

# Genome evolution, taxonomy, and transmission of potexviruses in cacti (*Alphaflexiviridae*)

Alexa Tyszk<sup>1,†</sup>, Karolis Ramanauskas<sup>1</sup>, and Boris Igić<sup>1</sup>

<sup>1</sup>Department of Biological Sciences, University of Illinois at Chicago, 840 West Taylor St. MC067, Chicago, IL 60607, United States of America

<sup>†</sup>Author for correspondence.

## ABSTRACT

*Potexvirus* species members are positive-sense single-stranded RNA viruses known to infect many flowering plants, including cacti (Cactaceae). The current viral taxonomic naming schemes in this group often employ informal or outdated host plant names (synonyms), which complicate systematic study. One such group, often named with a suffix "Virus X," presents a further complication—nearly all of its published sequences are from infections of cultivated plants, in which infections may dramatically affect yield. Because their host-specificity is broad, the source of infections, the natural distribution of this group, and the significance of infections in wild species of cacti all remain unclear. The lack of clarity is partly related to low sampling across the Potexviruses that infect cacti. And yet, the availability of sampled plant transcriptomes, all of which are practically metatranscriptomes, has recently exploded, along with the decreasing expense and difficulty of conducting RNAseq experiments. Here, we harness these new tools and perform phylogenetic analyses aimed at clarifying taxonomic diversity, quantifying patterns of tissue expression, diversity, and examining selective pressures across viral genomes. The results suggest a novel mode of transmission by sex (pollination) for this viral group, based on significant expression in pollen. We examine and discuss the implications of our key results for the taxonomy of *Potexviruses* that infect Cactaceae, noting their vastly understudied ecological significance.

## INTRODUCTION

Molisch's (1885) discovery of "protein bodies" on several species of cacti was one of the first documented descriptions of plant viruses. For over a century, subsequent studies were limited to observational data (), often augmented with clever experimental approaches (). A transformative advancement in virology—and all of biology—has been the advent of massively parallel DNA and RNA sequencing. It is readily accessible, inexpensive, with a massively higher throughput, and able to rapidly (albeit crudely) identify organisms from seemingly any sampled surface of the Earth. One common thread is that virtually every macro-organism genome study uncovers a micro-organismal metagenome, composed of both targeted host sequences and those from myriad co-existing organisms. Metagenomic samples have yielded an enormous number of genomes and have vastly expanded the tree of life (Schulz et al., 2017; Hug et al., 2016; Gregory et al., 2019; Lefeuvre et al., 2019; Shi et al., 2016).

The unprecedented amounts of data resulting from metagenomic studies have caused significant policy changes and revisions by the International Committee on Taxonomy of Viruses (ICTV) policy (International Committee on Taxonomy of Viruses Executive Committee, 2020; Simmonds et al., 2017), but nearly all viruses remain named by their original description of host, location, and/or symptoms. The historic naming conventions are ill-suited for host plants whose own taxonomic placement is uncertain, which has been particularly true for rapidly diversified groups such as Cactaceae (cacti).

Molisch's "protein bodies" are now widely understood to be comprised of plant-infecting potexviruses (*Tymovirales*, family *Alphaflexiviridae*). Their positive-sense, single-stranded RNA genomes consist of roughly 6,600 bp of positive-sense single-stranded RNA. Generally presenting as elongated, rod-shaped filamentous viruses, they express five primary open reading frames (ORFs): Replicase (Rep), Triple gene block (TGB), Coat protein (CP), coded in the 5' direction as well as two smaller overlapping ORFs coded

