Analysis of the genetic diversity of Grapevine rupestris stem pitting-associated virus in Ontarian vineyards and construction of a full-length infectious clone

by

Julia C. Hooker

A thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Master of Science
in
Molecular and Cellular Biology

©Julia Hooker, April, 2017
ABSTRACT

Analysis of the genetic diversity of Grapevine rupestris stem pitting-associated virus in Ontarian vineyards and construction of a full-length infectious clone

Julia Catherine Hooker
University of Guelph, 2017

Advisor:
Dr. Baozhong Meng

Advisory Committee:
Dr. Peter Krell
Dr. Annette Nassuth

Grapevine rupestris stem pitting-associated virus (GRSPaV; Betaflexiviridae, Foveavirus) has been associated with a number of diseases including Syrah decline. Previous findings show GRSPaV sequence variants cluster into three or more main phylogroups, regardless of geographical location. Here, the genetic diversity of GRSPaV isolates from Ontarian vineyards was analyzed using broad-spectrum primers targeting the viral polymerase coding sequence. It was hypothesized that GRSPaV variants in Ontario are diverse and GRSPaV-SY variants are involved in Syrah decline symptoms. In total, 169 cDNA clones from 21 Vitis sources were used for phylogenetic analysis. Similarly to previous reports, four major lineages were observed; GRSPaV-PN, -SG1, -SY, and –GG. Variants of the GRSPaV–SY lineage were confirmed in all 14 sources tested with SY-specific primers. Syrah cultivars expressing red canopy decline symptoms had more clones clustering with the GRSPaV-SY lineage than those without observable decline symptoms. GRSPaV clones from 8 hybrid sources mostly clustered with GRSPaV-SY, followed distantly by GRSPaV-BS. A full-length infectious cDNA clone of a GRSPaV-SY-related variant is being constructed to further investigate this relationship.
ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Baozhong Meng, for providing me with the opportunity to delve into the world of grapevine virology and for helping me pursue my passions in molecular virology. I would like to thank my committee members, Dr. Peter Krell and Dr. Annette Nassuth, for their knowledge and guidance throughout my research. I would also like to thank NSERC for funding through the Discovery grant and the NSERC Engage program.

A big thank you to all my lab members, past and present. Thank you for years of laughter, dancing, and most importantly for going through the experience of this learning process with me. A special thank you to Dr. Huogen Xiao for all his time, patience, and lab lessons, in addition to assistance in collecting grapevine source material and RNA extractions. Thank you to Sunny Li for all the years of guidance and encouragement, and for help with troubleshooting for the full-length clone and sequencing work. Thank you to Mehdi Shabanian for his help in source material collection and providing cDNA from hybrid sources.

A very special thank you to Deb Flett and Paula Russell for their guidance through each TA-ship and thesis writing. I truly do not think I could have made it without you two! Thank you to Erin and Diana at Student Counselling Services for their personal help.

I would like to say a special thank you to my loving parents, siblings, friends, and cats for their continued love, encouragement, and support. A special shout out to my partner Nick for being my number 1 fan and for being such an important emotional support through all my research woes and breakthroughs.
# TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................. viii
LIST OF FIGURES ................................................................................................................ ix
LIST OF ABBREVIATIONS ................................................................................................. xi
LIST OF VIRUS ABBREVIATIONS ..................................................................................... xii

CHAPTER 1: INTRODUCTION ............................................................................................... 1

1.1 Grapevine history: Cultivation and exchange of viral disease ....................................... 1
1.2 Grapevine viruses ........................................................................................................... 2
   1.2.1 Fanleaf degeneration ............................................................................................. 3
   1.2.2 Grapevine leafroll disease complex ...................................................................... 4
1.3 GRSPaV implications in disease .................................................................................... 6
1.4 GRSPaV taxonomy, genome and virion structure .......................................................... 12
1.5 Replication and life cycle ............................................................................................... 17
1.6 GRSPaV genetic diversity and phylogeny ..................................................................... 21
1.7 GRSPaV phylogroup disease associations ................................................................... 30
1.8 Full-length infectious clones ......................................................................................... 31
1.9 FLCs in literature .......................................................................................................... 34
1.10 Rationale ....................................................................................................................... 37
1.11 Hypotheses .................................................................................................................... 38
1.12 Objectives ..................................................................................................................... 38

