1. Cover Page

INTERIM Report: OGWRI

Project Title: **Development of NGS based multi-plex assay for non-regulated grapevine virus**

Applicant Name: **Dr. Sudarsana Poojari**

Report Number: **Interim #6**

Reporting Period: 1 **November 2021 and 31 March 2022**

Date of Submission: 10-June-2022

Program Coordinator: **Eleanor Hawthorn**

2. Executive Summary

Viral diseases in cultivated grapevines have been considered a limiting factor for the production of quality wines as well as to the lifespan of the vineyard. Detection is the first step in the management of viral diseases. Although, the traditional polymerase chain reaction (PCR) based detection methods are available, they are virus specific and less sensitive when compared to the next-generation sequencing (NGS). The availability of NGS platforms provides an opportunity for a broad-spectrum detection of viruses with less time and more sensitivity in comparison to tradition PCR or serological techniques. Here we started to develop an NGS based multiplex assay (NGS-MA) for the detection of non-regulated grapevine viruses in Canada.

List the project objectives:

1. Develop Next Generation Sequencing-based multiplex assay (NGS-MA) to identify non-regulated grapevine viruses.
2. Standardize and validate the NGS-MA and compare the sensitivity and accuracy of NGS-MA with domestic quarantine standards and facilitate the rapid adaption of NGS-MA in characterization and detection of non-regulated grapevine viruses
3. Determine the cost comparison analysis of the NGS-MA with high-end NGS technologies as well as traditional and quantitative PCR methods
4. Provide the science-based information and expertise for the need of promoting virus-free propagating material through the use of NGS-MA methods.
5. Contribute to maintaining the highest quarantine standards to support the domestic certification agencies and regulatory bodies.

Objectives or goals accomplished - Overall

1. Double standard RNA (dsRNA) protocols and total nucleic acids (TNA) isolation protocols were standardized from and successfully isolated from 50+ grapevines samples
2. Custom probes were designed for the panel of 32 different virus species (non-regulated) and six house-keeping genes (Vitis vinifera) based on the 800+ full-length genome sequences available in the public genomic database.
3. This project is in collaboration with Fed-Dev Ontario which provided the funding to purchase two units of high throughput sequencing equipment from Illumina through Brock's Validation and Prototype Manufacturing Institute (VPMI).
4. $20,000 cash contributions from Fed-Dev leveraged for $35,000 funding from OGWRI for the years 2019-2020. Among, the $35,000 from OGWRI, computer accessories were purchased.
5. Using our standardized TNA isolation protocol, we have successfully performed NGS on 50+ grapevine samples on Illumina's MiSeq platform using 150bp True-seq paired-end method
6. NGS data analysis (De novo assembly, trimming, aligning, quality control, taxonomic classification, and reference-based analyses) were conducted using Virtool (https://www.virtool.ca/) and CLC genomics workbench from Qiagen

Issues that have affected the success of the project to date: Due to limited laboratory access restrictions due to the COVID-19 pandemic from 2019 to 2021 the progress was slower with limited access to the laboratory and data analysis software.

3. Detailed description of the Project

a) Outputs Nov 2021-October 2022:

1. 72 pairs of probes specific to grapevine viruses and six housekeeping genes were synthesized at Illumina and are currently being used in the NGS runs on the MiSeq platform.
2. Due to changes in the Miseq running protocols from Illumina, we re-structured the protocol for NGS runs
3. One major change is the use of dsRNA as starting material which resulted in the highest number of quality sequence reads and increased sensitivity of virus detection in grapevine samples
4. Sequence data 50+ samples using both Virtool and Genomics workbench software indicated the presence of 12 different grapevine viruses and four viroid species were identified.
5. Table 1 represents the viruses and target regions of custom probes included in the target enrichment panel used for the NGS multiplexing assay

RNA Viruses			
Virus Name	**Region Selected**	**Nucleotide Position**	**Number of Reference Sequences**
Arabis mosaic virus (RNA 1 and 2)	CP	2485-2598	20
	RDRP	5171-5282	20
Grapevine fleck virus	CP	6367-6467	4
	REP	781-918	4
Grapevine fanleaf virus (RNA 1)	RDRP	5030-5138	90
	VPg	4098-4175	90
Grapevine fanleaf virus (RNA 2)	CP	2788-2902	146
	MP	1230-1349	146
Grapevine leafroll-associated virus 2	HSP70h	10639-10809	25
	CP	15200-15305	25
Grapevine leafroll-associated virus 3	HSP70h	11354-11474	67
	CP	13645-13773	67
Grapevine leafroll-associated virus 4	HSP70h	9606-9713	11
	CP	12861-12994	11
Grapevine leafroll-associated virus 5	HSP70h	8549-8673	6
	CP	11750-11865	6
Grapevine leafroll-associated virus 6	HSP70h	9579-9698	2
	CP	12300-12409	2
Grapevine leafroll-associated virus 7	HSP70h	9050-9160	4
	CP	12330-12431	4
Grapevine leafroll-associated virus 9	HSP70h	8050-8198	1

