Journal of Visualized Experiments

Combining analysis of DNA in a crude virion extraction with analysis of RNA from infected leaves to discover new virus genomes.

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE57855R2
Full Title:	Combining analysis of DNA in a crude virion extraction with analysis of RNA from infected leaves to discover new virus genomes.
Keywords:	virus purification; plant virus; Next generation sequencing; badnavirus
Corresponding Author:	Jeanmarie Verchot, PhD Texas A&M AgriLife Dallas, TX UNITED STATES
Corresponding Author's Institution:	Texas A&M AgriLife
Corresponding Author E-Mail:	jm.verchot@ag.tamu.edu
First Author:	Jeanmarie Verchot, PhD
Other Authors:	Aastha Thapa
	Dulanjani Wijayasekara
	Peter Hoyt
Author Comments:	Dear Editorial Team This manuscript is a revised submission as an invited manuscript by Jaydev Upponi, the science editor for Immunology and Infection. We are honored to have our methods featured in this journal and we hope that this method will be valuable to present to the broader scientific community as a simple and optimal method for virus discovery that makes sense. We would be honored for you to review this and provide comments how me may improve the manuscript Sincerely Jeanmarie Verchot
Additional Information:	
Question	**Response**
If this article needs to be "in-press" by a certain date, please indicate the date below and explain in your cover letter.	

March 3, 2018

Dear Editor

I am submitting a revised manuscript entitled "**Combining analysis of DNA in a crude virion extraction with analysis of RNA from infected leaves to discover new virus genomes**". **The original title was: Combining standard virus particle extraction method with next generation sequencing to discover new virus genomes".** We are excited for the opportunity to have this work featured in JoVE where we can present a technology for virus discovery. We present a pipeline where we prepare a crude preparation of virions, isolate and sequence DNAs in that preparation. Then purify RNA and following ribosome depletion carry out RNA-seq. By comparing the 2 datasets we were able to discover and confirm the existence of two virus genomes. We are excited to share this method and hope other researchers find this approach useful.

The reviewers comments were extensive and we have taken the time to diligently address every comment. I decided the best approach is to annotate the responses to the reviewers line-by-line and attach the document and responses. The manuscript has all edit tracking intact so you can follow the changes

Thank you for inviting this manuscript and we hope the changes are acceptable.

Sincerely

Jeanmarie Verchot

TITLE:

Combining Analysis of DNA in a Crude Virion Extraction with the Analysis of RNA from Infected Leaves to Discover New Virus Genomes

AUTHORS & AFFILIATIONS:

Jeanmarie Verchot[1], Aastha Thapa[2], Dulanjani Wijayasekara[3], and Peter R. Hoyt[4]

[1]Texas A&M Agrilife Center at Dallas, 17360 Coit Rd, Dallas TX 75252
[2]Noble Research Center, Oklahoma State University, Stillwater, Oklahoma, 74075
[3]Department of Biology, Oliphant Hall 304, College of Engineering and Natural Sciences, The University of Tulsa, 800 S. Tucker Dr., Tulsa, OK 74104
[4]Bioinformatics and Genomics Core Facility, Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, OK 74078

EMAIL ADDRESSES OF CO_AUTHORS:

Jeanmarie Verchot (jm.verchot@ag.tamu.edu)
Aastha Thapa (Aastha@okstate.edu)
Dulanjani Wijayasekara (dsw1451@utulsa.edu)
Peter Hoyt (peter.r.hoyt@okstate.edu)

CORRESPONDING AUTHOR:

Jeanmarie Verchot (jm.verchot@ag.tamu.edu)
Phone number: 9729529224

KEYWORDS:

Virus purification, plant virus, next-generation sequencing, badnavirus, virus metagenomics, virus

SHORT ABSTRACT:

Here we present a new approach to identify plant viruses with double-strand DNA genomes. We use standard methods to extract DNA and RNA from infected leaves and carry out next-generation sequencing. Bioinformatic tools assemble sequences into contigs, identify contigs representing virus genomes and assign genomes to taxonomic groups.

LONG ABSTRACT:

This metagenome approach is used to identify plant viruses with circular DNA genomes and their transcripts. Often plant DNA viruses that occur in low titers in their host or cannot be mechanically inoculated to another host are difficult to propagate to achieve a greater titer of infectious material. Infected leaves are ground in a mild buffer with optimal pH and ionic composition recommended for purifying most bacilliform Para retroviruses. Urea is used to break up inclusion bodies that trap virions and to dissolve cellular components. Differential centrifugation provides further separation of virions from plant contaminants. Then proteinase K treatment removes the capsids. Then the viral DNA is concentrated and used for next-generation sequencing (NGS). The NGS data are used to assemble contigs which are submitted

to NCBI-BLASTn to identify a subset of virus sequences in the generated dataset. In a parallel pipeline, RNA is isolated from infected leaves using a standard column-based RNA extraction method. Then ribosome depletion is carried out to enrich for a subset of mRNA and virus transcripts. Assembled sequences derived from RNA sequencing (RNA-seq) were submitted to NCBI-BLASTn to identify a subset of virus sequences in this dataset. In our study, we identified two related full-length badnavirus genomes in the two datasets. This method is preferred to another common approach which extracts the aggregate population of small RNA sequences to reconstitute plant virus genomic sequences. This latter metagenomic pipeline recovers virus related sequences that are retro-transcribing elements inserted into the plant genome. This is coupled to biochemical or molecular assays to further discern the actively infectious agents. The approach documented in this study, recovers sequences representative of replicating viruses that likely indicate active virus infection.

INTRODUCTION:

Emerging plant diseases drive researchers to develop new tools to identify the correct causal agent(s). Initial reports of new or recurring virus diseases are based on commonly occurring symptoms such as mosaic and malformations of the leaf, vein clearing, dwarfism, wilting, lesions, necrosis, or other symptoms. The standard for reporting a new virus as the causal agent for a disease is to separate it from other contaminating pathogens, propagate it in suitable host, and reproduce the disease by inoculating into healthy plants of the original host species. The limitation in this approach is that many genera of plant viruses depend upon an insect or other vectors for transmission to a suitable host or back to the original host species. In this case, the search for the appropriate vector can be prolonged, there may be difficulties to establish laboratory colonies of the vector, and further efforts are necessary to devise a protocol for experimental transmission. If the conditions for successful laboratory transmission studies cannot be achieved, then the work falls short of the standard for reporting a new virus disease. For viruses that occur in their natural hosts at very low titers, researchers must identify alternative hosts for propagation to maintain sufficient infectious stocks for carrying out research. For virus species that infect only a few plants this can also be an obstacle for growing stock cultures[1].

In recent years, scientists are more often employing high-throughput NGS and metagenomic approaches to uncover virus sequences that are present in the environment, which may exist unrelated to a known disease, but can be assigned to taxonomic species and genera[2-4]. Such approaches to the discovery and categorization of genetic materials in a distinct environment provide a way to describe virus diversity in nature or their presence in a certain ecosystem but does not necessarily confirm to a framework for defining causal agents for an apparent disease.

The *Badnavirus* genus belongs to the family *Caulimoviridae* of pararetroviruses. These viruses are bacilliform in shape with circular double strand DNA genomes of approximately 7 to 9 kb. All pararetroviruses replicate through an RNA intermediate. Pararetroviruses exist as episomes and replicate independent of the plant chromosomal DNA[5,6]. Field studies of virus populations indicate that these virus populations are genetically complex. In addition, information obtained across a range of plant genomes by high throughput sequencing have uncovered numerous

examples of badnavirus genome fragments inserted by illegitimate integration events into plant genomes. These endogenous badnavirus sequences are not necessarily associated with infection[7-11]. Subsequently, the use of NGS to identify new badnaviruses as the causal agent of disease is complicated by the subpopulation diversity of episomal genomes as well as the occurrence of endogenous sequences[12,13].

While there is not one optimal pipeline for the discovery of novel pararetrovirus genomes, there are two common approaches to identify these viruses as causal agents for disease. One method is to enrich for small RNA sequences from infected leaves and then assemble these sequences to reconstitute the virus genome(s)[14-17]. Another approach is the rolling circle amplification (RCA) to amplify circular DNA virus genomes[18]. The success of RCA depends upon the age of the leaf and the virus titer in the selected tissue. The RCA products are subjected to restriction digestion and cloned into plasmids for direct sequencing[19-21].

Canna yellow mottle virus (CaYMV) is a badnavirus and is described as the etiological cause of yellow mottle disease in canna, although only a 565 bp fragment of the genome has been previously isolated from infected cannas[22]. A contemporary study identified CaYMV in *Alpinia purpurata* (flowering ginger; CaYMV-Ap)[23]. The goal of this study was to recover complete badnavirus genome sequences from infected canna lilies. We describe a protocol for purifying virus from plant contaminants, and then isolating viral DNA from this preparation, and prepare a DNA library for use in NGS. This approach eliminates the need for intermediate molecular amplification steps. We also isolate mRNA from infected plants for RNA-seq. NGS, which includes RNA-seq was carried out using each nucleic acid preparation. Assembled contigs were found to relate to the *Badnavirus* taxon in both datasets using the National Center for Biotechnology and Information (NCBI) basic local alignment search tool for nucleic acids (BLASTn). We identified the genomes of two badnavirus species[24].

PROTOCOL:

1. General Virus Purification by Differential Centrifugation Using Standard Method by Covey *et al.*[25]

1.1) First, cut 80–100 g of leaves from diseased plants and grind in a waring blender at 4 °C using 200 mL grinding buffer (0.5 M NaH_2PO_4, 0.5 M Na_2HPO_4 (pH 7.2). and 0.5% (w/v) Na_2SO_3). Wear a laboratory coat and gloves for all steps of this procedure.

1.2) Then, transfer the homogenate (300 mL) to a 1.0 L beaker. Add 18 g of urea and 25 mL of 10% nonionic detergent (*t*-Oct-C_6H_4-$(OCH_2CH_2)_9OH$) to the homogenate inside a chemical hood.

Note: For this step it is best to wear safety goggles and a simple breathing mask for personal protection.

1.3) Stir with a magnetic stirrer briefly in the hood and cover the beaker with the foil. Then transfer the foil covered beaker to a cold room and stir with a magnetic stirrer overnight at 4 °C.

1.4) Transfer the homogenate to centrifuge rotor bottles (250 mL containers) and centrifuge in a fixed angle rotor at 4,000 x g for 10 min at 4 °C. In a chemical fume hood, recover the supernatant and filter through 4 layers of cheesecloth.

1.5) Divide the homogenate among 38.5 mL polypropylene centrifuge tubes and centrifuge for 2.5 h at 40,000 x g at 4 °C. Typically, check for the presence of a green pellet at the bottom of the tube and a white pellet along the length of the tube. Pour off the supernatant and retain both pellets; place samples on ice.

Note: The green pellet contains chloroplasts, starch, and other organelles.

1.6) Working in a chemical hood, use a rubber policeman to separate the pellets. Resuspend the white pellet in each rotor bottle in 1 mL of ddH_2O over the course of 1–2 h while maintaining the suspensions overnight at 4 °C to allow the materials to fully dissolve into solution. Centrifuge the suspension at 6,000 x g and at 4 °C for 10 min to remove the remaining debris.

1.7) Centrifuge the concentrated suspension at 136,000 x g for 2 h at 4 °C to pellet virions. Resuspend the pellets in 1 mL of buffer (50 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$).

