Cowpea Mosaic Virus (CPMV) Sequence Analysis over the Course of an Infection in Cowpea Plant

Nicole R. Jones
Point Loma Nazarene University
Mentor: Dr. Kris Koudelka
Committee: Dr. Kris Koudelka, Dr. Ryan Botts, Dr. April Cordero, Dr. Lori Carter

Abstract:

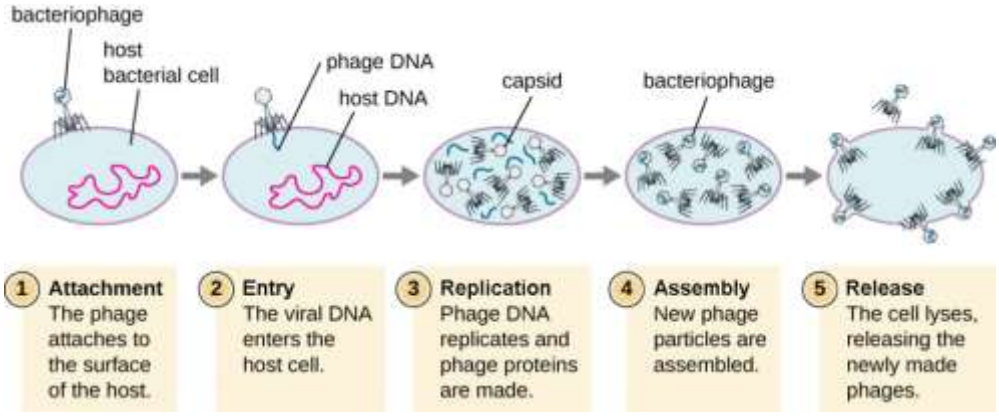
Viruses remain a highly studied particle in the biochemical field. In a time where SARS-CoV-2, the virus that causes COVID-19, plagues the globe, the understanding of viral genome evolution is vital. Here, cowpea mosaic virus is utilized to analyze the diversity of viral RNA over the course of an infection of cowpea plant. Extraction and sequencing of cowpea mosaic virus RNA was accomplished using a Qiagen Viral RNA Mini preparation as well as direct RNA sequencing using minION from Oxford Nanopore Technologies. Two new programs known as AssociVar and NanoR were utilized to analyze the computational output of minION in an attempt to identify the mutations in three samples from different times of infection. The entirety of the conclusion is not yet complete; however, the implementation of AssociVar and NanoR indicate that longer run times need to be completed on cowpea mosaic virus in order to have sufficient output data for further computational analysis.

Introduction:

In a time where viruses such as COVID-19 have plagued the normality of day to day life, the mutation and evolution of viruses seems as relevant as ever. It is well known that viruses evolve over time, resulting in new mutations of the same virus. It is for this reason that the common cold or flu is seen every year; a new year brings new strains. Part of this viral evolution is genetic drift, which can involve the selection for or against a certain allele. As a result, viruses that are closely related to the original virus are produced. While this is widely accepted, there is less research on the topic of viral diversity within an organism during the course of an infection. This exact concept is the very focus of these experiments.

A virus is a nonliving particle made up of nucleic acid and protein that is unable to replicate full infectious virions by themselves (7). As a result, a virus must invade a living cell and take over its metabolic machinery in order to replicate in a process called the viral reproductive cycle (7). As seen below in **Figure 1**, a lytic bacteriophage begins by attaching to the surface of a host cell, inserting its viral DNA into the host cell, and seizing control of the cell's method of replication in order to replicate its own genome (1,7). After replicating its genome, the virus can utilize the protein production mechanics of the host cell to produce more virus proteins (1). Once the host cell is full of replicated virus, the cell bursts open and releases the newly made viruses so that each can find a new host and repeat the process (1,7).

Figure 1: The Viral Life Cycle (1)