	CP	11175-11296	1
Grapevine Pinot gris virus	CP	6865-6975	50
	MP	5640-5744	50
Grapevine Rupestris stem pitting associated virus	CP	8321-8428	124
	RDRP	3524-3362	124
Grapevine Syrah virus 1	CP	5958-6068	9
	MP	950-1056	9
Grapevine virus A	CP	6642-6748	21
	MP	5790-5919	21
Grapevine virus B	CP	6659-6765	10
	MP	6301-6405	10
Grapevine virus D	CP	200-319	1
	RNA binding Protein	519-637	1
Grapevine virus E	CP	6754-6868	13
	MP	6130-6250	13
Grapevine virus F	CP	6762-6892	5
	MP	6440-6553	5
Grapevine virus G	CP	6726-6838	9
	MP	5774-5880	9
Grapevine virus H	CP	6525-6643	3
	MP	5835-5953	3
Grapevine virus I	CP	6599-6733	2
	MP	5730-5848	2
Grapevine virus J	CP	6525-6643	2
	MP	5700-5833	2
Grapevine virus K	CP	6780-6898	3
	MP	5999-6118	3
Grapevine virus L	CP	6896-7000	4
	MP	6057-6192	4
Grapevine virus M	CP	6615-6733	1
	MP	5820-5938	1
	DNA Viruses		
Virus Name	**Region Selected**	**Nucleotide Position**	**Number of Reference Sequences**
Grapevine geminivirus A	CP	68-182	23
	Pre-CP-V2	415-521	23
Grapevine badnavirus 1	ORF2	1010-1151	1
	ORF3	5850-6000	1
Grapevine roditis leaf discoloration-associated virus	ABY30 gp3	3400-3538	4
	ABY40 gp4	6179-6298	4

Grapevine vein clearing virus	ORF3	5029-5148	12
	ORF2	1175-1282	12
Grapevine red blotch virus	CP	984-1092	131
	REP	2680-2798	131

b) Methodology

1. Primers/probes specific to grapevine viruses and their variants were designed based on the consensus genome sequence alignments of available sequences from the public database. Custom probes were designed for the panel of 32 different virus species (non-regulated) based on the 800+ full-length genome sequences.
2. The efficiency of the probe panel is being evaluated on both MiSeq and iSeq100 flatforms.
3. Leaf and petiole samples collected from grapevines from the phytotron facilities as well as the sample collected from commercial vineyards were used for TNA and dsRNA extractions.
4. Total nucleic acids were isolated using a spectrum total RNA isolation kit (Sigma) with modifications and the quality and quantities of TNAs were measured using both nanodrop and qubit (Thermo Fisher Scientific) instruments.
5. Samples were probed singly and multiplexed for sequencing using the Illumina MiSeq platform via v3 chemistry with 2X150 bp read lengths
6. NGS runs were performed AmpliSeq library preparation and then analyzed by TrueSeq library preparation on the Illumina MiSeq platform.
7. De novo assembly, trimming, aligning, quality control, taxonomic classification, and reference-based analyses were conducted using Virtool (Rott, et al., 2017) on the Paired-end analysis setting with the host genome of Vitis Vinifera subtracted from the data for analysis. Virtool (https://www.virtool.ca/) was built upon various Linux-based bioinformatics pipelines and alignment programs that ran in the background and assisted in the analysis. These programs included Skewer, FastQC, Bowtie2, SPAdes, HMMER, FLASH, and AODP. Following the identification of specific Viruses using Virtool, CLC Genomics Workbench v11 (QIAGEN, Redwood City, CA) was utilized to scan the sequences obtained to pinpoint and extract the exact region of overlap between the viruses and original sequenced data. Identifying the sequences of the viruses provided more contextualized results for the virus sequences obtained. Further analysis of the virus sequences was carried out using BLAST to investigate data that was represented in each dataset.

4. Results to Date (short-term)

1. Ampliseq NGS-MA assay was developed and successfully applied to more than 70 grapevine samples
2. Ampliseq NGS-MA assay was modified to suit dsRNA as the source of starting genetic material which resulted in increased sensitivity.
3. Results of a sub-set of 11 samples were shown below to indicate the sensitivity and comprehensiveness of the PCR-based detection and NGS assay.
4. 10 of the 11 samples sequenced using high throughput sequencing had at least one target virus detected using endpoint PCR amplification as reported previously in ***Table 2***. 11 samples were processed using the standard Truseq library preparation protocol reported previously in the methods section. 6 of the 12 viruses previously detected were sequenced, which included 2 of 3

samples with previously detected GLRaV-1 virus, 1 of 2 samples with previously detected GLRaV-2 virus, 2 of 3 samples with previously detected GLRaV-3 virus, and 1 of 4 samples previously detected GRBV virus. Previously undetected pathogens using serological PCR methods were sequenced, including 1 sample with previously undetected GLRaV-1 virus, and 6 samples with previously undetected GPGV virus. ***Table 3*** below reports the results obtained from the analysis using high throughput sequencing. The results reported in ***Table 3*** include a report of the coverage and virus titer to allow for assessment of a feasible positive call for each virus.