CHAPTER 2: MATERIALS AND METHODS ........................................................................ 40

2.1 GRSPaV Diversity ......................................................................................................... 40
   2.1.1 Plant Total RNA extraction .................................................................................... 40
   2.1.2 RT-PCR of RdRp fragments for sequencing ......................................................... 40
   2.1.3 Cloning of RdRp PCR products into pGEM-T-Easy ............................................. 41
   2.1.4 Sanger sequencing of plasmid DNA ...................................................................... 42
   2.1.5 Phylogenetic analysis of sequenced fragments ..................................................... 43
   2.1.6 CP diversity ........................................................................................................... 43
2.2 Full-Length Clone ........................................................................................................... 45
   2.2.1 Primers ................................................................................................................... 45
   2.2.2 PCR Amplification and Cloning of CaMV 35S Promoter ....................................... 45
2.2.3 RNA Extraction and RT-PCR of FLC Fragments .................................................. 49
2.2.4 Cloning FLC fragments into pGEM-T-Easy .......................................................... 50
2.2.5 Overlap PCR of F2 and F3 .................................................................................... 51
2.2.6 Sanger sequencing of FLC fragments in pGEM-T-Easy ......................................... 53
2.2.7 Cloning F4 and F3 into pBS35S ............................................................................ 56

CHAPTER 3: RESULTS .................................................................................................. 57
3.1 GRSPaV Diversity .................................................................................................... 57
  3.1.1 Genetic diversity of GRSPaV from Vitis varieties ................................................. 57
  3.1.2 RSP35/RSP36 amino acid sequence alignment .................................................. 64
  3.1.3 RSP35-SY/RSP36 to specifically target GRSPaV-SY ........................................... 66
  3.1.4 Syrah cultivars with and without red canopy ....................................................... 68
  3.1.6 CP Diversity sequencing outcomes .................................................................... 74
3.2 Full-Length Clone .................................................................................................... 75
  3.2.1 Sequence analysis of CaMV 35S promoter and HDVnos terminator sequences ...... 75
  3.2.2 Sequence analysis of pGEM-F2a and pGEM-F2b .................................................. 76
  3.2.3 Sequence analysis of pGEM-RSP13-36, pGEM-F3a and pGEM-F3b ...................... 77
  3.2.4 Sequence analysis of pGEM-F4 ........................................................................... 79
  3.2.5 Attempt to clone F4 into pBS-35S failed ............................................................... 79
  3.2.6 Attempt to clone F3 into pBS-35S failed ............................................................... 81