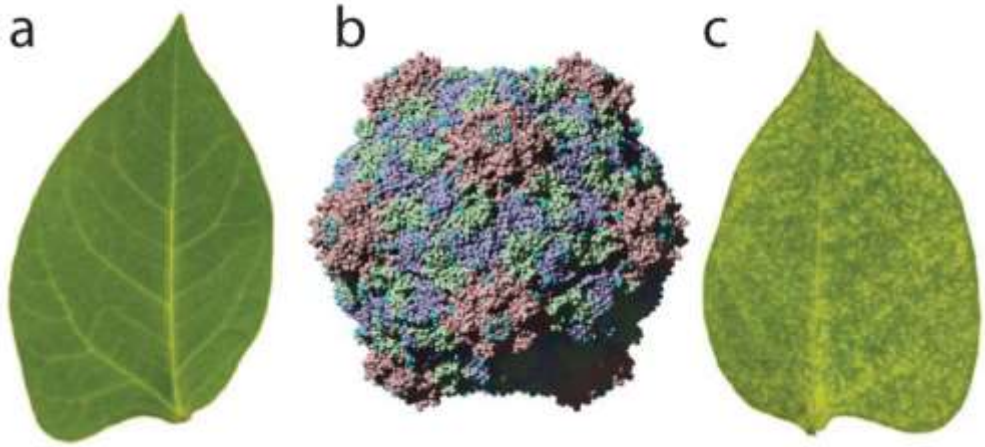


During RNA replication in a host cell, viral genomes undergo mutation (8). While error in genome replication may sound lethal, it is in fact the very way viruses survive and evolve; processes have

evolved in viral RNA replication where their RNA is intentionally mutated and acted on by natural selection with the idea that variation within the population will result in higher survival rates of the virus (8). The evolutionary implications of this are that the virus that an organism is infected with is theoretically not the same virus that that organism infects others with. By studying viral diversity during the course of an infection to better understand the process of error prone RNA replication within viral genomes, new viral treatment methods could eventually be produced.

Cowpea Mosaic Virus (CPMV) is a plant virus that is used as a viral nanoparticle. This virus was utilized due to its familiar purification and extraction protocols as well as the fact that it is single-stranded RNA. Plant viruses can have a narrow or broad range of hosts (1). Cowpea mosaic virus has a narrow host range as it is specific only to the Cowpea plant and cannot infect other organisms (2). CPMV is a part of the picornavirus family (9,10) which includes virus strains that result in the common cold, polio, and hand foot and mouth disease (11). Because of its specificity to plants, lack of harm to animals, and relationship to multiple common viruses that affect large populations of people, Cowpea mosaic virus is the perfect model organism to observe and analyze viral diversity in a controlled and safe lab environment.

Figure 2: Cowpea Mosaic Virus a) Cowpea plant pre-infection, b) computer generated structure of CPMV, c) Cowpea plant post-infection displaying the "mosaic" pattern that the virus is most commonly known for. (2)



As seen in **Figure 2c** above, CPMV causes a speckled, mosaic-like pattern on the leaves of its infected plant host. **Figure 2b** provides an animation of the virus, illustrating the capsid as a soccer ball-like lattice structure that surrounds the internal RNA of the virus. The transmission of CPMV, typically transferred through beetle bites (12), can be mimicked by creating lesions in the leaf tissue of the plant and directly applying the virus. These lesions are admitted to the primary leaves of the CPMV plant in order to infect the plant and allow the virus to multiply. As the virus replicates, the infection spreads to the forming secondary leaves of the plant where it can be collected. It should also be noted that CPMV is single-stranded RNA, which is more unstable than double stranded RNA or DNA; therefore, this fact makes the work of CPMV even more impressive. Cowpea plants were grown up and infected via primary leaves, followed by the virus being collected via secondary leaves, extracted, purified, and sequenced.

In order to analyze a viral genome, methods of RNA sequencing are needed. In a typical lab setting, one method of DNA sequencing is Illumina-based sequencing technologies, which generate great

quantities of highly accurate reads of DNA (4, 5). However, Illumina is limited by short reads and does not have the ability to sequence RNA directly (3). Therefore, a different method must be used to sequence viral RNA. In this experiment, Oxford Nanopore Technologies were utilized to accomplish direct sequencing of the Cowpea mosaic virus genome. Nanopore sequencing technologies provide a method of direct RNA sequencing with long read lengths and high yields (6). However, as these technologies are new to the biological field, the Nanopore technologies have relatively high error rates (3). This makes the resulting statistical analysis difficult when trying to identify mutations in the viral genome. In a paper published in October of 2019, a group of scientists created and tested a statistical program called AssociVar that uses MinION sequencing results to accurately discern real mutations in a RNA sample when compared to a reference genome (3). This program was utilized in order to analyze the diversity of Cowpea mosaic virus.

An additional program created by Davide Bolognini entitled NanoR was utilized to analyze the data obtained from MinION sequencing. This program is an R based program that incorporates code for six analysis tests, three of which pertained to our data specifically. These programs included NanoPrepareG, NanoStatsG, and NanoTableG. NanoPrepareG organizes output data to then be input into NanoStatsG, which gives statistics on the output data, and NanoTableG, which outputs a metadata table (19).