46 in the 3' direction: ORF6 and ORF7 (Evallo et al., 2021; Liou et al., 2004; Martelli et al., 2007). Members  
47 of this group produce variably symptomatic infections in cacti, and many infected plants show no external  
48 signs of viral infection (Bos, 1977; Liou et al., 2004). Neither the significance of their infections in nature,  
49 nor relative modes of transmission are clear. Reports of symptomatic plants range from 0%-5.5% in wild  
50 species in the southwestern United States (Attathom et al., 1978) to 44% in agricultural fields on Hainan  
51 Island, China (Peng et al., 2016). The most commonly recognized symptoms of the disease are mosaic,  
52 mottling, stunted growth, and distortion (Attathom et al., 1978; Maliarenko and Mudrak, 2013; Peng  
53 et al., 2016). Infection through grafting and mechanical contact, particularly following stem injury and  
54 human-mediated or hemipteran insect-mediated sap inoculation, is well-documented (Liou et al., 2004;  
55 Maliarenko and Mudrak, 2013; Park et al., 2018). Grafting is a primary means of propagation among crop  
56 cacti (Park et al., 2018), and *Selenicereus* is a commonly chosen graft stock. However, there are reports of  
57 other members within the family *Alphaflexiviridae* transmitting via insect and seed vectors (Martelli et al.,  
58 2007), and pre-DNA studies tentatively suggest that in the wild, pollen may transmit CVX (Attathom et al.,  
59 1978).

60 *Schlumbergera truncata* (Haworth) Moran has undergone a number of name changes, including  
61 *Epiphyllum truncatum* Haworth in 1819, *Cactus truncatus* (Haworth) Link in 1822, and *Zygocactus*  
62 *truncatus* (Haworth) K. Schumann in 1890, dramatically confusing subsequent viral taxonomy. This  
63 naming scheme may cause confusion as it results in many distinct viruses having the same name. Currently  
64 accepted names in the *Potexvirus* group thus include *Cactus Virus X* (CVX), *Zygocactus Virus X* (ZyVX),  
65 and *Schlumbergera Virus X* (SchVX), each of which was likely characterized on the same host genus  
66 (and possibly species). The Baltimore classification system standardizes viral classification by intrinsic  
67 morphological characteristics of a virus' replication machinery. It has been integrated into the ICTV  
68 guidelines to better reflect viral evolutionary relationships (International Committee on Taxonomy of  
69 Viruses Executive Committee, 2020). The term "plant virus" in itself is problematic since there is strong  
70 evidence to suggest that viruses frequently spill over from fungal or invertebrate hosts (Lefeuve et al.,  
71 2019). Additionally, many plant viruses that infect agriculturally important species are named using the  
72 common name of a plant, which carries its own problems, for example: *Pitaya Virus X* is named for the  
73 common name "Pitaya" which can refer to as many as thirty-one species within the genus *Selenicereus*  
74 (Korotkova et al., 2017; Guerrero et al., 2019; Le Bellec and Vaillant, 2011). The matter is further  
75 complicated by basic viral ecology, because one virus may infect many hosts, and one host may be  
76 co-infected by many viruses. A single-stranded RNA virus has a faster rate of evolution than a host plant.  
77 It engages in a different mode of reproduction, making a direct assignment of viruses and their hosts  
78 difficult (citation needed). There is no guarantee that viral evolution and speciation follow linearly behind  
79 plant evolution and speciation—especially due to viral host-switching. These problems persist throughout  
80 the genus *Potexvirus* and are especially prominent in cactus-infecting *Potexvirus* species. We suggest a  
81 phylogeny-based approach to remedy some prominent taxonomic issues within this specific clade that  
82 cause naming inconsistencies.

83 Knowledge about cactus-infecting *Potexviruses* contributes to a growing yet biased study of plant  
84 viruses. Human-assisted dispersal, grafting, and cultivation obscures the evolutionary history of these  
85 viruses, which parallels the disproportionate sampling representation of plants raised in greenhouses or  
86 for agricultural production. However, *Cactus Virus X* and associated viruses seem restricted to cactaceous  
87 hosts for unknown reasons—every sample of CVX or CVX-related viruses has come from cacti. The few  
88 studies that have investigated wild *Potexviruses* of cacti predate DNA methods and have yet to identify the  
89 origin. Recent sequencing efforts have revealed multiple inconsistent virus-host pairs on cacti. Although  
90 many metagenomic studies capture environmental, genetic information that allows for virus identification,  
91 tissue type may bias expression rates of viruses (Lacroix et al., 2016). The pursuit of wild cactus-infecting  
92 *Potexviruses* expands our evolutionary knowledge of viral evolution, host selection, and transmission  
93 mechanics. The relationships of the virus can be investigated with a thorough phylogenetic approach,  
94 using available virus samples. In this study we present the largest to date phylogeny of cactus-infecting  
95 *Potexviruses*. We attempt to use this expanded phylogeny to answer relevant questions about *Potexvirus*  
96 evolutionary relationships and revisit the utility of decades-old taxonomy in current virus research.