Note: An optional step is to treat virions with DNAse I (10 µg/mL) for 10 min at 37 °C to remove non-encapsidated DNA, *i.e.,* contaminating chloroplast and mitochondrial DNA. Then, inactivate the DNAse I by adding EDTA to 1 mM.

1.8) Disrupt virions with 40 µL of 2 µg/µL proteinase K at 37 °C for 15 min.

1.9) Work inside a chemical hood to recover virion DNA by organic extraction. Wear a face shield, gloves, and a laboratory coat during the extraction for protection against potential acute health effects. Add 1 volume of phenol-chloroform-isoamyl alcohol (49:50:1) to the sample and shake by hand for 20 s. Centrifuge at room temperature for 5 min at 16,000 x g. Remove the upper aqueous phase and transfer to a new tube. Repeat this extraction two or more times. Dispose of the organic phase by placing it in a glass waste bottle for proper institutional chemical disposal[26].

1.10) Concentrate the DNA using ethanol precipitation. Use 0.3 M final concentration of sodium acetate (pH 5.2) and 2.5 volumes of 95% ethanol. Place samples at -20 °C for 30–60 min and centrifuge at 13,000 x g for 10–20 min to pellet the DNA[26].

1.11) Working at a laboratory bench, resuspend the DNA pellet in 1 mL of 0.1 mM TE buffer (pH 8.0). Filer the suspension through a commercial gel filtration column (normally used for polymerase chain reaction (PCR) clean up) to eliminate salts and low molecular weight material that might impede NGS.

1.12) Analyze the samples by 1% agarose gel electrophoresis using ethidium bromide staining to view the quality of the preparations. Assess the quality of DNA using a nanodrop

spectrophotometer.

Note: A ratio of sample absorbance at 260 λ and 280 λ between 1.85 and 2.0 typically indicates that the preparation is "clean" of impurities and is of the desired quality.

1.13) Analyze the quality of DNA (use 5 pg to 10 ng) using a chip based capillary electrophoresis instrument.

Note: Quality output shows clean peaks, representing DNA fragments distributed by size along an X-axis. Peak height indicates abundance of the fragment. Jagged peaks indicate partially degraded fragments or chemical contaminants. Round curves represent a smear of DNA indicating poor quality

2. Library Preparation Using DNA and Emulsion-based Clonal Amplification (emPCR Amplification)

Note: The library is typically prepared by a NGS facility which carries out customer-oriented work.

2.1) Shear a solution of DNA (> 200 ng) using a nebulizer which converts the DNA to fragments. Ligate the commercial adapters according to the manual's instructions[27].

2.2) Carry out emPCR amplification of the DNA sample according to the manufacturer's instructions[28-30]. Repeat the wash step three times, and after each wash, pellet the beads in a minicentrifuge for 10 s. Discard the supernatant after each wash.

Note: The procedure begins with preparation of the capture beads by washing in the commercial wash buffer provided with the kit. emPCR is commonly used for template amplification for NGS.

2.3) Heat denature the DNA or RNA at 95 °C for 2 min and then 4 °C until ready to use. Use 200-million molecules of DNA/RNA to 5 million capture beads in a final volume of 30 µL. Prepare a mock sample alongside the DNA/RNA sample and carry out the following steps with the nucleic acid sample as well as the mock sample.

2.4) Perform emulsification by vortexing the tube of emulsion oil for 10 s at maximum speed, then pour the entire content (4 mL) into a plastic stirring tube that is compatible with a platform homogenizer. Place the stirring tube on the platform to mix the emulsion at 2,000 rpm for 5 min.

2.5) Dispense 100 µL aliquots of emulsion into 8-strip cap tubes or into a 96-well plate. Cap the tubes or seal the plate and carry out emPCR using the manufacturer's recommended program[28].

Note: After the PCR is complete, check the wells to see if the emulsion is intact and then proceed. Discard the entire well if the emulsion is broken.

2.6) Wear a lab coat and work in a chemical hood to collect the amplified DNA beads (ADB). Vacuum aspirate the emulsion from the wells and collect the beads in a 50 mL tube. Rinse the wells twice with 100 µL of isopropanol and aspirate the rinse to the same 50 mL tube.

2.7) Vortex the collected emulsions and resuspend the ADB with isopropanol to a final volume of 35 mL. Pellet the ADB at 930 x g for 5 min. Remove the supernatant and add 10 mL of enhancing buffer. Vortex the ADB and then wash by adding isopropanol to 40 mL final volume Centrifuge and discard the supernatant after each wash and repeat the wash step twice.

2.8) Carry out a final wash using ethanol in place of isopropanol. Add enhancing buffer to 35 mL final volume, vortex, and pellet the beads at 930 x g for 5 min. Remove the supernatant but leave 2 mL of enhancing buffer.

2.9) Transfer the suspension to a microcentrifuge tube and briefly centrifuge to pellet the ADB. After discarding the supernatant, rinse the ADB pellet twice with 1 mL of enhancing buffer. Centrifuge and discard the supernatant after each wash.

2.10) To prepare for the DNA library bead enrichment, add 1 mL of 1 N NaOH to the beads. Vortex the ADB and then incubate for 2 min at room temperature. Centrifuge and discard the supernatant. Repeat this wash step once.

2.11) Add 1 mL of annealing buffer, then vortex the ADB and incubate for 2 min at room temperature. Briefly centrifuge and discard the supernatant. Repeat this step again using 100 µL of annealing buffer.

2.12) To anneal sequencing primers to the DNA, add 15 µL of Seq Primer A and 15 µL of Seq Primer B provided in the kit. Briefly mix by vortexing and place the microcentrifuge tube in a heat block at 65 °C for 5 min. Transfer to ice for 2 min.

2.13) Wash three times with 1.0 mL of annealing buffer. Vortex for 5 s and discard the supernatant each time.

2.14) Before sequencing, measure the number of beads using a commercial bead counter. There should be at least 500,000 enriched beads.

Note: The bead counter is a special device that measures the beads in a supplied microcentrifuge tube.

3. General mRNA Isolation and dsDNA Synthesis Starting with Infected Canna Leaves that Test by RT-PCR for CaYMV Using Reported Diagnostic Primers

3.1) Wear a laboratory coat and latex gloves for personal protection in all subsequent steps. Working at a laboratory bench, collect 12 samples from the leaves and plunge the samples into liquid nitrogen. Use a bead mill for homogenization. Use a commercial kit that provides a

standard column-based method for total plant RNA isolation. Add the guanidine-isothiocyanate lysis buffer provided by the kit to the ground sample and shake for 20 s.

3.2) Add ethanol and mix thoroughly, according to kit instructions. Add each homogenate to a spin column that binds RNA to the membrane. Wash three times and elute the RNA into a recovery tube[24].

3.3) Quantify the RNA using a spectrophotometer to measure the ratio of absorbance at 260 λ and 280 λ. Verify the RNA integrity using 1% agarose gel electrophoresis stained with ethidium bromide.

Note: An absorbance ratio between 1.85 and 2.0 indicates that the preparation is of the desired quality. Treat RNA with DNase I (10 µg/mL) for 10 min at 37 °C. Use a commercial spin column to concentrate RNA in RNase-free water[31]. Pool RNA samples before proceeding.

3.4) Use an rRNA removal kit to remove plant ribosomal RNA. Aliquot magnetic beads to the microcentrifuge tube and wash twice with RNase-free water. Vortex the tube aliquot to resuspend, place the tube on a magnetic stand and wait for liquid to clear. Discard the supernatant and replace with the magnetic bead resuspension solution. Vortex to resuspend and add 1 µL of RNase Inhibitor.

Note: Such kits use oligo-dT bound to magnetic beads which hybridize to mRNA. The method uses standard magnet bead separation technology to recover transcripts[24].

3.5) Combine 500 ng to 1.25 µg of RNA, RNase-free water, and reaction buffers provided by the kit. Place the mixture for 10 min at 50 °C. Remove from heat and add washed magnetic beads in RNAse free water. Vortex briefly and set at room temperature for 5 min.

3.6) Place on a magnetic stand and wait for the liquid to clear. Transfer the supernatant to a fresh microcentrifuge tube. Set on ice.

3.7) Use a solution-based capture method for enrichment of exosomes and 200 ng of RNA to prepare the cDNA library.

Note: The double strand cDNA library is typically prepared by a NGS facility which carries out customer-oriented work.

3.8) Fragment the RNA using a commercial RNA fragmentation solution (0.136 g $ZnCl_2$ and 100 mM Tris-HCl pH 7.0). Add 2 µL of solution to 18 µL of RNA (200 ng total). Spin tubes briefly in the microcentrifuge, place the samples at 70 °C for 30 s, and transfer to ice. Stop the reaction using 2 µL of 0.5 M EDTA pH 8.0 and 28 µL of 10 mM Tris-HCl pH 7.5.

3.9) Bind RNA to magnetic beads by mixing at room temperature for 10 min. Use a magnetic concentrator to collect the beads and discard the supernatant. Wash the beads three times with

200 µL of 70% ethanol. Discard each wash and then air dry the pelleted beads at room temperature for 3 min. Resuspend in 19 µL of 10 mM Tris-HCl pH 7.5.

3.10) Anneal random primers to fragmented RNA by heating to 70 °C for 10 min and then place the tube on ice for 2 min. Prepare the first strand and second strand cDNA using a standard commercial cDNA synthesis kit.

3.11) Purify the double-strand cDNA using a magnetic bead concentrator. Wash with 800 µL of 70% ethanol three times. Discard each wash and air dry the pellets at room temperature for 3 min. Resuspend in 16 µL of 10 mM Tris-HCl pH 7.5. Use the magnetic bead concentrator to separate the beads from the double-stranded cDNA, which is now in solution. Remove the cDNA by pipetting into a new 200 µL PCR tube.

3.12) Carry out fragment end repair using Taq polymerase and a mixture of deoxyribonucleotides provided by a commercial library preparation kit. The commercial kit provides pre-diluted adapters to add to each end of the double-strand cDNA using commercial ligase at 25 °C for 10 min.

4. NGS of DNA Library Prepared from Crude Virus Preparation and dsDNA Library Prepared from mRNA

4.1) Use a standard high throughput pyrosequencing instrument and follow all recommended manufacturers' protocols to generate direct readouts of DNA sequences. Use commercial sequencing reagents, including fluorescently labeled nucleotides.

Note: For details refer manufacturer's instructions provided with the instrument.

4.2) Carry out post-sequencing analysis using genome assembly software which automatically assembles reads to produce the first set of contigs with an average length of < 700 bp. Use the FastQC software on the iPlant/CyVerse website which performs quality control checks on the raw sequence data[32]. Select sequences with Phred scores ≥ 30 to continue to reconstruct longer sequences from smaller sequence reads[24] using mapping and amplicon software.

Note: For details refer manufacturer's instruction.

4.3) Submit these assembled contigs to NCBI-BLASTn analysis using MEGABLAST default module as well as Viridplantae (TaxID: 33090) and Viruses (TaxID: 10239) as the limiting organismal names[33]. Gather the subpopulation of contigs that show high similarity to reported *Badnavirus* genomes into a report.