The addition of carrier RNA is an important step in this RNA sequencing and purification process. Carrier RNA plays two major roles: 1) it enhances the collection of viral nucleic acids and 2) it reduces the chances of RNA degradation (13). RNA is known for being more unstable than DNA due to its single-stranded nature and 2' hydroxyl group (17), therefore, carrier RNA binds to the single-stranded RNA samples and resolves this problem through the creation of double stranded RNA. Because of the rapid degradation of RNA, it was asked how much of a difference RNA carrier would make on the minION sequencing reads. As a result, two control samples were run for the initial RNA sequencing, one with the addition of carrier RNA and one with no carrier RNA.

As a whole, this experiment works to test multiple pieces of the story. Here, we explore the protocols of RNA extraction and sequencing from CPMV to create a uniform extraction and sequencing protocol for the Koudelka Research Lab. We also look at the effects of carrier RNA on the output of minION. Lastly, we utilize the program AssociVar and NanoR to try and create a computational protocol for the accurate analysis of CPMV RNA to then analyze the diversity of viral RNA in Cowpea mosaic virus over the course of a 12 day infection.

Materials and Methods:

Plant Growth and Infection:

Approximately 120 Cowpea seeds were planted in 40 small pots and grown over the course of 4 weeks in order to collect the needed CPMV samples. The plants were grown for approximately 2 weeks under greenhouse lighting until the sprouting of the secondary leaves. Once the secondary leaves of the cowpea plants were sprouted, the cowpea plants were infected. To infect the leaves, powder carborundum was wrapped in cheesecloth and sprinkled onto the primary leaves of each Cowpea plant. The sprinkle of carborundum was rubbed into the surface of the leaves to create small lesions in the leaf tissue. Following this, a control of secondary CPMV-infected leaves from 7/11/16 stored at -80^0C were crushed and filtered through cheesecloth. This cheesecloth was rubbed onto each individual leaf post-carborundum to infect the Cowpea plants, which were then left to grow for an additional 2 weeks. A fraction of the leaves used

for infection were kept as a control. The primary and secondary leaves of 60 plants were collected at day 6 of the experiment and again at day 12 of the experiment to act as the experimental samples. The collected leaves were labeled and stored at -80^{0}C for 3 months until purification.

Preparation #1: Purification of Cowpea Mosaic Virus Infection Control Group

The infection control CPMV leaves were weighed (46.00 g) and the virus was extracted using the following protocol: The leaves were homogenized in 3 times the weight of pre-chilled (4^{0}C) 0.1 M potassium phosphate, pH 7.0, 0.3%-mercaptoethanol with a pestle and mortar (150 mL buffer, 450 µL mercaptoethanol). The homogenate was filtered through 2 layers of cheesecloth. The filtrate was centrifuged for 20 minutes at 9,500 rpm (SLA-3000). CHCl3 and n-butanol were added (1:1) in a volume equal to that of the volume of the supernatant and stirred at 500 rpms for 30 minutes at 4℃. The solution was then centrifuged at 5,700 rpm (SLA-3000) for 10 minutes. The aqueous phase was collected. NaCl was added to obtain a 0.2 M solution and PEG8000 was added to 8% to the aqueous phase (1.17 g NaCl / 100 mL, 8 g PED800 / 100 mL). The solution was stirred at 4℃ for 30 minutes at 300 rpm. The solution was then centrifuged at 9,000 rpm (SLA-3000) for 15 minutes. The pellet was resuspended in 0.1 M K phosphate, pH 7 (15 mL per 1 g of pellet) and centrifuged at 9,300 rpm (SA-600) for 15 minutes. 3 mL of 30% sucrose was layered in phosphate buffer under the supernatant with a pipette, and then the tube was filled to the top with phosphate buffer and centrifuged at 42,00 rpm (50.2Ti) for 3 hours. The resulting pellet was resuspended in 1 mL phosphate buffer and allowed to dissolve overnight at 4^{0}C. The following day acrylic sucrose gradient makers were used to create 10-30% sucrose gradients. The resuspended pellet was loaded into the gradients and spun at 28,000 rpm for 3 hours. A fluorescent light source was used to visualize and remove bands of the virus with a needle. The extracted virus bands were added to KPO4 buffer and spun at 42,000 rpm (50.2 Ti) for 3 hours. The pellet was resuspended in 1 mL KPO4 buffer and allowed to dissolve overnight at 4℃ on a shaker.