## MATERIALS AND METHODS

### Host Material Collection Species and Tissues

We recovered viral sequences on *Schlumbergera truncata* (Haworth) Moran, commonly known as ‘crab cactus’ or ‘false Christmas cactus,’ one of the most widely cultivated species of cacti. Our samples were sourced from a haphazardly collected personal plant collection (B.I.), purchased or found abandoned around the city of Chicago, IL (USA). Although there are dozens of named varieties of this species, nearly all commercially grown plants are of uncertain provenance. They almost certainly trace to a handful of plants collected in their native Atlantic forests of Brazil and brought to England in the early 1800s (Boyle, 2003). Plants are easily grown from cuttings and the species has been extensively hybridized across Western Europe and exported across the world, prized for their showy winter (short-day) displays. Most of the plants were seemingly asymptomatic at the time of tissue collection, which was aimed at the study of self-incompatibility in angiosperms (Ramanauskas and Igić, 2021), and we fortuitously discovered viral infections in the course of that study.

### RNA Sequencing

Pistils (without ovaries), pollen, leaf, and root tissues were removed and submerged in 1.5 ml of RNeasy<sup>TM</sup> lysis reagent (Qiagen). Samples were held at room temperature for 30-60 minutes and then moved to a -80 °C freezer for storage. Approximately 100 mg of tissue was ground to a fine powder in 1.5 ml tubes submerged in liquid nitrogen. Total RNA was isolated using Total RNA Mini Kit (Plant kit; IBI Scientific, Cat. No. IB47341) following manufacturer’s instructions. We assessed RNA concentration and purity with a NanoDrop<sup>TM</sup> Lite Spectrophotometer (Thermo Scientific). The XX samples used in this study were sequenced as part of a larger sequencing effort which consisted of XXX separate sequencing runs and included additional samples from other plant species. Sequencing libraries were prepared using the KAPA Stranded mRNA-Seq (Roche) These libraries were sequenced on a single lane of Illumina HiSeq 4000 or Illumina NovaSeq 6000 platform (paired-end 150 bp reads) at the Duke University Center for Genomic and Computational Biology. The number of resulting read pairs (for the XX samples presented here) ranged from X,XXX,XXX to X,XXX,XXX with a median of X,XXX,XXX and average of X,XXX,XXX (Table S1).

### RNAseq Assemblies

Raw paired-end Illumina reads were first processed using Rcorrector v1.0.4 (Song and Florea, 2015) to infer and correct sequencing errors. Reads were next trimmed with Trimmomatic v0.39 (Bolger et al., 2014) to remove any read containing bases with Phred scores lower than 20, low quality reads less than 50 bp long, and any adapter or other Illumina-specific sequences that were still present. The remaining reads were filtered with Kraken 2 (Wood et al., 2019) to remove small and large subunit ribosomal RNA (using the SILVA database; Quast et al. 2013) and contaminating reads (minikraken2\_v2 database). We used custom-built databases, derived from RefSeq libraries: UniVec\_Core, viral, mitochondrion, plastid, plasmid, archaea, bacteria, protozoa, human, and fungi to minimize the number of contaminating and non-nuclear reads (Ramanauskas and Igić, 2021). Only paired reads were used for transcriptome assemblies. *Schlumbergera truncata* filtered reads were combined across all samples into a single RNA-seq data set. We conducted a *de novo* transcriptome assembly using Trinity v2.8.5 (Grabherr et al., 2011) to generate a single reference transcriptome assembly for *Schlumbergera truncata*.

### NCBI Data Collection and Compilation

We collected publicly available genomes, complete proteins, gene annotations, and available metadata from Potexviruses (NCBI:txid12176) (NCBI: www.ncbi.nlm.nih.gov/, accession numbers provided in Supplemental Data). The untranslated regions (UTRs) were trimmed from the sequences to provide consistency.