4.4) Verify that the joined scaffolds that represent one or more candidate full-length virus genomes, correctly produce in-frame sequences that have the same organization as the standard badnavirus genome. To do this, input the candidate full length virus genome into a plasmid drawing software. Then confirm the first 15 nucleotides consists of a $tRNA^{met}$

(TGGTATCAGAGCGAG) which is highly conserved among badnaviruses. Locate the potential polyadenylation signal near the 3' end of the genome. Annotate the complete genome to identify the presence of two small ORFs and one large ORF encoding a polyprotein. Then use the ExPASy portal translate tool to identify the badnavirus ORF1, ORF2, and ORF3 translation products[34].

Note: This scientific software is free and will generate circular DNA, identify all open reading frames, and provides an immediate output to verify that the sequence represents the full length circular DNA genome.

4.5) Use open source multiple sequence comparison tools, MUSCLE and CLUSTALW, to compare the virus genomes obtained from DNA and RNA analyses[35,36].

4.6) Search the NCBI nucleotide database to obtain the full genome sequences of 30 badnavirus species and export them as a document in .fasta format. Upload sequences to a software that conducts evolutionary genetic analysis of sequences along with the virus genome sequences obtained by NGS. Generate multiple sequence alignments and Maximum likelihood trees using MUSCLE[37].

5. Quality Assessment of *De Novo* Sequencing by PCR Amplification of Virus Genomes from Infected Plants

5.1) Input the newly identified full-length badnavirus genome sequences (.fasta format) into the free online Primer3 tool to derive PCR primers[38]. Identify primer sets that will produce staggered products of 1,000–1,500 bp along the entire length of the virus genome(s). Send the sequences to a service facility that will synthesize and deliver PCR primers.

Note: The output identifies acceptable primer pairs with common and acceptable melting temperatures and precise primer locations along the introduced sequences.

5.2) Working at a laboratory bench and wearing a lab coat and gloves, isolate 5 µg of DNA from the virus-infected and healthy control leaves using an automated method that involves standard paramagnetic cellulose particles to isolate DNA from plant material[39] . Freeze leaf material (20–40 mg) in liquid nitrogen in a microcentrifuge tube and grind using a bead mill. Combine the sample with lysis buffer in a microcentrifuge tube and add RNase A to each sample. Vortex the sample for 10–20 s and briefly spin the sample to remove solid particles.

Note: Paramagnetic cellulose particles have high DNA-binding capacity and isolate high yields of pure DNA. The standard commercial silica column methods for DNA isolation do not efficiently extract DNA from a wide variety of plant species. Consequently, dozens of methods exist that are modifications of these procedures to improve the efficiency for individual plant species. The automated paramagnetic cellulose particle method was chosen because it yields more and higher quality DNA from more than 25 herbaceous angiosperm species[40].

5.3) Use commercial reagent cartridges for automated paramagnetic DNA isolation. Add 300

µL of nuclease free water to each commercial reagent cartridge and transfer plant lysate to the same cartridge. Place the cartridge in the cartridge rack, place a plunger in the well closest to the elution tube, and place elution buffer into the elution tube. Load cartridges into the automated nucleic acid isolation machine and run the plant DNA isolation protocol[41,42].

5.4) Carry out PCR to derive a set of overlapping PCR products. Use 5 µM of each forward and reverse primer with 35 cycles of PCR amplification. Use the following cycling conditions: denaturation at 95 °C for 60 s, annealing at 50 °C for 45 s, and extension at 72 °C for 1–2 min with a final extension at 72 °C for 7–10 min. Use a prepackaged gel filtration column to eliminate salts and low molecular weight material as in step 1.2[31].

5.5) Calculate a 3:1 molar ratio of PCR product to vector to determine the amount of PCR product to ligate to 50 ng of linearized pGEM plasmid[43]. Use a control insert DNA to determine whether the ligations work efficiently. Perform the ligation overnight using T4 DNA ligase (3 U/µL) at 4 °C. Then transform commercially prepared JM109 competent *Escherichia coli* cells. Use control 100 pg of uncut plasmid DNA as a positive control for efficient transformation. Plate 100 µL of transformed cells onto LB-agar plates with antibiotic and blue/white selection to recover ligated plasmids[26]. Incubate plates for 16–24 h at 37 °C.

Note: The pGEM vector has a lacZ gene which encodes β-galactosidase. Transformed bacteria grown on a plate containing 100 µg/mL ampicillin, 0.5 mM IPTG, 80 µg/mL 5-bromo-4-chloro-3-indoyl-β-D-galactopyranosidase (X-gal) will turn blue because of β-galactosidase activity. The pGEM plasmid is linearized in a manner that disrupts the *lacZ* gene. Colonies that contain the PCR product inserts disrupt the lacZ gene and do not metabolize X-gal. These colonies are white. Thus colonies with an insert can be differentiated from those without an insert by the color of the colony (white versus blue)[26].

5.6) Isolate DNA from three colonies using a standard column-based plasmid isolation kit[39]. Sequence three plasmids of transformation product. Compare each DNA sequence with the *de novo* assembled virus genomes produced by NGS. Use CLUSTALW to align the sequences and to ensure that they are appropriately ordered.

REPRESENTATIVE RESULTS:

This modified virus purification method provided an enrichment of virus DNAs useful for identifying two virus species by NGS and bioinformatics. After the homogenate was centrifuged at 40,000 x g for 2.5 h, there was a green pellet at the bottom of the tube and a white pellet along the length. The green pellet was resuspended into one microcentrifuge tube and the white pellet was resuspended into two microcentrifuge tubes. PCR was carried out using standard CaYMV PCR diagnostic primers, and products were detected in the solubilized white pellet and not the green pellet (**Figure 1A**). A sample of the crude preparation was examined by transmission electron microscopy and we observed bacilliform particles measuring 124–133 nm in length (**Figure 1B**). This is within the predicted modal length of most badnaviruses. DNA was extracted from the white and green pellets and resuspended separately. In **Figure 1C**, we loaded 5 µL of DNA extracted from the green and white pellet sample (1.6 µg of DNA for the green fraction and

3.1 µg of DNA for the white fraction) to 0.8% agarose gel electrophoresis and analyzed the DNA following ethidium bromide staining. The green fraction contained low molecular weight DNA whereas the white fraction produced two bands of higher molecular weight DNA, as well as the lower molecular weight DNA (**Figure 1C**). The gel presented in **Figure 1C** was run for 40 min at 100 V and the smear in lane 3 suggests that the gel voltage should be lowered to produce clearer bands. These data suggest that the white pellet was enriched for virions. The DNA (0.6 µg/mL) concentration extracted from the white sample was low, but adequate for NGS, which requires a minimum of 10 ng of DNA to proceed. Fragmented DNAs were used to prepare a library for NGS.

In parallel, RNA was extracted from infected canna plants (**Figure 1D**) for high-throughput RNA-seq. A standard workflow was carried out for library preparation, NGS, creating contigs, and identifying viral genome sequences (**Figure 1E**). The output results from using DNA and RNA as starting materials were compared.

We obtained 188,626 raw DNA reads by NGS using DNA isolated from crude virus preparation. Reads were assembled into 13,269 contigs and BLASTn was used to search the NCBI dataset of nucleotide sequences (using Viridplantae TaxID: 33090 and Virus TaxID: 10239 as the limiting organisms) (**Figure 1E**). The NCBI-BLASTn results revealed that 93% of *de novo* assembled contigs were cellular sequences, 22% were unknown, and 0.3% were virus contigs (**Figure 2A**). The majority of contigs categorized as cellular sequences were identified as mitochondrial or chloroplast DNA. Within the dataset of virus contigs, 32% of the virus contigs were related to members of *Caulimoviridae* (that were not Badnavirus sequences) and 58% of these were related to *Badnavirus.* Of the virus contigs, 29% were highly similar ($e < 1 \times 10^{-30}$) to CaYMV isolate V17 ORF3 gene (EF189148.1), *Sugarcane bacilliform virus* isolate Batavia D, complete genome (FJ439817.1), and *Banana streak CA virus* complete genome (KJ013511). Within this population, there were long contigs that resembled two full length genomes.

High-throughput RNA-seq produced 153,488 cleaned individual sequence reads with an average read length of < 500 bp. Contig assembly reduced this to 8,243 contigs. These were submitted to NCBI-BLASTn (using Viridplantae TaxID: 33090 and Virus TaxID: 10239 as the limiting organisms) and the outputs placed 76% of the contigs in a category of plant cellular sequences, 23% were unknown, and 0.1% were categorized as virus contigs (**Figure 2B**). Closer examination of the population of the 0.1% population of virus contigs determined that 68% of these were assigned to *Caulimoviridae* (**Figure 2B**). Three large contigs within this population were identified with high similarity ($e < 1 \times 10^{-30}$) to CaYMV isolate V17 ORF3 gene (EF189148.1), *Sugarcane bacilliform virus* isolate Batavia D, complete genome (FJ439817.1) and *Banana streak CA virus* complete genome (KJ013511). Examining the three contigs, we manually joined two of these to produce a full-length virus genome.

We compared the virus genome length contigs produced by DNA and RNA sequencing as a mutual scaffold to confirm the presence of two full-length virus genomes. One full-length virus genome of 6,966 bp was tentatively named *Canna yellow mottle associated virus* 1 (CaYMAV-1) (**Figure 3A**). The second genome was 7,385 bp and a variant of CaYMV infecting *Alpinia purpurata*

(CaYMV-Ap01) (**Figure 3A**).

Finally, PCR primers which were designed to clone ~1,000 bp fragment of each virus, were used to differentially detect both genomes in a population of 227 canna plants representing nine commercial varieties. In many instances individual plants were infected with both viruses. We provide an example of RT-PCR detection of CaYMAV-1 and CaYMV-Ap01 in the 12 plants. Three of these were positive only for CaYMV-Ap01 and nine were positive for both viruses (**Figure 3B**).

FIGURE AND TABLE LEGENDS:

Figure 1: Virus nucleic acid preparations and NGS workflow. (**A**) Agarose (1.0%) gel electrophoresis of 565 bp PCR fragments of CaYMV genomes. Two PCR products were detected in samples prepared from the white pellet (lanes 1, 2) but not in the green pellet sample (lane 3). Positive control (+) represents a PCR product amplified from infected plant DNA that was isolated using an automated method involving standard paramagnetic cellulose particles. Lane L contains the DNA ladder used as a standard for measuring the size of linear DNA bands in sample lanes. (**B**) Example of virus particle viewed by transmission electron microscopy in the white pellet recovered by crude fractionation of infected canna leaves. (**C**) Agarose (0.8%) gel electrophoresis of DNA recovered from the green (lane 1) and white (lane 2) pellets that tested positive by PCR in panel A. The red and yellow dots next to lane 2 identify two high molecular weight DNA bands that occur in the white fraction. (**D**) Agarose (1%) gel electrophoresis of total RNA recovered by column-based RNA purification. Lane L contains the DNA ladder used as a standard for measuring the size of linear bands in sample lanes. Lane 1–6 contains RNA isolated from infected canna leaves which were pooled to a single sample for ribo-depletion and RNA-seq. (**E**) Schematic pipeline of nucleic acid isolations, library preparation, sequencing, contig assembly, and virus genome discovery.

Figure 2: Krona charts visualizing the taxonomic categories of contigs. (**A**) The chart on the left shows the abundance and taxonomic distribution of contigs assembled from the crude virus preparation. The right chart depicts the proportions of virus contigs associated with the *Caulimoviridae* family, *Badnavirus* genus, and three closely related species. (**B**) The panel on the left shows the abundance of contigs derived from RNA-seq based on their taxonomic distribution. On the right is the graph depicting the abundance of contigs within the population of virus contigs associated with the *Caulimoviridae* family, *Badnavirus* genus, and three closely related species.