Preparation #2: Purification of Cowpea Mosaic Virus Sample Groups

The procedure described above was repeated on the four sample leaf groups. The weights and volumes of materials added are as followed:

Table 1: Experimental Leaves: Weight, Buffer, and Mercaptoethanol

Sample Leaves:	Mass of Leaves:	Phosphate Buffer:	Mercaptoethanol:
1º - Day 6	36.86 g	120 mL	360 µL
2º - Day 6	30.41 g	90 mL	270 µL
1º - Day 12	7.73 g	30 mL	9 µL
2º - Day 12	40.22 g	120 mL	360 µL

Table 2: Addition of NaCl and PEG8000

Sample:	Volume:	NaCL added:	PEG800 added:
1º - Day 6	150 mL	1.755 g	12 g
2º - Day 6	65 mL	0.7605 g	5.2 g

1º - Day 12	55 mL	0.6435 g	4.4 g
2º - Day 12	57 mL	0.6669 g	4.56 g

The concentration of each sample was measured on a Nanodrop2000 to confirm the presence of the virus. Each sample was stored at 4^0C until extraction.

Combination of Control Samples of CMPV:

The two original CMPV samples that were used to create the infection control sample having concentrations of 5.6 mg/mL and 1.22 mg/mL were combined to act as a control sample for the extraction and sequencing of viral RNA via minION. A Nanodrop2000 and the formula for absorption ($A = \varepsilon l c$) were utilized to evaluate the concentration of the combined sample. The extinction coefficient of CPMV is 8.1 mL/mg*cm and the output of the methods described above is as follows:

Table 3: Concentration of Combined CMPV Control Sample

Concentration:	3.846 mg/mL
A280:	17.405
260/280:	1.79
A260:	31.155

It was concluded that the sample may have crystalized, so the sample was spun at max speed for 10 minutes in an Eppendorf Centrifuge and the supernatant was saved and evaluated on the Nanodrop2000 with the following output:

Table 4: New Concentration of Combined CMPV Control Sample

Concentration:	2.983 mg/mL
A280:	14.164
260/280:	1.68
260:	23.796
Volume:	0.15 mL
Yield:	0.4407 mg

Extraction of CPMV in Control Sample:

Extraction of viral RNA from the CPMV sample was accomplished through the utilization of the *Qiagen Viral RNA Mini Protocol* (13) along with the *QIAamp Viral RNA Mini Kit*. One sample was prepared with carrier RNA (approximately 100 bp to 10 kb in length) and a second sample was prepared without carrier RNA to determine if the addition of carrier RNA was necessary in the protocol to achieve a readout on the minION. After extraction of the viral RNA, both samples were evaluated on the Nanodrop2000 with the following output:

Table 5: Yield Data of Sample with Carrier RNA

Concentration:	144.2 ng/µL
A260:	3.604
A280:	1.461
260/280:	2.47
260/230:	1.44
Volume:	~90 µL
Yield:	~12,978 ng

Table 6: Yield Data of Sample without Carrier RNA

Concentration:	34.2 ng/µL
A260:	0.855
A280:	0.362
260/280:	2.36
260/230:	0.20
Volume:	100 µL
Yield:	~3,420 ng

Sequence-Specific Direct RNA Sequencing - Nanopore Protocol:

The 2 RNA combined CPMV control samples were prepared and sequenced on minION following the Nanopore Protocol (14), the Sequence Specific Direct RNA Sequencing Kit, and the Flow Cell Priming Kit from Oxford Nanopore Technologies. According to the protocol, 500 ng of RNA was needed in order to sequence the samples properly, so concentrations were used to calculate the volume of sample to input into the flow cell. Cowpea mosaic virus with carrier was run on minION for approximately 24 minutes while Cowpea mosaic virus was run for approximately 20 minutes. The data was collected to be analyzed. The remainder of the extracted samples from each time stamp for both primary and secondary leaves were saved for further analysis later after the application of the protocol was perfected.

FASTQ to FASTQ File

The output of the minION sequencing was a fastq file. According to the literature that the computational analysis was based upon, this file needed to be converted to a fasta file. This was accomplished utilizing the command: sed '/^@/!d;s//>/;N' fastq > fasta. Both the fastq and fasta files were run on BLAST to identify a reference Cowpea mosaic virus genome. The fastq file had two matches while the fasta file had no matches on BLAST. In an attempt to try and run the program, the output from the fastq file was used.

BLAST and BLAST+:

BLAST is an online database search tool provided by the National Center for Biotechnology that allows the user to compare biological sequences and look for sequence similarity to reference genomes available on the database (15). BLAST was utilized to compare the output FASTQ file from the RNA sequence to a reference genome of CPMV. Two matches were found: Cowpea mosaic virus (CPMV) middle-component RNA (M RNA) and Cowpea mosaic virus RNA 1, complete sequence. From these, the complete sequence was chosen to move forward in order to run the AssociVar program.
The AssociVar program highly recommends utilizing BLAST+, a downloadable version of the BLAST database that can be utilized from the command line. In order to successfully run the program, BLAST+ was downloaded and researched to discover how to utilize the program.