We also searched the NCBI Sequence Read Archive (SRA) database (www.ncbi.nlm.nih.gov/sra) for RNA-sequencing (RNA-seq) data within the flowering plant order Caryophyllales (NCBI:txid3524) that had been sequenced using the Illumina library sequencing platform. For each identified SRA run accession (SRR), viral RNA that matched sample cactus-infecting Potexvirus RNA (accession numbers provided in Supplemental Data) was identified, extracted, and assembled using the kakapo 0.7.3-dev pipeline (http://flightless.one) with Kraken2 viral filters disabled. The SAM files produced through kakapo were

loaded through Geneious 11.1.5 along with the *Schlumbergera* reads. These sequences were annotated using the Geneious 11.1.5 "Find ORFs" function.

The complete dataset comprises 37 existing Potexvirus genomes and proteins, four new viral sequences located within original *Schlumbergera truncata* RNA-seq data, and 52 viral sequences found within NCBI Caryophyllales RNA-seq data.

### Sequence Alignment and Phylogenetic Analyses

Sequence alignments were performed through MAFFT v7.429 ((Kato, 2002)) using the full dataset. The aligned sequences were divided by ORF using the annotations to produce five partial sequence alignments corresponding to each ORF to accompany the full-sequence alignment. The individual proteins were exported to FASTA files, then gaps at the start of the sequence and stop codons were removed manually. Phylogenetic relationships and bootstrap values were calculated using IQtree v1.6.12 ((Nguyen et al., 2015)), ModelFinder ((Kalyanamoorthy et al., 2017)), and UFBoot ((Hoang et al., 2018)) for both the individual gene/protein alignments and the full sequence alignment. Trees were visualized in R version 4.0.3 using ggtree v2.4.2 ((R Core Team, 2020; Yu et al., 2017)). Host information was obtained through reported metadata and mapped onto the phylogeny. Species groupings were determined using the existing species boundaries compared to the phylogenetic branch lengths within the Potexvirus genus. These groupings were generally consistent with most recent branch lengths over 0.1 subs/site and this value was therefore used as a cutoff. Pairwise distance analysis was conducted on the sequence alignments in R using the ape v5.5 dist.dna() function with a raw model. For each defined clade, nonzero pairwise distances between each possible combination of tips was averaged. Expanded phylogenetic trees and individual gene/protein trees are available in the Supplementary Data.

## RESULTS AND DISCUSSION

### Characterization

The collection of *Schlumbergera* samples and thorough investigation of previously published data on Cactaceae resulted in XX new virus lineages. (Figure 1) The genome sizes of X,XXX - X,XXX bp were consistent with published genomic *Potexvirus* data, ranging from X.Xk - X.Xk bp. The XXX sample reads recovered XX percent of the CVX genome for the newly discovered viruses with *Selenicereus* hosts. All of the publicly available new viral lineages were found on *Selenicereus* hosts from SRA XXXXX. We annotated the open reading frames of the viruses to recover all seven Potexvirus proteins.

The *Schlumbergera*-infecting viruses were found in high amounts on pollen and style tissue. The viral loads of each *Selenicereus* sample that was found to have viruses ranged from XX-XX percent of all reads. This was a relatively high recovery rate and in some pollen tissues there existed more viral reads (XX,XXX reads) than *Schlumbergera* reads (XX,XXX reads).