Figure 3. Characterization of CaYMAV-1 and CaYMV-Ap01 genomes. (**A**) Diagrammatic representation of *Canna yellow mottle associate virus 1* (CaYMAV) and *Canna yellow mottle virus* similar to the genome isolated from *Alpinia purpurata (CaYMV-Ap01).* Nucleotide positions 1–10 is identified as the start of the genome and contains a $tRNA^{met}$ anticodon site typical of most badnavirus genomes. The stop and start positions for translation of open reading frame (ORF) 1 and 2 are adjacent. These proteins have unknown functions. ORF3 is a polyprotein containing zinc finger (ZnF), protease (Pro), reverse transcriptase (RT), and RNAse H domains. A 3' poly(A) signal sequence is conserved for both virus genomes. (**B**) RT-PCR analysis was carried out using RNA isolated from virus infected leaves and primers that detect CaYMAV and CaYMV-Ap01. In the same population of 12 plants, three were infected with CaYMV-Ap01 only, whereas the

remaining were infected with both CaYMAV and CaYMV-Ap01. (+) indicates positive control and (-) indicates negative control. This figure is reproduced/modified from Wijayasekara *et al.*[24] with permission.

DISCUSSION:

In recent years a variety of methods have been employed to study plant virus biodiversity in natural environments which include enriching for virus-like particles (VLP) or virus specific RNA or DNA[2,3,44-46] . These methods are followed by NGS and bioinformatic analysis. The goal of this study was to find the causal agent of a common disease in a cultivated plant. The disease was reported to be the result of an unknown virus that has non-enveloped bacilliform particles, and for which only a 565 bp fragment has been cloned[47]. This information was sufficient for prior researchers to hypothetically assign the virus to the genus *Badnavirus* within the family *Caulimoviridae*. While prior reports hypothesized that canna mottle disease in canna lilies was the result of a single badnavirus, using the metagenomics approach outlined in this study, we determined that the disease was caused by two tentative badnavirus species[24]. Thus, the strength of using a metagenome approach to discover the causal agent of a disease is that we can now identify situations where there may be more than one cause.

Our approach combining DNA and RNA sequencing data is thorough and also demonstrates that the outcomes using two approaches yielded consistent results and confirmed the presence of two related viruses. We employed a modified procedure for isolation of caulimoviruses and produced a sample that was enriched for virus associated nucleic acids and that were protected within the virus capsid. A service laboratory was contracted to carry out DNA sequencing. The essential concept for *de novo* sequencing is that DNA polymerase incorporates the fluorescent labeled nucleotides into a DNA template strand during sequential cycles of DNA synthesis. The contigs assembled followed by NGS were submitted into a bioinformatic workflow producing a few contigs that were identified as virus contigs. Further confirmation of two virus genomes[10,24,48-50] was obtained through bioinformatic analysis of RNA-seq data obtained from ribo-depleted RNA preparations. One interesting outcome was to learn that the populations of sequences recovered by DNA and RNA sequencing provided similar distributions of non-viral and viral nucleic acids. For DNA and RNA sequencing, < 0.5% of sequences were of virus origin. Within the population of virus sequences 78–82% belonged to the family *Caulimoviridae*. By comparing the assembled virus contigs from DNA and RNA sequencing, we confirmed that the two assembled genomes occurred in both datasets.

A concern of using only DNA sequencing to identify the new virus genomes is that the badnavirus genome is an open circular DNA. We surmised that sequences overlapping discontinuities in the genome might present obstacles for genome assembly from contigs. Initial examination of the DNA sequencing results revealed two similar virus genomes. We hypothesized that these genomes either represented genetic diversity of a species that has not been studied, or represented two species co-infecting the same plant[24]. Therefore, the collective bioinformatic analysis of datasets obtained by NGS DNA and RNA sequencing, enabled the confirmation of the presence of two full length genomes.

There is another report which developed an alternative method for extracting VLP and nucleic acids from plant homogenates for metagenomic studies, based on procedures to recover DNA from *Cauliflower mosaic virus* (CaMV; a caulimovirus)[3]. This approach identified novel RNA and DNA virus sequences in non-cultivated plants. The steps derived from the caulimovirus isolation procedure used in this study to discover the causal agent of a disease of cultivated plants are unlike the steps derived for extracting VLP from naturally infected plants[24]. The success of both modified methods suggests that the framework procedure for caulimovirus isolation may be a valuable starting point for metagenomic studies of plant viruses in general.

ACKNOWLEDGMENTS:

Research was funded by Oklahoma Center for Advancement of Science and Technology Applied Research Program Phase II AR 132-053-2; and by the Oklahoma Department of Agriculture Specialty Crops Research Grant Program. We thank Dr. HongJin Hwang and the OSU Bioinformatics Core Facility which was supported by grants from NSF (EOS-0132534) and NIH (2P20RR016478-04, 1P20RR16478-02 and 5P20RR15564-03).

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES:

1 Dijkstra, J. & Jager, C. *P*. *Practical Plant Virology : Protocols and Exercises*. 1 edn, (Springer-Verlag Berlin Heidelberg, 1998).

2 Roossinck, M. J. Plant virus metagenomics: biodiversity and ecology. *Annu Rev Genet.* **46** 359-369, doi:10.1146/annurev-genet-110711-155600, (2012).

3 Melcher, U. *et al.* Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from Ambrosia psilostachya. *J Virol Methods.* **152** (1-2), 49-55, doi:10.1016/j.jviromet.2008.05.030, (2008).

4 Stobbe, A. H., Schneider, W. L., Hoyt, *P*. R. & Melcher, U. Screening metagenomic data for viruses using the e-probe diagnostic nucleic Acid assay. *Phytopathology.* **104** (10), 1125-1129, doi:10.1094/PHYTO-11-13-0310-R, (2014).

5 Borah, B. K. *et al.* Bacilliform DNA-containing plant viruses in the tropics: commonalities within a genetically diverse group. *Mol Plant Pathol.* **14** (8), 759-771, doi:10.1111/mpp.12046, (2013).

6 Bousalem, M., Douzery, E. J. & Seal, S. E. Taxonomy, molecular phylogeny and evolution of plant reverse transcribing viruses (family Caulimoviridae) inferred from full-length genome and reverse transcriptase sequences. *Arch Virol.* **153** (6), 1085-1102, doi:10.1007/s00705-008-0095-9, (2008).

7 Geering, A. D. *et al.* Banana contains a diverse array of endogenous badnaviruses. *J Gen Virol.* **86** (Pt 2), 511-520, doi:10.1099/vir.0.80261-0, (2005).

8 Kunii, M. *et al.* Reconstruction of putative DNA virus from endogenous rice tungro bacilliform virus-like sequences in the rice genome: implications for integration and evolution. *BMC Genomics.* **5** 80, doi:10.1186/1471-2164-5-80, (2004).

9 Laney, A. G., Hassan, M. & Tzanetakis, I. E. An integrated badnavirus is prevalent in **Figure** germplasm. *Phytopathology.* **102** (12), 1182-1189, doi:10.1094/PHYTO-12-11-0351,

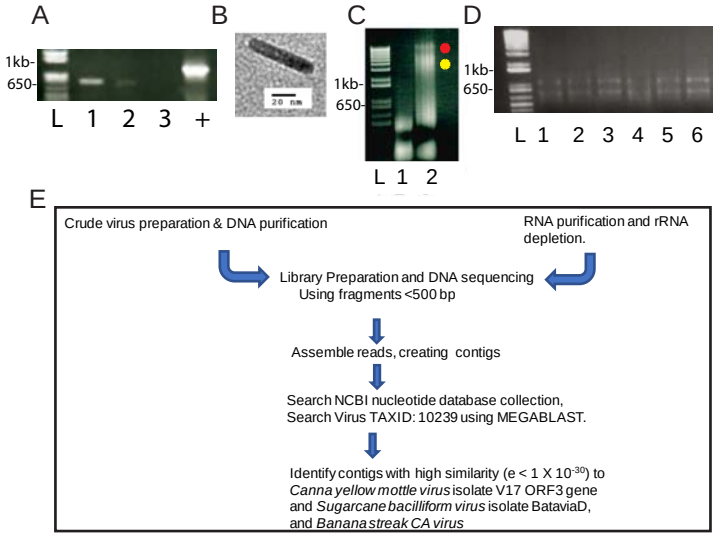
(2012).
10 Gambley, C. F., Geering, A. D., Steele, V. & Thomas, J. E. Identification of viral and non-viral reverse transcribing elements in pineapple (Ananas comosus), including members of two new badnavirus species. *Arch Virol.* **153** (8), 1599-1604, doi:10.1007/s00705-008-0160-4, (2008).
11 Gayral, *P*. *et al.* A single Banana streak virus integration event in the banana genome as the origin of infectious endogenous pararetrovirus. *J Virol.* **82** (13), 6697-6710, doi:10.1128/JVI.00212-08, (2008).
12 Lyttle, D. J., Orlovich, D. A. & Guy, *P*. L. Detection and analysis of endogenous badnaviruses in the New Zealand flora. *AoB Plants.* **2011** plr008, doi:10.1093/aobpla/plr008, (2011).
13 Le Provost, G., Iskra-Caruana, M. L., Acina, I. & Teycheney, *P*. Y. Improved detection of episomal Banana streak viruses by multiplex immunocapture PCR. *J Virol Methods.* **137** (1), 7-13, doi:10.1016/j.jviromet.2006.05.021, (2006).
14 Singh, K., Talla, A. & Qiu, W. Small RNA profiling of virus-infected grapevines: evidences for virus infection-associated and variety-specific miRNAs. *Funct Integr Genomics.* **12** (4), 659-669, doi:10.1007/s10142-012-0292-1, (2012).
15 Alfson, K. J., Beadles, M. W. & Griffiths, A. A new approach to determining whole viral genomic sequences including termini using a single deep sequencing run. *J Virol Methods.* **208** 1-5, doi:10.1016/j.jviromet.2014.07.023, (2014).
16 Kreuze, J. F. *et al.* Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology.* **388** (1), 1-7, doi:10.1016/j.virol.2009.03.024, (2009).
17 Zheng, Y. *et al.* VirusDetect: An automated pipeline for efficient virus discovery using deep sequencing of small RNAs. *Virology.* **500** 130-138, doi:10.1016/j.virol.2016.10.017, (2017).
18 James, A. *P*., Geijskes, R. J., Dale, J. L. & Harding, R. M. Molecular characterisation of six badnavirus species associated with leaf streak disease of banana in East Africa. *Annals of Applied Biology.* **158** (3), 346-353, doi:DOI 10.1111/j.1744-7348.2011.00466.x, (2011).
19 Baranwal, V. K., Sharma, S. K., Khurana, D. & Verma, R. Sequence analysis of shorter than genome length episomal Banana streak OL virus like sequences isolated from banana in India. *Virus Genes.* **48** (1), 120-127, doi:10.1007/s11262-013-0984-1, (2014).
20 Sukal, A., Kidanemariam, D., Dale, J., James, A. & Harding, R. Characterization of badnaviruses infecting Dioscorea spp. in the Pacific reveals two putative novel species and the first report of dioscorea bacilliform RT virus 2. *Virus Res.* **238** 29-34, doi:10.1016/j.virusres.2017.05.027, (2017).
21 Bömer, M., Turaki, A. A., Silva, G., Kumar, *P*. L. & Seal, S. E. A sequence-independent strategy for amplification and characterisation of episomal badnavirus sequences reveals three previously uncharacterised yam badnaviruses. *Viruses.* **8** (7), doi:10.3390/v8070188, (2016).
22 Momol, M. T., Lockhart, B. E. L., Dankers, H. & Adkins, S. Canna yellow mottle virus detected in Canna in Florida. *Plant Health Progress.* (August), 2-4, doi:10.1094/PHP-2004-0809-01-HN., (2004).
23 Zhang, J. *et al.* Characterization of Canna yellow mottle virus in a new host, Alpinia purpurata, in Hawaii. *Phytopathology.* **107** (6), 791-799, doi:10.1094/PHYTO-04-16-0160-