AssociVar:

The AssociVar program and codes described in the literature were found, fully accessible, on an online database (16). The provided codes are set to be run in Python. The necessary codes were created from and saved in the terminal as .py files to eventually be run once the BLAST+ output file was successfully created and the required parameters of the program input were determined. However, application of the AssociVar proved difficult given our dataset output. After many tries to apply this code, a new code was sought out for a simpler analysis of minION output data.

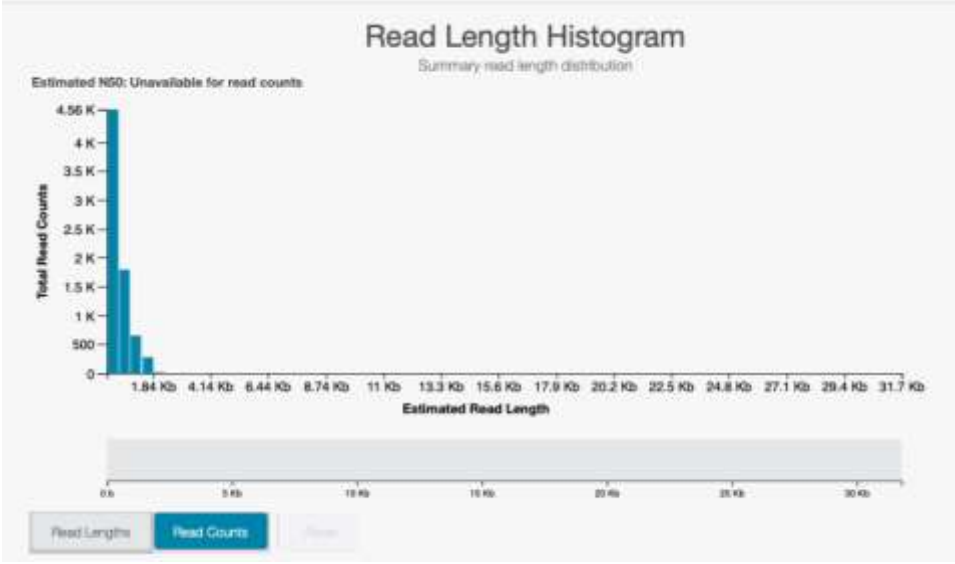
NanoR: R Package to Visualize ONT Data

A new program was researched to analyze the minION output data. NanoR incorporates code for six analysis tests, three of which pertained to our data specifically: NanoPrepareG, NanoStatsG, and NanoTableG. NanoPrepareG was created in R Studio and ran on both CPMV with carrier and CPMV without carrier. Both samples had an output summary with an empty summary table and a value of 0. Due to this output, NanoStatsG and NanoTable G could not be run.

Results:

Figure 3: a) Read length histogram of CPMV wildtype with carrier shows smaller variety but greater quantity of RNA read lengths on minION. **b)** Read Length Histogram of CPMV wildtype with no carrier shows larger size variety of RNA read lengths on minION.

a.



b.