Table 2: Virus diagnostic results for grapevine viruses in plant samples obtained utilizing endpoint PCR.

Sample	Sample Type	Variety	GRBV	GLRaV-1	GLRaV-2	GLRaV-3	GPGV
119A	Petiole	Cab Franc	**-ve**	**+ve**	**-ve**	**-ve**	**-ve**
119B	Petiole	Cab Franc	**-ve**	**+ve**	**-ve**	**-ve**	**-ve**
122C	Petiole	Cab Franc	**+ve**	**-ve**	**-ve**	**-ve**	**-ve**
1251A	Cane	Reisling	**-ve**	**-ve**	**-ve**	**+ve**	**-ve**
1683A	Cane	Semilion	**+ve**	**-ve**	**-ve**	**-ve**	**-ve**
1683B	Cane	Semilion	**+ve**	**-ve**	**-ve**	**-ve**	**-ve**
1764	Cane	Gamay	**+ve**	**-ve**	**-ve**	**-ve**	**-ve**
1879	Cane	Chardonnay	**-ve**	**+ve**	**-ve**	**+ve**	**-ve**
1953	Cane	Vidal	**-ve**	**-ve**	**+ve**	**+ve**	**-ve**
2027	Cane	Cab Franc	**-ve**	**-ve**	**+ve**	**-ve**	**-ve**
2024	Cane	Cab Franc	**-ve**	**-ve**	**-ve**	**-ve**	**-ve**
	Percent Positivity		**36.3%**	**27.3%**	**18.1%**	**27.3%**	**0%**

Table 3: Virus diagnostic results for grapevine viruses in plant samples obtained utilizing traditional high throughput TruSeq[a] sequencing.

Sample Tested	Virus Detected	Genome Length (nt)	Depth[b] of sequence	% sequence coverage	Virus weight[c] titer	No. of virus specific reads	% of total detected reads
119A	GLRaV-1	18863	6	87.9	0.200	212	0.56
	GLRaV-1	18863	3	75.3	0.482	253	48.3
	GPGV	7223	<1	2.1%	0.000256	<1	<1
119B							
	GPGV	7250	37	99.6	0.936	1369	93.6
	GRBV	3206	2	69.6	0.0000217	<1	<1
122C							
	GPGV	7250	<1	38.9	0.233	82	23.4
1251A	GLRaV-1	18731	3	72.1	0.163	57	16.2
1764	GPGV	7224	43	100	0.987	1383	98.7
	GPGV	7224	29	100	0.984	977	98.4
	GLRaV-3	18498	<1	2.3	0.00745	7	0.704
1879							
	GPGV	7250	18	99.5	0.255	400	25.5
	GLRaV-3	18498	6	94.2	0.702	1102	70.2
1953							
2027	GLRaV-2	16486	2	69.9	0.901	423	20.9
2104	GPGV	7250	<1	2	2.29×10^{-9}	<1	<1

[a] An Illumina library preparation protocol used to generate the high throughput sequencing data
[b] refers to a measure of how many times a genome is covered by mapped reads.
[c] refers to the calculated proportion of reads mapping to a virus. The weight is roughly proportional to the titre. Higher the titre, the higher the weight. A weight greater than 0.001 is a strong indicator of a positive detection. (https://www.virtool.ca/)

5. Reach and Communications (actual vs. expected)

1. Abdallah Meknas, a 4th year BTech student at Brock University started his research thesis work on this project in September 2020 and completed the research thesis and in April 2021.
2. Publications: Abdallah Meknas, Tony Wang, and Sudarsana Poojari. 2021. *One Test for All: Detection of Grapevine Viruses using High Throughput Sequencing*. Ontario Fruit and Vegetable Convention 2021. **2nd Place**. https://www.ofvc.ca/posters.html

6. Conclusion and Next Steps

3. The standardized NHS-MA would be used for comparative studies using funds from the Genome Canada project and determine which methods are the best suited for screening viruses in grapevine material in the clean plant program pipeline.
4. 2021-2022: Cost comparisons will be carried out between the MiSeq and Iseq100 flatforms using NGS-MA assays.
5. 2021-2022: NGS-MA will be applied to the field samples for compatibility and sensitivity studies.