovery.* Phytopathology, 2015. **105**(6): p. 716-27.
14. Hernandez-Zepeda, C., A. Varsani, and J.K. Brown, *Intergeneric recombination between a new, spinach-infecting curtovirus and a new geminivirus belonging to the genus Becurtovirus: first New World exemplar.* Arch Virol, 2013. **158**(11): p. 2245-54.
15. Jones, R.A. and B.A. Coutts, *Spread of introduced viruses to new plants in natural ecosystems and the threat this poses to plant biodiversity.* Mol Plant Pathol, 2015. **16**(6): p. 541-5.
16. Alexander, H.M., et al., *Plant-virus interactions and the agro-ecological interface.* European Journal of Plant Pathology, 2014. **138**(3): p. 529-547.
17. Bankevich, A., et al., *SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.* J Comput Biol, 2012. **19**(5): p. 455-77.
18. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J., *Basic local alignment search tool* J. Mol. Biol., 1990. **215**: p. 403-410.
19. Muhire, B.M., A. Varsani, and D.P. Martin, *SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation.* Plos One, 2014. **9**(9): p. e108277.
20. Hernandez-Hernandez, T., et al., *Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae).* Am J Bot, 2011. **98**(1): p. 44-61.
21. Varsani, A. and M. Krupovic, *Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family Genomoviridae.* Virus Evol, 2017. **3**(1): p. vew037.
22. Krupovic, M., et al., *Genomoviridae: a new family of widespread single-stranded DNA viruses.* Archives of virology, 2016. **161**(9): p. 2633-2643.
23. Yu, X., et al., *A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus.* Proc Natl Acad Sci U S A, 2010. **107**(18): p. 8387-92.