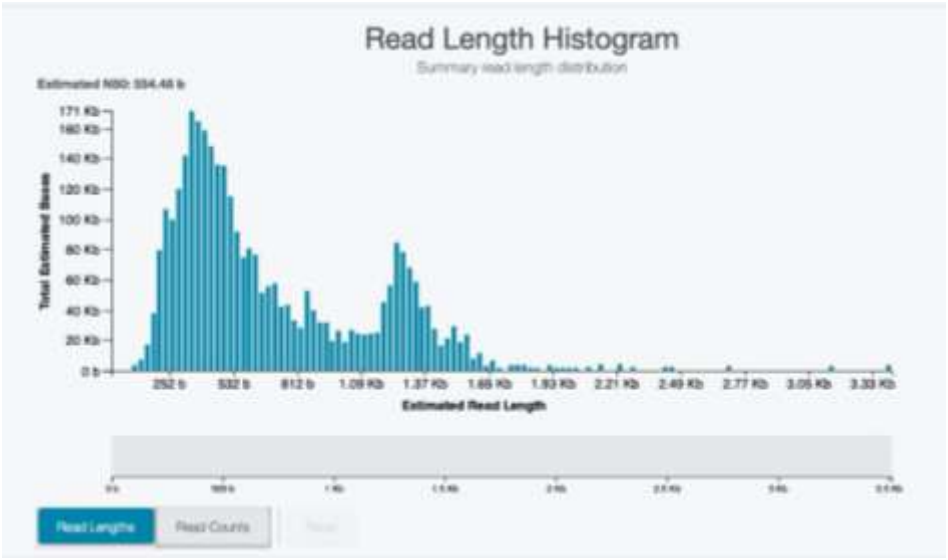
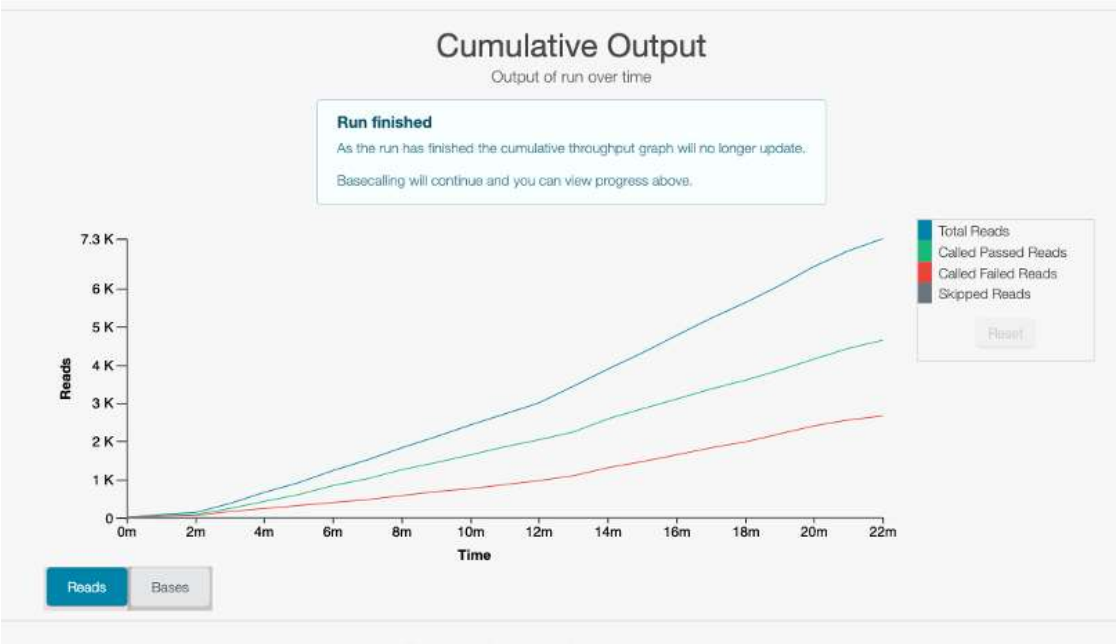


Figure 4: Cumulative output of reads is greater in CPMV with carrier than CPMV with no carrier. **a)** Cumulative output of reads of control sample with carrier shows total number of reads just under 7.3 K. **b)** Cumulative output of reads of control sample with no carrier shows total number of reads as approximately 6.24 K.

a.



b.

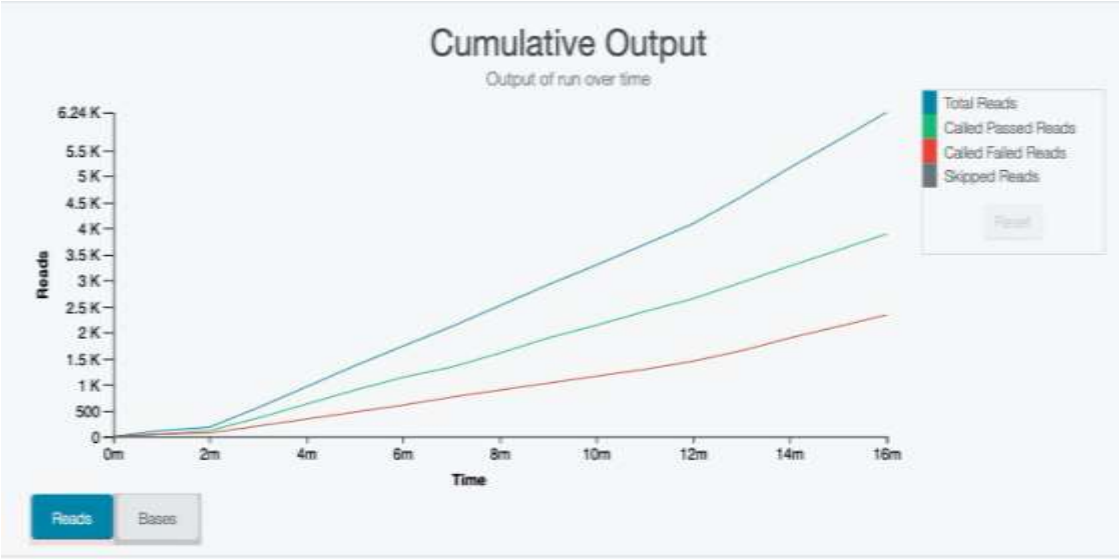


Table 7: Output of BLAST from online database shows two reference genome matches for both tested Cowpea mosaic virus with carrier and Cowpea mosaic virus without carrier.