R, (2017).
24 Wijayasekara, D. *et al.* Molecular characterization of two badnavirus genomes associated with Canna yellow mottle disease. *Virus Res.* **243** 19-24, doi:10.1016/j.virusres.2017.10.001, (2018).
25 Covey, S. N., Noad, R. J., al-Kaff, N. S. & Turner, D. S. Caulimovirus isolation and DNA extraction. *Methods Mol Biol.* **81** 53-63, doi:10.1385/0-89603-385-6:53, (1998).
26 Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular cloning: A laboratory manual*. 2nd edn, (Cold Spring Harbor Press, 1989).
27 Radford, A. D. *et al.* Application of next-generation sequencing technologies in virology. *J Gen Virol.* **93** (Pt 9), 1853-1868, doi:10.1099/vir.0.043182-0, (2012).
28 Kanagal-Shamanna, R. Emulsion PCR: Techniques and Applications. *Methods Mol Biol.* **1392** 33-42, doi:10.1007/978-1-4939-3360-0_4, (2016).
29 Getts, D. R. *et al.* Targeted blockade in lethal West Nile virus encephalitis indicates a crucial role for very late antigen (VLA)-4-dependent recruitment of nitric oxide-producing macrophages. *J Neuroinflammation.* **9** 246, doi:10.1186/1742-2094-9-246
1742-2094-9-246 [pii], (2012).
30 van Dijk, E. L., Jaszczyszyn, Y. & Thermes, C. Library preparation methods for next-generation sequencing: tone down the bias. *Exp Cell Res.* **322** (1), 12-20, doi:10.1016/j.yexcr.2014.01.008, (2014).
31 Healthcare, G. *Gel filtration principles and methods*, 2010).
32 Goff, S. *et al.* The iPlant Collaborative: Cyberinfrastructure for Plant Biology. *Frontiers in Plant Science.* **2**, doi:10.3389/fpls.2011.00034, (2011).
33 Lin, Z. *et al.* Next-generation sequencing and bioinformatic approaches to detect and analyze influenza virus in ferrets. *J Infect Dev Ctries.* **8** (4), 498-509, doi:10.3855/jidc.3749, (2014).
34 Artimo, *P*. *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40** (Web Server issue), W597-603, doi:10.1093/nar/gks400, (2012).
35 Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* **5** 113, doi:10.1186/1471-2105-5-113, (2004).
36 Hung, J. H. & Weng, Z. Sequence Alignment and Homology Search with BLAST and ClustalW. *Cold Spring Harb Protoc.* **2016** (11), pdb.prot093088, doi:10.1101/pdb.prot093088, (2016).
37 Sohpal, V. K., Dey, A. & Singh, A. MEGA biocentric software for sequence and phylogenetic analysis: a review. *Int J Bioinform Res Appl.* **6** (3), 230-240, doi:L323363736563620 [pii], (2010).
38 Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res.* **40** (15), e115, doi:10.1093/nar/gks596, (2012).
39 Dhaliwa, A. DNA extraction and purification. *Mater Methods.* **3** 191, doi:dx.doi.org/10.13070/mm.en.3.191, (2013).
40 Moeller, J. R., Moehn, N. R., Waller, D. M. & Givnish, T. J. Paramagnetic cellulose DNA isolation improves DNA yield and quality among diverse plant taxa. *Appl. Plant Sci.* **2(10)** (2014).
41 Moeller, J. R. *e. a.* Paramagnetic cellulose DNA isolation improves DNA yield and quality among diverse plant taxa. *Appl. Plant Sci.* **2(10)** (2014).

42 Grooms, K. Review: Improved DNA Yield and Quality from Diverse Plant Taxa. (2015).
43 Nishimori, A. *et al.* In vitro and *in vivo* antivirus activity of an anti-programmed death-ligand 1 (PD-L1) rat-bovine chimeric antibody against bovine leukemia virus infection. *PLoS One.* **12** (4), e0174916, doi:10.1371/journal.pone.0174916, (2017).
44 Rojas, M. R. & Gilbertson, R. L. in *Plant Virus Evolution* Vol. 1 (ed M. J. Roossinck) 27-51 (Springer-Verlag, 2008).
45 Roossinck, M. J. The big unknown: plant virus biodiversity. *Curr Opin Virol.* **1** (1), 63-67, doi:10.1016/j.coviro.2011.05.022, (2011).
46 Roossinck, M. J., Martin, D. *P*. & Roumagnac, *P*. Plant Virus Metagenomics: Advances in Virus Discovery. *Phytopathology.* **105** (6), 716-727, doi:10.1094/PHYTO-12-14-0356-RVW, (2015).
47 Momol, M. T., Lockhart, B. E. L., Dankers, H. & Adkins, S. in *Plant Health Progress* (Online, 2004).
48 Eni, A., Hughes, J. D., Asiedu, R. & Rey, M. Sequence diversity among badnavirus isolates infecting yam (Dioscorea spp.) in Ghana, Togo, Benin and Nigeria. *Archives of Virology.* **153** (12), 2263-2272, doi:DOI 10.1007/s00705-008-0258-8, (2008).
49 Harper, G. *et al.* The diversity of Banana streak virus isolates in Uganda. *Arch Virol.* **150** (12), 2407-2420, doi:10.1007/s00705-005-0610-1, (2005).
50 Muller, E. & Sackey, S. Molecular variability analysis of five new complete cacao swollen shoot virus genomic sequences. *Arch Virol.* **150** (1), 53-66, doi:10.1007/s00705-004-0394-8, (2005).

Figure

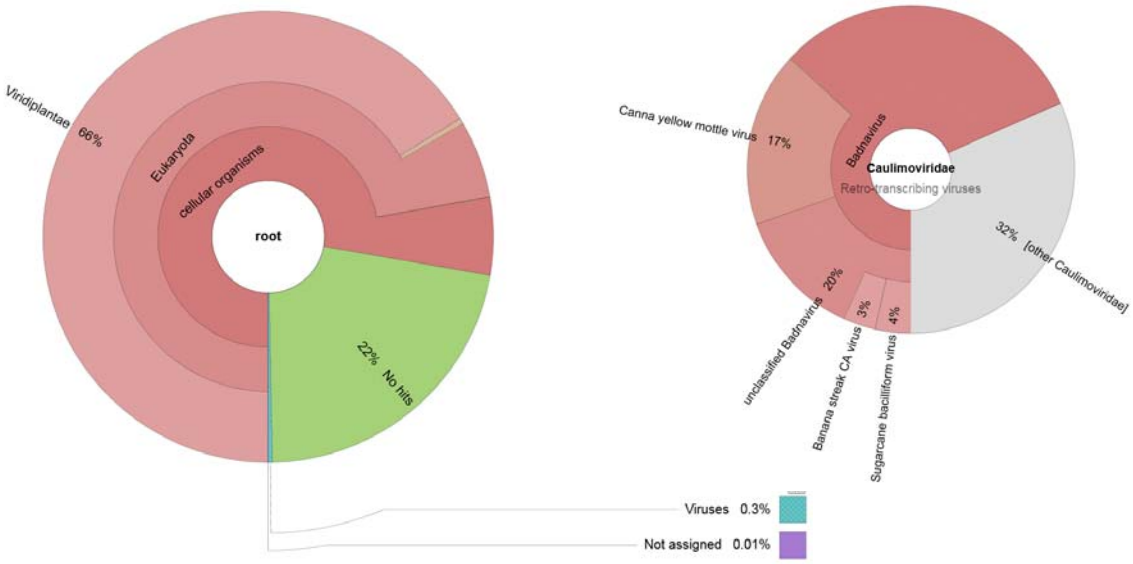
Click here to download Figure JOVE_Fig1_rev.pdf



