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 X 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