a) Cowpea Mosaic Virus Control with Carrier

Sample match	Max (Bit) score	Total score	Query cover	E value	Percent Identity
CPMV middle-component RNA	1844	2.140e+05	9%	0	90.01%
CPMV RNA1	1748	3.172e+05	13%	0	92.06%

complete sequence					

b) Cowpea Mosaic Virus without Carrier

Sample match	Max (Bit) score	Total score	Query cover	E value	Percent Identity
CPMV middle-component RNA	1775	2.865e+05	12%	0	89.02%
CPMV RNA1 complete sequence	1616	2.9232e+05	13%	0	87.37%

Discussion:

As a whole, we have only scratched the surface of the diversity of RNA over the course of an infection. An extraction protocol producing clean readouts was successfully created as described above. The computational analysis work continues to be practiced and perfected with the hopes of success in the Koudelka Research Lab.

From the data obtained so far, some loose conclusions can be made. First, **Figure 3** displays the read lengths of both control samples of CPMV. It should be noted that the scales of the x and y axis in this figure differ for 3a and 3b. This is important to note in order to better understand what the figure is displaying. From these graphs provided by minION, it is obvious that the RNA sample with the addition of carrier RNA has longer read lengths than that of the RNA sample without carrier. It is known that CPMV RNA has a read length of 3.5 kb (18) and that the carrier in the Qiagen protocol ranges in length between 100 bp to 10 kb (13). This is important to note because RNA with carrier should appear near this length on the figures if degradation of the unstable nucleic acid structure was avoided. Acknowledging this, it can also be seen that the sample with carrier has a shorter read length spread or a higher frequency of similar read lengths than that of the sample without carrier. This distribution could very well be due to the RNA degradation discussed above. By including carrier RNA, degradation of the viral genome is minimized, resulting in unfragmented RNA. On the contrary, a lack of carrier would result in fragmented RNA, which is exactly what **Figure 3** displays. CPMV without carrier has a large variety of read lengths, indicating fragmented RNA due to degradation while CPMV with carrier has a high frequency of consistent read lengths, indicating little to no RNA degradation. As a result, it is highly beneficial and necessary to include the addition of carrier RNA in the protocol of CPMV sequencing.

Second, **Figure 4** displays the total number of reads obtained from each sequence run on minION. According to the graphs, it appears that RNA with carrier had approximately 7.3 K total reads while RNA without carrier had approximately 6.24 K total reads. However, it should be noted that the RNA with carrier was sequenced for 4 minutes longer than the RNA without carrier; therefore, the

number difference is most likely a result of the difference of sequencing length and cannot be attributed to the presence or lack of carrier RNA.

Lastly, **Table 7** displays the output of the online BLAST analysis on both the Cowpea mosaic virus with and without carrier. In this table, it can be seen that both samples matched with the same two reference genomes but at different rates. In BLAST, a good match has a high percent identity, large query cover, low E value, and large Bit score. These parameters were utilized when analyzing the BLAST output in **Table 7**. CPMV with carrier matched two reference genomes, with both matches having a percent identity above 90%. The control sample with carrier had a higher percent identity to the complete sequence reference genome. On the other hand, CPMV without carrier RNA had the same two matches, but with a percent identity of less than 90% and a higher match with the middle component RNA reference genome. In both samples, a low query cover is seen, with every match being between 9-13%. In both samples it is also seen that an E value of 0 was obtained. Lastly, Bit scores for both samples were large, with CPMV with carrier having larger Bit scores than that of CPMV without carrier. Overall, between the two samples, the control sample with carrier had better BLAST match results than that of the control sample without carrier; however, both samples have matches of enough quality to continue on in the computational sequence analysis. For the remainder of the sequence analysis, it was concluded that the complete sequence Cowpea mosaic virus reference genome will be utilized as the reference genome in AssociVar.

After success was not obtained through AssociVar, the NanoR R package was utilized to analyze the output data. As seen in methods and materials, the output of NanoPrepareG was an empty summary table with a value of 0. This output proved to be inconsistent with the results obtained via BLAST. It is believed that this output is due to insufficient run times. Given the circumstance of the pandemic, our run times were shortened and elongated run times were not able to be obtained before the completion of this project. As a whole, this entails that longer run times need to be obtained to successfully run the summary R packages provided in NanoR.

In conclusion, the RNA sequencing on minION appeared to be successful in the sense that clean reads in the real-time Nanopore program were obtained using carrier RNA. However, the output FASTQ files of the RNA sequencing is not adequate enough to continue on in the computational analysis of the Cowpea mosaic virus genome to identify the mutation of the genome over the course of infection. With this being said, adequate RNA can be successfully obtained with elongated run times on Nanopore's minION. Much work still needs to be completed to make further conclusions and to create a proper protocol for the Koudelka Research lab for further research, but the results so far appear to be a promising start.