Figure

Click here to download Figure JOVE_Fig2.pdf

A



B

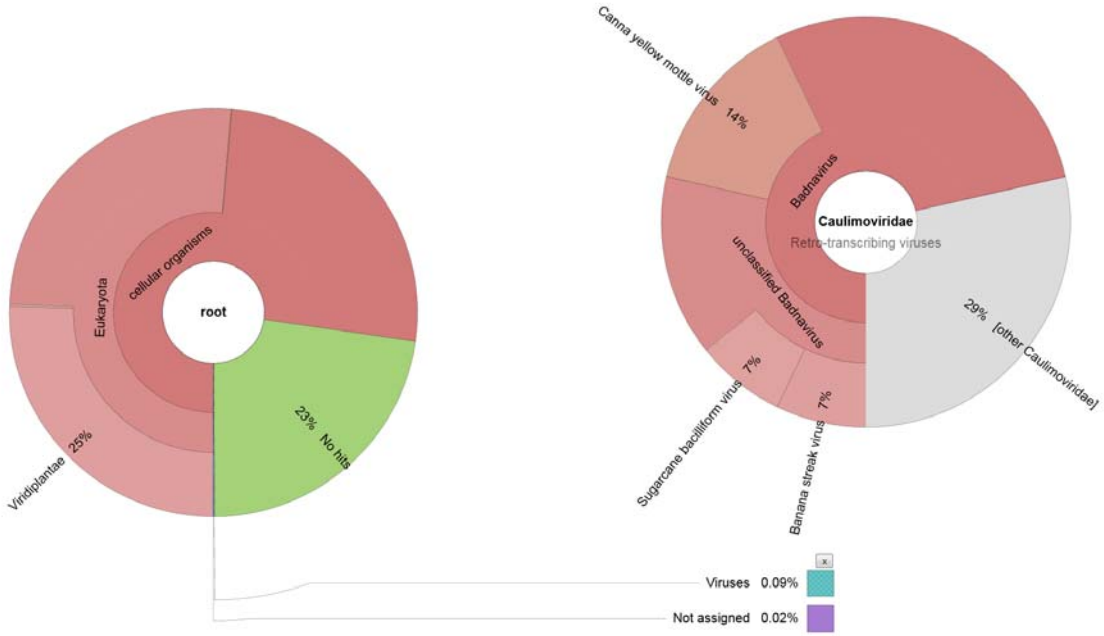


Figure 1

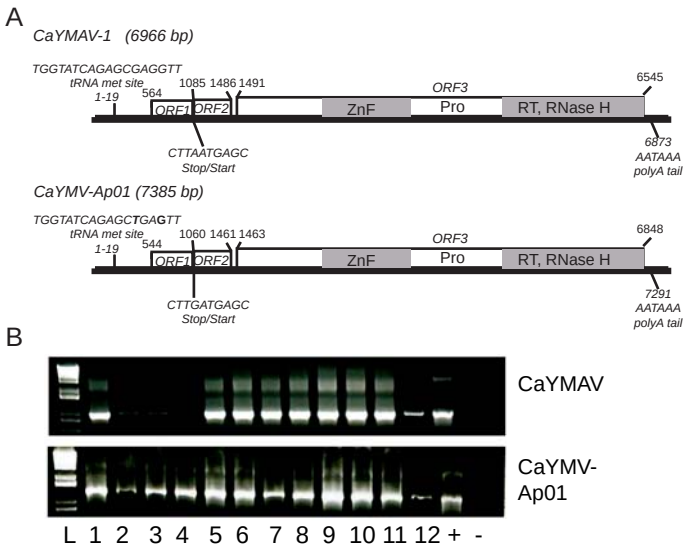


Figure 3

Materials Table

Click here to download Materials Table Materials3.xlsx

Material/Equipment	Company
NaH2PO4	Sigma-Aldrich St. Louis MO
Na_2HPO_4	Sigma-Aldrich
Na_2SO_3	Thermo-Fisher Waltham, MA
urea	Thermo-Fisher
Triton X-100	Sigma-Aldrich
Cheesecloth	VWR Radnor, PA
Tris	Thermo-Fisher
$MgCl_2$	Spectrum, Gardena, CA
EDTA	Spectrum
Proteinase K	Thermo-Fisher
phenol:chloroform:isoamylalcohol	Sigma-Aldrich
DNAse I	Promega
95% ethanol	Sigma-Aldrich
Laboratory blender	VWR
Floor model ultracentrifuge &Ti70 rotor	Beckman Coulter, Irving TX
Floor model centrifuge and JA-14 rotor	Beckman Coulter
Magnetic stir plate	VWR
Rubber policeman	VWR
2100 bioanalyzer Instrument	Agilent Genomics, Santa Clare, CA
2100 Bioanalyzer RNA-Picochip	
2100 Bioanalyzer DNA-High Sensitive chip	
Nanodrop spectrophotometer	Thermo-Fisher
Plant total RNA isolation kit	Sigma-Aldrich
RNase-free water	VWR
RNA concentrator spin column	Zymo Research, Irvine, CA
rRNA removal kit	Illumina, San Diego, CA
DynaMag-2 Magnet	ThermoFisher
RNA enrichment system	Roche
Agarose	Thermo-Fisher
Ethidium bromide	Thermo-Fisher
pGEM-T +JM109 competent cells	Promega, Madison, WI
pFU Taq polymerase	Promega
dNTPs	Promega
PCR oligonucleotides	IDT, Coralvill, IA
Miniprep DNA purification kit	Promega
PCR clean-up kit	Promega
pDRAW32 software	ACAClone

MEGA6.0 software	MEGA
Primer 3.0	Simgene.com
Quant-iT™ RiboGreen™ RNA Assay Kit	Thermo-Fisher
GS Junior™ pyrosequencing System	Roche
GS Junior Titanium EmPCR Kit (Lib-A)	Roche
GS Jr EmPCR Bead Recovery Reagents	Roche
GS Junior EmPCR Reagents (Lib-A)	Roche
GS Jr EmPCR Oil & Breaking Kit	Roche
GS Jr Titanium Sequenicing kit*	Roche
GS Jr. Titanium Picotiter Plate Kit	Roche
IKA Turrax mixer	
IKA Turrax Tube (specialized mixer)	
GS Nebulizers Kit	Roche
GS Junior emPCR Bead Counter	Roche
GS Junior Bead Deposition Device	Roche
Counterweight & Adaptor for the Bead Deposition Devices	Roche
GS Junior Software	Roche
GS Junior Sequencer Control v. 3.0	Roche
GS Run Processor v. 3.0	Roche
GS De Novo Assembler v. 3.0	Roche
GS Reference Mapper v. 3.0	Roche
GS Amplicon Variant Analyzer v. 3.0	Roche

Catalog Number	Comments/Description
S5976	Grinding buffer for virus purification
S0751	Grinding buffer for virus purification
28790	Grinding buffer for virus purification
PB169-212	Homogenate extraction
X-100	Homogenate extraction
21910-107	Filter homogenate
BP152-5	Pellet resuspension& DNA resuspension buffers
M1035	Pellet resuspension buffer
E1045	Stops enzyme reactions
25530	DNA resuspension buffer
P2069	Dissolve virion proteins
M6101	Degrade cellular DNA from extracts
6B-100	Virus DNA precipitation
58984-030	Grind leaf samples
A94471	Separation of cellular extracts
369001	Separation of cellular extracts
75876-022	Mixing urea into samples overnight
470104-462	Dissolve virus pellet
G2939BA	Sensitive detection of DNA and RNA quality and quantity
5067-1513	Microfluidics chip used to move, stain and measure RNA quality in a 2100 Bioa
5067-4626	Microfluidics chip used to move, stain and measure DNA quality in a 2100 Bioa
ND-2000	Analysis of DNA/RNA quality at intermediate steps of procedures
STRN50-1KT	Isolate RNA for RNA-seq
10128-514	Resuspension of DNA and RNA for NGS
R1013	Prepare RNA for RNA-seq
MRZPL116	Prepare RNA for RNA-seq
12321D	Prepare RNA for RNA-seq
7277300001	Prepare RNA for RNA-seq
16500100	Gel analysis of DNA/RNA quality at intermediate steps of procedures
15585011	Agarose gel staining
A3610	Clone genome fragments
M7741	PCR amplify virus genome
U1511	PCR amplify virus genome
Custom order	PCR amplify virus genome
A1330	Plasmid DNA purification prior to sequencing
A9281	Prepare PCR products for cloning
	Computer analysis of circular DNA and motifs

	Molecular evolutionary genetics analysis
R11490	Fluorometric determination of RNA quantity
5526337001	Sequencing platform
5996520001	Reagents for emulsion PCR
5996490001	Reagents for emulsion PCR
5996538001	Reagents for emulsion PCR
5996511001	Reagents for emulsion PCR
5996554001	Includes sequencing reagents, enzymes, buffers, and packing beads
5996619001	Sequencing plate with associated reagents and gaskets
3646000	Special mixer used with Turrax Tubes
20003213	Specialized mixing tubes with internal rotor for creating emulsions
5160570001	Nucleic acid size fractionator for use during library preparations
05 996 635 001	Library bead counter
05 996 473 001	Holder for Picotiter plate during centrifugation
05 889 103 001	Used to balance deposition device with picotiter plate centrifugation
05 996 643 001	Software suite for controlling the instrument, collecting and analyzing data
	(Included in item 05 996 643 001 above)
	(Included in item 05 996 643 001 above)
	(Included in item 05 996 643 001 above)
	(Included in item 05 996 643 001 above)
	(Included in item 05 996 643 001 above)

nalyzer
nalyzer

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: Combining Standard virus partial extracti.....

Author(s): Jeanmarie Verchot

Item 1 (check one box): The Author elects to have the Materials be made available (as described at http://www.jove.com/author) via: ☑ Standard Access ☐ Open Access

Item 2 (check one box):

☑ The Author is NOT a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. Defined Terms. As used in this Article and Video License Agreement, the following terms shall have the following meanings: **"Agreement"** means this Article and Video License Agreement; **"Article"** means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; **"Author"** means the author who is a signatory to this Agreement; **"Collective Work"** means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; **"CRC License"** means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; **"Derivative Work"** means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; **"Institution"** means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; **"JoVE"** means MyJove Corporation, a Massachusetts corporation and the publisher of *The Journal of Visualized Experiments;* **"Materials"** means the Article and / or the Video; **"Parties"** means the Author and JoVE; **"Video"** means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. Background. The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. Grant of Rights in Article. In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

jove
VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

11. Indemnification. The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in *JoVE* or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

12. Fees. To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

13. Transfer, Governing Law. This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement required per submission.

CORRESPONDING AUTHOR:

Name: Jeanmarie Verchot

Department: The Dallas Research & Extension Center

Institution: Texas A&M Agrilife

Article Title: Combing standard virus extracts with NGS...

Signature: Jeanmarie Verchot Date: 1/9/18

Please submit a signed and dated copy of this license by one of the following three methods:

1) Upload a scanned copy of the document as a pfd on the JoVE submission site;
2) Fax the document to +1.866.381.2236;
3) Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02139

For questions, please email submissions@jove.com or call +1.617.945.9051

Dear Editor

Thank you for taking the time to read the paper and provide the extensive reviewer comments. I have carefully gone through each comment and made the appropriate changes. Given the extensive comments, I have chosen to highlight my rebuttal in red below. I uploaded a revised Figure 1 and a new figure 3 in response to reviewer comments. These are explained below.
I hope that you find this revision acceptable for publication

Sincerely

Jeanmarie Verchot

Editorial comments:
Changes to be made by the Author(s):
1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues. The JoVE editor will not copy-edit your manuscript and any errors in the submitted revision may be present in the published version. Carefully edited language for grammar, tense, voice and the markup is provided.
2. Please ensure that the references appear as the following: [Lastname, F.I., LastName, F.I., LastName, F.I. Article Title. Source. Volume (Issue), FirstPage – LastPage, doi: DOI (YEAR).] For more than 6 authors, list only the first author then et al. Corrected bibliography and deleted duplicate references.
3. Please define all abbreviations before use. This includes Polymerase chain reaction. Notably the abbreviations spelled out in the abstract, where then re-defined in the introduction and the remaining body of the text. In the abstract we spelled out and provided abbreviations for NBCI and BLASTn and NGS and RNA-seq. Then in the introduction this was repeated, treating the summary as a separate entity.
4. Please revise the table of the essential supplies, reagents, and equipment. The table should include the name, company, and catalog number of all relevant materials in separate columns in an xls/xlsx file. Changed. Revised table to use generic terms and added more catalog numbers
5. Please rephrase the Short Abstract to clearly describe the protocol and its applications in complete sentences between 10-50 words: “Here, we present a protocol to …” Changed.
6. The current Long Abstract is over the 150-300-word limit. Please rephrase the Long Abstract to more clearly state the goal of the protocol. Done
7. JoVE cannot publish manuscripts containing commercial language. This includes trademark symbols (™), registered symbols (®), and company names before an instrument or reagent. Please remove all commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials and Reagents.
For example: Triton X, Agilent 2100 Bioanalyzer, The Spectrum™ Plant Total RNA kit, Ribo-zero™ Plant rRNA removal kit (Illumina), SeqCap™ RNA enrichment system (Roche Biomedical), the Oklahoma State University 207 Bioinformatics and Genomics Core Facility, Roche 454-Junior Genomic Sequencer, Roche Genome Assembly software (ver 2.7), (Promega), Nanodrop, DynaMag-2, Megan6, etc.
8. Please remove trademark (™) and registered (®) symbols from the Table of Equipment and Materials. Done
9. Please ensure that all text in the protocol section is written in the imperative tense as if telling someone how to do the technique (e.g., “Do this,” “Ensure that,” etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Avoid usage of phrases such as “could be,” “should be,” and “would be” throughout the Protocol. Any text that cannot be written in the

imperative tense may be added as a “Note.” However, notes should be concise and used sparingly. Please include all safety procedures and use of hoods, etc.
10. The Protocol should be made up almost entirely of discrete steps without large paragraphs of text between sections. Please simplify the Protocol so that individual steps contain only 2-3 actions per step and a maximum of 4 sentences per step. DONE
11. The Protocol should contain only action items that direct the reader to do something. DONE
12. Please add more details to your protocol steps. Please ensure you answer the “how” question, i.e., how is the step performed? Alternatively, add references to published material specifying how to perform the protocol action. Elaborated the steps for organic extraction of DNA, ethanol precipitation and added Sambrook as a reference. Added references to each step in Sections 2, 3 and 4,
13. 1.1: Which leaves are used? I used leaves from canna plants. But this could be from any plant. I would prefer not to name the plant in this experiment and leave it open to use any plant.
14. 1.6: How is Phenol: Chloroform: Isoamyl Alcohol purification performed? How do you use gel filtration desalting column? If you are interested in filming these steps, then please provide details. What do you resuspend DNA in to proceed for NGS? Better elaborated the organic extraction in line 1.10. Then 1.11 explains ethanol precipitation. Added safety to these steps. Also added references since this is typically a kit that you purchase.
15. 1.7: What do you look for during in the spectrophotometer reading? How do you check the DNA quality using Bioanalyzer?Changed the nanodrop in section 1.13 and 2.2 to indicate that OD ration of 1.85 and 2.0 is desired. The bioanalyzer is explained in section 1.14
16. 2.1: Please provide details on RNA isolation steps. We need hard experimental steps to be filmed. These steps are elaborated in more detail in steps 2.1, 2.2, and 2.3