### Distribution of Genetic Distances

The highly similar and well-clustered newly discovered viruses displayed low diversity within the clusters. Therefore, these additions to the *Potexvirus* family tree do not drastically alter the tree structure. Since each sampled cactus in SRA XXXXX was located close to other sampled cacti, this low diversity potentially represents the first example of background mutation among viruses incurred due to host infection. The average nonzero pairwise distance between the included subset of related *Potexviruses* was 0.256 (maximum = 0.492). When the outgroup (including *Plantago asiatica* MV, *Alternanthera* MV, *Papaya* MV, etc.) was excluded from pairwise analysis, the average nonzero pairwise distance value was 0.177 (maximum = 0.326). When these cactus-infecting *Potexviruses* were subdivided into six groups of relatively recent diversification, the average nonzero pairwise distance for full-genome sequences among groups was always above 0.015. The newly discovered *Schlumbergera*-infecting viruses displayed XX percent similarity, and the *Selenicereus*-infecting viruses from existing cactus samples showed XX percent similarity. For the genes RNA-dependent RNA polymerase (RdRp) and Coat protein (CP), which the ICTV recommends be analyzed for species delimitation, the average nonzero pairwise distance was always above 0.02 (Figure 2). This correlates to roughly greater than 97.5 and 98 nucleotide identity.

The ICTV guidelines for *Potexviruses* indicate that less than 72 percent nucleotide sequence identity (or 80 percent amino acid identity) between the CP or Rep genes demarcates separate viral species. Because we compare closely related *Potexviruses*, it might be expected that members of the same putative species would have higher than 72 percent nucleotide identity and members of different putative species would have

200 lower than 72 percent sequence identity. However, the low pairwise distances between *Potexviruses* cause  
201 very few cactus-infecting *Potexviruses* to be demarcated as separate species, even when only considering  
202 previously described species compared to each other. Examples here.

### 203 **Phylogenetic Relationships**

204 A well-supported phylogenetic tree was recovered including closely related non-cactus-infecting *Po-*  
205 *texviruses*. The phylogenetic tree (Figure 1) places the new viral sequences from *Schlumbergera* and  
206 *Selenicereus* near existing viral species within *Potexvirus* (Figure 1). Phylogenetic analysis recovered  
207 defined monophyletic groups corresponding to five or six major groups of cactus-infecting *Potexviruses*,  
208 with *Cactus Virus X* displaying two branching subgroups. *Opuntia Virus X* was the sister clade to the rest  
209 of the cactus-infecting *Potexviruses*. The *S. truncata* samples located within the *Cactus Virus X* clade  
210 appear to represent the first known discovery of *Cactus Virus X* on *Schlumbergera*.

211 Reported host genera are presented for each viral sample. The reported taxonomy of each existing  
212 sample aligns well with the tree structure, but this monophyly is not recovered for hosts. Putative viral  
213 species appear to infect as few as one genus, in the case of *Opuntia Virus X*, or as many as three in the  
214 case of *Schlumbergera Virus X*. Further, the three genera are evolutionarily distinct and polyphyletic in  
215 phylogenetic analyses of *Cactaceae*. This relationship raises questions about a *Potexvirus*'s ability to  
216 switch hosts. All three genera that are reported hosts of *Schlumbergera Virus X* are ornamental crops:  
217 *Selenicereus*, *Schlumbergera*, and *Opuntia*. Extended greenhouse contact between the three genera may  
218 have resulted in viral spillover of *Schlumbergera Virus X*. However, it is also important to note that the  
219 present viral taxonomy is not infallible. The putative species known as *Schlumbergera Virus X* may have  
220 been initially - incidentally - found on a spillover host rather than a member of an actual viral circulating  
221 population. Unfortunately there is no way to know the "original host" of any virus that possesses the  
222 mechanisms to infect multiple cacti genera. Further testing of known host genera is necessary, and  
223 metagenomic sequencing of closely related genera to monitor potential opportunistic spillover.

224 Taxonomic descriptions must also be analyzed for accuracy and reflection of actual viral activity. A  
225 virus named for its first known host may not represent the evolutionary history of the virus. Although  
226 phylogenetic analysis is likely to produce monophyletic clades of viruses that have each been named in  
227 succession for the first known member, this produces inconsistent and confusing names. *Zygocactus Virus*  
228 *X* is a clear example: Although the name *Zygocactus* is outdated in reference to the host genera, now  
229 classified as *Schlumbergera*, the viral name remains. This is exacerbated by the presence of a separate viral  
230 species, *Schlumbergera Virus X*. The ICTV strongly opposes unnecessary name changes, but it is unclear  
231 what should be done with outdated and confusing names that describe an extant clade of viruses. Mixed  
232 infections and the lack of a one-to-one correlation between virus and host complicate naming endeavors  
233 further, and more sampling will uncover novel hosts and mixed infections.