References:

1. "The Viral Life Cycle." Microbiology. Lumen, https://courses.lumenlearning.com/microbiology/chapter/the-viral-life-cycle/, Accessed 1 May 2020.
2. Pouwels, Jeroen, Jan E. Carette, Jan Van Lent and Joan Wellink. *Cowpea mosaic virus: effects on host cell processes*. 31 October 2002, https://doi.org/10.1046/j.1364-3703.2002.00135.x.
3. Harel, Noam, Moran Meir, Uri Gophna, and Adi Stern. *Direct sequencing of RNA with MinION Nanopore detecting mutations based on associations*. Nucleic Acids Research, 2019, vol. 47, No. 22, 3 October 2019.
4. "What is the Illumina method of DNA sequencing?" yourgenome.org, 25 February 2015, https://www.yourgenome.org/facts/what-is-the-illumina-method-of-dna-sequencing.
5. "An Introduction to Next-Generation Sequence Technology." Illumina Inc, 2017, www.illumina.com/technology/next-generation-sequencing.html.
6. Oxford Nanopore Technologies. 2020, https://nanoporetech.com.
7. Brooker, Widmaier, Graham,and Stiling. "The Simpler Genetic System of Viruses and Bacteria." *Principles of Biology, Second Addition*. McGraw-Hill Education, 2017, p. 337-344.
8. Everett Clinton Smith. *The not-so-infinite malleability of RNA viruses: Viral and cellular determinants of RNA virus mutation rates*. PLOS Pathogens, 27 April 2017, https://doi.org/10.1371/journal.ppat.1006254.
9. Carette, Jan E., Marchel Stuiver, Jan Van Lent, Joan Wellink, Ab Van Kammen. *Cowpea Mosaic Virus Infection Induces a Massive Proliferation of Endoplasmic Reticulum but Not Golgi Membranes and Is Dependent on De Novo Membrane Synthesis*. Journal of Virology, American Society of Microbiology, July 2000, https://jvi.asm.org/content/74/14/6556.
10. Koudelka, Kristopher J., Giuseppe Destito, Emily M. Plummer, Sunia A. Trauger, Gary Siuzdak, and Marianne Manchester. *Endothelial Targeting of Cowpea Mosaic Virus (CPMV) via Surface Vimentin*. PLOS Pathogens, 5 May 2009, 10.1371/journal.ppat.1000417.
11. Tuthill T.J., Groppelli E., Hogle J.M., Rowlands D.J. (2010) Picornaviruses. In: Johnson J. (eds) Cell Entry by Non-Enveloped Viruses. Current Topics in Microbiology and Immunology, vol 343. Springer, Berlin, Heidelberg, Curr Top Microbiol Immunol. 2010; 343: 43–89.
12. A. van Kammen, J. van Lent, and J. Wellink. *Cowpea mosaic virus*. AAB: Association of Applied Plant Biologists, April 2001, http://www.dpvweb.net/dpv/showdpv.php?dpvno=378.
13. *QIAamp Viral RNA Mini Handbook*. QIAGEN, March 2018.
14. *Nanopore Protocol: Sequence-Specific Direct RNA Sequencing (SQK-RNA002*). Version: DSS_9081_v2_revH_14Aug2019, Oxford Nanopore Technologies, 14 August 2019.
15. *Using BLAST*. BIOSEQ: Bioinformatics Activity, https://ase.tufts.edu/chemistry/walt/sepa/Activities/BLASTpractice.pdf, accessed on 4 May 2020.
16. AssociVar. GitHub, 11 Aug 2019, https://github.com/SternLabTAU/AssociVar/tree/master/association_tests.
17. Rana, Ajay K. , and Serge Ankri. *Reviving the RNA World: An Insight into the Appearance of RNA Methyltransferases*. Frontiers in Genetics, 6 June 2016, 10.3389/fgene.2016.00099.
18. Hesketh, E.L., Meshcheriakova, Y., Thompson, R.F. *et al.* The structures of a naturally empty cowpea mosaic virus particle and its genome-containing counterpart by cryo-electron microscopy. *Sci Rep* 7, 539 (2017). https://doi.org/10.1038/s41598-017-00533-w

19. Davide Bolognini, Niccolò Bartalucci, Alessandra Mingrino, Alessandro Maria Vannucchi, Alberto Magi. NanoR: A user-friendly R package to analyze and compare nanopore sequencing data. PLoS One. 2019 May 9.