17. 2.2: How do you do so? What do you dissolve the RNA pellet in? Revised to resuspend in RNAse free water.
18. 2.3: Please provide hard experimental steps for RNA concentrating. Elaborated now in step 2.2. I deleted the original step 2.2 which was redundant with step 2.3 and added confusion.
19. 2.4: How do you use this kit to remove plant ribosomal RNA? What is magnetic bead used for. Please provide hard experimental steps in the order of the steps being performed. e.g. load column, spin the tube, add buffer, wash the column, spin the tube, add water onto the column, spin and recover RNA, etc. This method uses a magnetic stand to separate ribosome from mRNA. I elaborated this in section 2.5 and 2.6. I added a note to explain the kit processes.
20. 2.5: This talks about ds DNA library? Did you perform RT for the RNA isolated in the above steps? How is the library generated? The core facility generates the library. We send them the RNA and they perform the RT reaction. However, we added a section 2.9 to explain how this is done.
21. 3.1: This can be coverted to a note without mention of the core facility. Added notes to 2.9 and 3.1
22. 3.2: Please remove hyperlinks from the protocol steps and sufficiently refer it in the reference or table of materials section. Please provide steps as to how the NGS is performed. This can be in the form of hard experimental steps, button clicks, GUI etc. In its present format it will be difficult to film. The sequencing is carried out by loading samples to a machine. We added some details to step 3.1 and then the post sequencing analysis in 3.2 was changed to reference rather than hot links to websites.
23. 4.1: how? What are the primers used? How do you differentiate the virus genome from the plant genome? Do you use any positive/negative control? what program is used? Section 4.1 is better elaborated.
24. At present section 3 cannot be filmed. It is pretty boring. A flow chart of the work may be the best way to present this.
25. 4: Please provide hard experimental steps. I did this and expanded the section to include more steps. This section now has 6 steps, notes, and references.

26. Please highlight 2.75 pages or less of the Protocol (including headings and spacing) that identifies the essential steps of the protocol for the video, i.e., the steps that should be visualized to tell the most cohesive story of the Protocol. The highlighted steps should form a cohesive narrative with a logical flow from one highlighted step to the next. Remember that non-highlighted Protocol steps will remain in the manuscript, and therefore will still be available to the reader. DONE
27. Please include all relevant details that are required to perform the step in the highlighting. For example: If step 2.5 is highlighted for filming and the details of how to perform the step are given in steps 2.5.1 and 2.5.2, then the sub-steps where the details are provided must be highlighted. I did not break these out into substeps but these are obvious sentences in the steps.
28. For the representative result section, please refer figures in order of their numbering. I added 2 new panels to Figure 1, to address comments for reviewer 1 (see below). Then I moved the last 2 panels of figure 1 to a new figure 3 in order for things to be put in proper order.
29. Each Figure Legend should include a title and a short description of the data presented in the Figure and relevant symbols. This is all changed in Figure 1, 2, and 3.
30. Please explain clearly how RNA and DNA are processed and what results are obtained from them individually. Reorganized the results line 346-366 to explain the workflow of identifying the best DNA fraction for preparing a library and then sequencing. I added figures 1A, B, and C to more clearly show the decision processes for picking the correct fraction. Then added a paragraph at lines 368-371 explaining RNA is extracted in parallel and both samples were inputted to similar workflow (Figure 1E). Reorganizing this section should make it clearer to the reader. Also the methods are more elaborated for clarity for DNA preparation method and RNA preparation method.
31. Please obtain explicit copyright permission to reuse any figures from a previous publication. Explicit permission can be expressed in the form of a letter from the editor or a link to the editorial policy that allows re-prints. Please upload this information as a .doc or .docx file to your Editorial Manager account. The Figure must be cited appropriately in the Figure Legend, i.e. “This figure has been modified from [citation].” I obtained this permission and am attaching this here.
32. As we are a methods journal, please revise the Discussion to explicitly cover the following in detail in 3-6 paragraphs with citations:
a) Critical steps within the protocol
b) Any modifications and troubleshooting of the technique
c) Any limitations of the technique
d) The significance with respect to existing methods
e) Any future applications of the technique

Reviewers' comments:

Reviewer #1:
Major Concerns:
No evidence (for example, TEM images) is presented that virions of CaYMV were actually isolated by the procedure used. No evidence is presented that the top band in Figure A lane 2 consists of viral DNA and not mainly of host plant DNA. Added a panel A to figure which is a bacilliform virus particle of the correct length.
The virus isolation procedure cited (Covey et al) on pg 11 line 474 was designed to isolate virions of caulimoviruses that may accumulate in virus-encoded (ORF 5) cytoplasmic inclusion bodies. Disruption of these structures would increase the yield of virions. Badnaviruses are not known to either produce or

accumulate in any similar cellular structures. I tried to better address this comment in the discussion to incorporate the notion that virus yields were low because the method is designed to break up inclusions which may not exist. On the other hand we used a crude preparation. So if we took the process all the way through to final purification by differential centrifugation, we would have lost the badnaviruses from the preparation. In fact we did this several times and learned to stop at the crude preparation. The protocol used for pelleting of virions (pg 4 line 167) 136000G for 2 hours will pellet out materials of S values above 75S (based on a clearing factor of 150). This is sufficient to pellet out host cellular debris containing DNA. It is highly unlikely that the quantity of DNA shown in the top band in Fig 1A, lane 2, could have been obtained from CaYMV virions occurring in 100g of infected leaf tissue. This figure also lacks a virus-free plant control. I modified the gel labelling. There are 2 HMW bands which are identified by red and yellow dots. We show in Fig 2 that the sample has an abundance of cellular DNA, so we do not dispute this. Luckily we were able to capture virus DNA in the preparation. We argue that the method provides an enrichment for virus DNA that is necessary for discovery of the new genomes. Bacilliform particles measuring 124-133 nm in length were observed using transmission electron microscopy, which is within the predicted modal length of most badnaviruses (Fig. 1*A*). DNA extracted from the virion preparation produced a concentration of 0.6 µg/mL, which is a low concentration. This procedure is designed for isolating virions of caulimoviruses which accumulate in virus-encoded (ORF5) inclusion bodies. This method of DNA preparation disrupts the inclusion bodies and typically yields greater than 2 mg/ml of caulimovirus virions [23]. The yield of badnavirus DNA is likely lower because badnaviruses are not know to either produce or accumulate in similar cellular structures. However, the recovered amount of DNA is adequate for NGS, which requires a minimum of 10 ng of DNA. The first paragraph of the discussion addresses these concerns. Figure 1A-C also add data to support our assertion that this fraction had virus sequences, although also cellular sequences, in it.

Reviewer #2:

Manuscript Summary:

The manuscript describes the protocol for crude purification of badnaviruses from plant tissues and subsequent DNA isolation from the preparations, followed by high throughput sequencing and sequence data analysis for identification of viruses in such samples. Simultaneously, the protocol for sequencing ribosomal RNA depleted total RNA from the same plants is also described and again the data analyzed with the purpose of virus discovery and characterization.

Major Concerns:

1. I understand the rationale behind using the partial virion purification and NGS to somehow fish-out the episomal viruses, and this also connects nicely to the introduction section. However, I don`t understand how sequencing rRNA depleted total RNA (including viral transcripts) will help to achieve this task; as mentioned in the introduction (in the section about sRNA sequencing) some integrated viral sequences can be transcribed, thus the result here should be more or less similar as when sequencing sRNAs (which are derived from transcripts in this case). Focusing only on the first protocol (virion purification + NGS) would make the story clearer. **This becomes a point that could be debatable on many levels. We would expect tat endogenous sequences, whether obtained from the DNA or RNA sequencing would have some non-viral sequences attached to them in they were integrated sequences. One of these data sets should have given evidence of a virus end joining to cellular DNA or transcripts. We never found any evidence to indicate added non-viral sequences at either end of the virus genome. So, we are not arguing that there is not endogenous sequences in the plant. We are arguing that we could recover intact virus genomes.**

2. How certain you can be that your sequencing results (for the DNA protocol in this case) are really reflecting only (or mostly) virion-derived DNA and not nuclear DNA? Do you have any control for this?

Can you observe virions in partial purifications under the electron microscope? We present an example of particles seen by TEM. We added new panels to Figure 1 demonstrating enrichment for virus particles and virus DNA. In figure 2, we scanned through the sequences and the majority of cellular sequences were organellar, very little nuclear got carried through. Added a statement at line 377 to this point. Took out "eliminating majority of nuclear DNA inline 468 of discussion since we did not present that data.

3. There are several points (especially in the bioinformatics processing of the data), which should be revised for more accuracy and better clarity. Please see specific comments about this below.

-L161: Point 1.4: from the text, it is not totally clear to me how this is done: what are you doing for 1-2h and what for the remaining night? Also when going to point 1.5 (L167): what is transferred to new centrifugation? Changed this and is now in 1.6

-L173: Point 1.6: please refer to published protocol for DNA extraction or provide more details; can the DNA extraction kits be used? Or is the problem in the too high volume of the sample? Maybe explain this here. The point being we were fractionating virus from chromosomal DNA using a standard method that eliminates nuclei and captures cytoplasmic DNA. The fractions cytoplasmic from organellar DNA. These methods have been around for 30 years for virus fractionation. Then you capture the virus nucleic acids from the fraction.

-L175: Give more details about gel filtration desalting columns (also include them in the table of materials or made clear to which item do they refer). This refers to Wizard PCR clean up kit. This is now line 185, 1.11. Added parenthetical statement (normally use kit for PCR clean up).

-L178: Point 1.7: what is the expected result? This is now section. 1.13. The Note indicates the type of output.

-L186-201: This part (2.) is not very clear; e.g. why is RNA participated again after the extraction with columns (is it because of the pooling of extractions?), deleted that statement which was in error. why 12 leaves are used and why SeqCap RNA enrichment system is used (as I understand it would enrich specific selected target transcripts - was this really the case?)…We used it because it was recommended for NGS.

-In L215 the assembly is mentioned and then again in L220 the same is repeated, clarify.