### 234 **Recombination and selection analysis**

235 Recombination events, how they were detected. Selection analysis goes here.

### 236 **Grouping**

237 The new samples fall into already existing species groups. Pairwise distances between species were  
238 calculated for six groups, with a value of 0 indicating identical sequences and a value of 1 indicating  
239 entirely dissimilar sequences.

### 240 **Concluding Remarks**

241 X new viruses were included as part of this study, which represents a multifaceted approach to viral  
242 discovery using metagenomic techniques for already available public data as well as newly collected  
243 data. Mixed infections, co-infection dynamics. Viral abundance, viral loads, tissue type, and diversity.  
244 Transmission: Sexual Taxonomic recommendations - the current species names should not be used in the  
245 future.

### 246 **ACKNOWLEDGMENTS**

247 Acknowledgments text.

## REFERENCES

- Attathom, S., Weathers, L. G., and Gumpf, D. J. (1978). Occurrence and distribution of a virus induced disease of barrel cactus in California. *Plant disease reporter*.
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120.
- Bos, L. (1977). *SYMPTOMS OF VIRUS DISEASES IN PLANTS*. Research Institute for Plant Protection.
- Boyle, T. H. (2003). Identification of self-incompatibility groups in *Hatiora* and *Schlumbergera* (Cactaceae). *Sexual Plant Reproduction*, 16(3):151–155.
- Evallo, E., Taguiam, J. D., and Balendres, M. A. (2021). A brief review of plant diseases caused by Cactus virus X. *Crop Protection*, 143:105566.
- 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., Palma, F. d., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7):644–652.
- Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., Ardyna, M., Arkhipova, K., Carmichael, M., Cruaud, C., Dimier, C., Domínguez-Huerta, G., Ferland, J., Kandels, S., Liu, Y., Marec, C., Pesant, S., Picheral, M., Pisarev, S., Poulain, J., Tremblay, J.-É., Vik, D., Acinas, S. G., Babin, M., Bork, P., Boss, E., Bowler, C., Cochrane, G., de Vargas, C., Follows, M., Gorsky, G., Grimsley, N., Guidi, L., Hingamp, P., Iudicone, D., Jaillon, O., Kandels-Lewis, S., Karp-Boss, L., Karsenti, E., Not, F., Ogata, H., Pesant, S., Poulton, N., Raes, J., Sardet, C., Speich, S., Stemmann, L., Sullivan, M. B., Sunagawa, S., Wincker, P., Babin, M., Bowler, C., Culley, A. I., de Vargas, C., Dutilh, B. E., Iudicone, D., Karp-Boss, L., Roux, S., Sunagawa, S., Wincker, P., and Sullivan, M. B. (2019). Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell*, 177(5):1109–1123.e14.
- Guerrero, P. C., Majure, L. C., Cornejo-Romero, A., and Hernández-Hernández, T. (2019). Phylogenetic Relationships and Evolutionary Trends in the Cactus Family. *Journal of Heredity*, 110(1):4–21.
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, 35(2):518–522.
- Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., HERNSDORF, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D. A., Finstad, K. M., Amundson, R., Thomas, B. C., and Banfield, J. F. (2016). A new view of the tree of life. *Nature Microbiology*, 1(5):16048.
- International Committee on Taxonomy of Viruses Executive Committee (2020). The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nature Microbiology*, 5(5):668.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6):587–589.
- Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14):3059–3066.
- Korotkova, N., Borsch, T., and Arias, S. (2017). A phylogenetic framework for the Hylocereae (Cactaceae) and implications for the circumscription of the genera. *Phytotaxa*, 327(1):1.
- 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. *Applied and environmental microbiology*, 82(6):1966–1975.
- Le Bellec, F. and Vaillant, F. (2011). Pitahaya (pitaya) ( *Hylocereus* spp.). In *Postharvest Biology and Technology of Tropical and Subtropical Fruits*, pages 247–273e. Elsevier.
- Lefeuvre, P., Martin, D. P., Elena, S. F., Shepherd, D. N., Roumagnac, P., and Varsani, A. (2019). Evolution and ecology of plant viruses. *Nature Reviews Microbiology*, 17(10):632–644.
- Liou, M. R., Chen, Y. R., and Liou, R. F. (2004). Complete nucleotide sequence and genome organization of a Cactus virus X strain from *Hylocereus undatus* (Cactaceae). *Archives of Virology*, 149(5):1037–1043.
- Maliarenko, V. M. and Mudrak, T. P. (2013). Cactus viruses in fasciated plants. *Biologija*, 59(2).
- Martelli, G. P., Adams, M. J., Kreuze, J. F., and Dolja, V. V. (2007). Family *Flexiviridae* : A Case Study in Virion and Genome Plasticity. *Annual Review of Phytopathology*, 45(1):73–100.
- Molisch, H. (1885). Über merkwürdige geformte Proteinkörper in den Zweigen von *Epiphyllum*. *Berichte der Deutschen botanischen Gesellschaft in Berlin*, 3:195–202.
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A Fast and

303 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology*  
304 *and Evolution*, 32(1):268–274.

305 Park, C. H., Song, E. G., and Ryu, K. H. (2018). Detection of Co-Infection of *Notocactus leninghausii* f.  
306 *cristatus* with Six Virus Species in South Korea. *The Plant Pathology Journal*, 34(1):65–70.

307 Peng, C., Yu, N. T., Luo, Z. W., Fan, H. Y., He, F., Li, X. H., Zhang, Z. L., and Liu, Z. X. (2016). Molecular  
308 Identification of Cactus virus X Infecting *Hylocereus polyrhizus* (Cactaceae) in Hainan Island, China.  
309 *Plant Disease*, 100(9):1956.

310 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F. O.  
311 (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based  
312 tools. *Nucleic Acids Research*, 41(D1):D590–D596.

313 R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for  
314 Statistical Computing, Vienna, Austria.

315 Ramanauskas, K. and Igić, B. (2021). Rnase-based self-incompatibility in cacti. *New Phytologist*,  
316 231(5):2039–2049.

317 Schulz, F., Elie-Fadrosh, E. A., Bowers, R. M., Jarett, J., Nielsen, T., Ivanova, N. N., Kyrpides, N. C., and  
318 Woyke, T. (2017). Towards a balanced view of the bacterial tree of life. *Microbiome*, 5(1):140.

319 Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., Qin, X.-C., Li, J., Cao, J.-P., Eden, J.-S.,  
320 Buchmann, J., Wang, W., Xu, J., Holmes, E. C., and Zhang, Y.-Z. (2016). Redefining the invertebrate  
321 RNA virosphere. *Nature*, 540(7634):539–543.

322 Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Brister, J. R., Carstens, E. B., Davison, A. J.,  
323 Delwart, E., Gorbalenya, A. E., Harrach, B., et al. (2017). Virus taxonomy in the age of metagenomics.  
324 *Nature Reviews Microbiology*, 15(3):161–168.

325 Song, L. and Florea, L. (2015). Rcorrector: efficient and accurate error correction for Illumina RNA-seq  
326 reads. *GigaScience*, 4(1):48.

327 Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome*  
328 *Biology*, 20(1):257.

329 Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T. (2017). ggtree: an R package for visualization and  
330 annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology*  
331 *and Evolution*, 8(1):28–36.

332 **FIGURES**

333 Figure inventory list. 1. The updated cactus-infecting *potexvirus* phylogeny reflects distinct viral groupings  
334 which opportunistically infect multiple host plants. Asterisks (\*) represent samples recovered by this study.  
335 2. Genome organization of viruses in phylogeny, ORFs color-coded and arrows indicating which direction  
336 each gene is read. 3. Selection graph as displayed in B's r code 4. Table of pairwise distances in nucleotides  
337 5. Graph of pairwise distances 6. Proposed update to potexvirus taxonomy.  
338 Supplemental figures: 1. full gene phylogenies for each gene in genome 2. Tissue expression rates?