-L265: Showing the Bioanalyzer results would be clearer than the gels. Better elaborated section 2 and section about bioanalyzer

-BLASTn results: when talking about the blast results (especially when mentioning E-values) the database, which was used shoud always be specified - in protocol you mention BLASTn only against Viruses (L221), however, when reading the description on page 7, this does not seem to be the case.

-Also important, you are describing which fractions of sequences were classified as viral or plant taxa at several points in the text (e.g., L270-275, L324, L378-379 and elsewhere); from the text it seems that blast was performed using assembled contigs and in this case the % of classified contigs does not represent the real % of viral sequences in the samples (since, if there is a lot of viral sequences, nicely distributed over the viral genome, this could after de novo assembly result in only one or few contigs). Thus, the reads should be classified directly or the reads could be mapped back to the complete viral genomes to estimate the % of viral sequences in the sample. When using NCBI-BlastN, you can select an organism for preferential screening. We select viruses. The output typically includes sequence assignments to non-virus taxids as well. You are correct the percent of sequences after de novo assembly produces a few contigs. This is key to discover of 2 new virus sequences. I changed the explanation of Figure 2 outputs to reflect that the NCBI blast results were not about original sequences, but were the results of blasting assembled contigs lines 368-377. Figure 2 legend I also changed the description from sequences to contigs.

For the results around RNA sequencing in lines 380-389 I changed the word sequences to contigs. I added a sentence to explain that we recovered 3 contigs but two were manually joined to produce

one full length genome. Then I added a summary in lines 391-394 showing that by comparing the 2 datasets, we basically used them as virtual hybridization or as mutual scaffolds to confirm the presence of 2 new, very similar viruses. We could not have confirmed these 2 viruses without the 2 datasets because we were not certain that we were not looking at quasispecies population of sequences. Also, because NGS involves a single sequencing run, that is not a statistically valid approach to identifying a novel virus. But using 2 approaches (RNA and DNA sequencing), I felt more confident we were presenting 2 new genomes, not a quasispecies.
-L271: 76% of what (I am guessing viral fraction) and then, what is the remaining 24%? I rexamined the numbers in figure 2 and realized tht the explanation of the contigs and full length genome was not adequate. So I changed the paragraph lines 371-379 to be: The majority of contigs categorized as cellular sequences were identified as mitochondrial or chloroplast DNA. Within the dataset of virus contigs, 32 % of the virus contigs were related to members of *Caulimoviridae* (that were not Badnavirus sequences) and 58% of these were related to *Badnavirus.* Twenty nine % of the virus contigs were highly similar ($e < 1 \times 10^{-30}$) to *Canna yellow mottle virus* isolate V17 ORF3 gene (EF189148.1), *Sugarcane bacilliform virus* isolate Batavia D, complete genome (FJ439817.1) and *Banana streak CA virus* complete genome (KJ013511). Within this population were long contigs that resembled two full length genomes.

-Figure 1. A: it would be good to specify the sizes in the ladder on the picture; C: this panel is missing! C is actually what is described as D in the legend (this schemes need more explanation on what are the things designated on them). -L322: Blast of what?
-L324: This is referring to which sample?
-L352: Further comment on this would be beneficial: were there obstacles to whole genome assembly and were you able to get the complete circular genome sequence?
-L356: The use of MegaBLAST should be mentioned in protocol. Done
-L363-374: This seems to fit more in the results part. Changed to note
-L376-384: See the comment about BLAST results of contigs. Fixed
Figure 1 is modified to add more panels. Figure legend indicates ladder and annotated the sizes of 2 relevant ladder bands in each panel. Line 261 explains that assembled contigs were blasted. I also checked the use of word “sample” and corrected where it was overused.
Minor Concerns:
-In the abstract, the same information is repeated couple of times; it could be edited to become more concise. Edited and made concise
-There is a lot of switching of verb tenses, several grammatical errors and misused scientific formulations, which make manuscript a bit harder to read, it would be beneficial if you read the manuscript carefully through to eliminate such errors, few examples are below:
-L74: How can contigs be developed? Changed to assembled following NGS
-L235, 267 and elsewhere: you refer to the protocol for partial virion purification and sequencing simply as "NGS" and to sequencing of rRNA depleted totRNA as "RNA-seq"; both are sequenced using NGS (next generation sequencing approaches), the formulation should be changed to be correct and consistent throughout the paper. Changed throughout text
-L281: which mean? Average, median…? Average

All grammar was reviewed and corrected
Reviewer #3:
Manuscript Summary:
The authors describe a pipeline for quality de novo assembly of DNA virus genomic sequences that are low in titer

The review of the manuscript by Verchot et al., is mainly editorial in nature. The authors should take into consideration the following comments when revising the manuscript.

Major Comments:

1-Line 3: Delete "particle" as it is not needed here and other parts of the manuscript; Insert a dash between next and generation to read next-generation Done

2-Line 50: What is meant by the statement "may be difficult to propagate in the laboratory"? Re-write to be scientifically accurate as the manuscript deals with viruses, not bacteria, fungi, etc. eliminated the phrase “in another host”

3-Line 53: Replace "pioneered" with described Done

4-Line 63: Replace "concentration" with high titer. Please consult the chemical definition of concentration. Done

5-Line 64: What is meant by laboratory propagation for a virus? changed

6-Line 92: Revise "lesions, or necrosis" to lesions, necrosis or other symptoms. changed

7-Line 95: Revise "insect vector" to insect or other vectors changed

8-Line 139: Revise "carried out next generation sequencing (NGS) and RNA-seq" to: carried out NGS which also included RNA-seq changed

9-Line 235: "derived from NGS and RNA-seq" Do you mean: derived from NGS of DNA and RNA-seq? changed

10-Line 275: Insert "tentatively" before "named" to read tentatively named as the International Committee of Taxonomy of Viruses (ICTV) has to approve the name of a new virus species before it becomes official changed

11-Line 276 When a virus species is discussed taxonomically then it is written in italic, otherwise, it is not in italic according to ICTV. Please see: King, A.M.K, Adams, M.J., Carstens, E.B., Lefkowitz E.J., 2012. Virus Taxonomy: Ninth Report of the International Committee of Taxonomy of Viruses. Elsevier/Academic Press, London, UK. changed

12-Line 311: Figure 1C is missing! Please provide it or its figure legend should be deleted. In the latter case, figure legends D and E become C and D changed

13-Line 377: Change organism to non-viral changed

14-Line 490: Reference 29 is not complete, it should be completed changed

15-Figure 1: C should be D and D should be E. C is missing changed

16-Materials Table: Provide the location of the suppliers. The location of each supplier should be written only once when it is provided for the first time in the text. Done

Minor Comments:

1- Line 13: Provide email address and phone number Done

2-Line 32: Delete title (Director------) Done

3- Line 38: Delete Ph.D Done

4- Line 41: Insert a dash between next and generation; Insert (NGS) after next-generation sequencing Done

5-Line 46: Insert a dash between next and generation Done

6-Line 62: Replace "test" with host Done

7-Line 66: Replace "concentration" with titer Done

8-Line73: Delete next generation sequencing and leave (NGS) Done

9-Line 149: Replace the outer parenthesis ( ) with [ ] as two ( ) were used for (Ph.7.2) and (w/v). Delete the extra (in ((w/v) Done

10-Line 195: Delete "to" Done

11-Line 203: Change "Next generation sequencing" to NGS Done

12-Line 241: Location of Promega needs to be mentioned, it is located in Madison, WI All product

references removed
13-Line 344: Delete "been" after may be Done
14-Lines 354 and 360: Place a dash between next and generation Done
15: Figure 1 at the bottom of the page should be corrected to Figure 2 Done

Reviewer #4:
Manuscript Summary:
This paper presents an interesting approach to detect and sequence new viral sequences representative of replicating viruses from plant samples that could be useful for analyzing the risk of new diseases emergence. However, some majors flaws should be corrected and the paper should be reorganized.

Major Concerns:
1- The title did not reflect the combined approach of NGS on mRNA and DNA from plant samples to discover new viral sequences corresponding to replicating viruses. In fact, this paper described the research of badnaviral sequences in the sample and not completely new viruses.
New title:
Combining analysis of DNA in a crude virion extraction with analysis RNA from infected plants to discover new virus genomes.

2- The manuscript includes three different abstract, one on the first page (230 words), a short one (instead of the summary) and a long one (517 words) and they do not reflect the approach used in the paper. A lot of details from the protocol should not appear in the long abstract. The two sentences in the first abstract beginning by "an often preferred" and "However, this latter" correspond to introduction or discussion but not to the content of an abstract.

There are 2 abstracts, not 3 as required. These were modified as requested by reviewers 1-3. I changed the first sentences of the 2 abstracts to metagenome approach to identify plant viruses with double-strand DNA genomes.

3- The legend of figure 1 does not correspond entirely to the Figure 1. Figure 1E is missing that correspond probably to the legend described for 1C. All fixed

4-lanes 363-374 of the discussion does correspond to the results and not to the discussion Fixed

Minor Concerns:
In the protocol, paragraph 2, the dsDNA library should be mentioned as cDNA library and linked clearly to the RNA-seq procedure All fixed
In the protocol, paragraph3-2 (lane 212), it is explained that the Oklahoma State University Bioinformatics and Genomics Core Facility carries out the library preparation but paragraph 2-5, it seems that the library preparation is done with SeqCap RNA enrichment system. What has been done exactly?
Lane 229. Did the authors use MUSCLE or CLUSTALW? In the discussion, they spoke about CLUSTALW.
References 1, 7, 8, 11, 16, 17, 22, 27, 29 are not in a good format, "et al" should be detailed
All fixed

Jeanmarie Verchot

From: Permissions Helpdesk <permissionshelpdesk@elsevier.com>
Sent: Wednesday, February 28, 2018 9:34 AM
To: Jeanmarie Verchot
Subject: RE: request permission to reproduce a figure panel in JoVE

Dear Jeanmarie,

Elsevier journal authors retain the right to use or re-use portions or excerpts in other works, subject to proper acknowledgement; see https://www.elsevier.com/about/our-business/policies/copyright/personal-use. As this is a retained right, no written permission is necessary provided that the material is not credited to any third party and that the original source is properly acknowledged.

If I may be of further assistance, please let me know.

Best Wishes,
Laura

Laura Stingelin
Permissions Helpdesk Associate
ELSEVIER | Global E-Operations Books
+1 215-239-3867 office
l.stingelin@elsevier.com
Contact the Permissions Helpdesk
+1 800-523-4069 x3808 | permissionshelpdesk@elsevier.com

From: Jeanmarie Verchot [mailto:jm.verchot@ag.tamu.edu]
Sent: Tuesday, February 27, 2018 5:14 PM
To: Permissions Helpdesk <permissionshelpdesk@elsevier.com>
Subject: request permission to reproduce a figure panel in JoVE

***** External email: use caution *****

Dear Colleagues at Elsevier

I wrote a paper published in Virus Research vol 243 (2018) 19-24 entitled :Molecular characterization of two badnavirus genomes associated with Canna yellow mottle disease". By Wijayasekara et al.

In Figure 1 panel A is a diagrammatic representation of 2 virus genomes and I would like to reproduce this panel in another paper that I am submitting to JoVE this month. I would like to receive permission to reuse this portion of the figure. Please can you inform me of the process to gain permission?

Thank you

Jeanmarie Verchot
Texas A&M Agrilife Research
Center Director Dallas
17360 Coit Rd
Dallas TX 75252
(972) 952-9224