CHAPTER 4: DISCUSSION ............................................................................................ 83
4.1 GRSPaV variants in Ontario are genetically diverse ................................................. 83
4.2 GRSPaV-SY is likely involved in SD in Syrah, but not the sole causal agent .............. 88
4.3 The GRSPaV-PN sequence may in fact be a chimera .............................................. 89
4.4 GRSPaV distribution in hybrid varieties differed from Vitis vinifera varieties ........... 91
4.5 The FLC sequences are most closely related to the GRSPaV-SY isolate .................... 92
4.6 A GRSPaV-SY-related infectious clone is being constructed .................................... 97
4.7 Troubleshooting ..................................................................................................... 98
  4.7.1 Sequencing issues ............................................................................................... 98
  4.7.2 Full-length cDNAs for FLC ............................................................................... 100
  4.7.3 Difficulties in obtaining large cDNA fragments and issues with ethanol ............ 101
  4.7.4 Overlap PCR conditions .................................................................................... 102
4.7.5 Failed Clal digestion during F3-pBS35S cloning ........................................103
4.8 Future directions ..................................................................................................104
  4.8.1 CP diversity ..................................................................................................104
  4.8.2 Steps for completing the GRSPaV-SY FLC .............................................105
  4.8.3 Potential future experiments involving the GRSPaV-SY FLC .................107
4.9 Concluding remarks ..........................................................................................109
REFERENCES ........................................................................................................111
APPENDIX I: Primers ...............................................................................................121
APPENDIX II: GRSPaV diversity: Supplemental phylogenetic trees ..................126
APPENDIX III: Nucleotide alignment of FLC fragments ..................................134
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nucleotide identity between GRSPaV isolates</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>GRSPaV-SY genome FLC primer pairs and optimal annealing temperatures</td>
<td>52</td>
</tr>
<tr>
<td>3</td>
<td>GRSPaV clone distribution table</td>
<td>61</td>
</tr>
<tr>
<td>4</td>
<td>GRSPaV distribution in hybrid sources using primers RSP35-RSP36</td>
<td>73</td>
</tr>
<tr>
<td>5</td>
<td>Comparative analysis of GRSPaV-PN with four sequence variants</td>
<td>93</td>
</tr>
<tr>
<td>6</td>
<td>Primer table</td>
<td>121</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Grapevine disease symptoms</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>Genome organization of <em>Betaflexiviridae</em> genera</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>GRSPaV genome organization</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>Viral replication cycle in a plant cell</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>GRSPaV distribution using primers targeting the HEL and CP</td>
<td>27</td>
</tr>
<tr>
<td>6</td>
<td>Target sequence for primers RSP35 and RSP36 across 5 GRSPaV sequence isolates</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>GRSPaV distribution of Italian isolates shows similar phylogroups</td>
<td>29</td>
</tr>
<tr>
<td>8</td>
<td>GRSPaV-SY FLC Restriction Map – Original Plan</td>
<td>47</td>
</tr>
<tr>
<td>9</td>
<td>GRSPaV-SY FLC Restriction Map – Final Design</td>
<td>48</td>
</tr>
<tr>
<td>10</td>
<td>F2 overlap PCR diagram</td>
<td>54</td>
</tr>
<tr>
<td>11</td>
<td>Primer map including primers used for internal overlap regions</td>
<td>55</td>
</tr>
<tr>
<td>12</td>
<td>Neighbour joining GRSPaV distribution in <em>Vitis</em> using primers RSP35-RSP36</td>
<td>60</td>
</tr>
<tr>
<td>13</td>
<td>Alignment of RSP35-RSP36 aa sequences representing each phylogroup</td>
<td>65</td>
</tr>
<tr>
<td>14</td>
<td>Neighbour joining distribution of RSP35-SY-RSP36 clones from 14 <em>Vitis</em> sources</td>
<td>67</td>
</tr>
<tr>
<td>15</td>
<td>Neighbour joining distribution in Syrah with red canopy using primers RSP35-RSP36</td>
<td>69</td>
</tr>
<tr>
<td>16</td>
<td>Neighbour joining distribution in Syrah without red canopy using RSP35-RSP36</td>
<td>70</td>
</tr>
<tr>
<td>17</td>
<td>Neighbour joining distribution in hybrid varieties using primers RSP35-RSP36</td>
<td>72</td>
</tr>
<tr>
<td>18</td>
<td>FLC fragment alignment and orientation</td>
<td>80</td>
</tr>
<tr>
<td>19</td>
<td>SalI-digested F3 from pGEM-F3 subcloned into SalI-digested pBS35S-F3</td>
<td>82</td>
</tr>
<tr>
<td>20</td>
<td>FLC cloning strategy</td>
<td>108</td>
</tr>
<tr>
<td>21</td>
<td>Alignment map of FLC fragments and internal primers with the GRSPaV-SY reference sequence</td>
<td>124</td>
</tr>
<tr>
<td>22</td>
<td>Primer map including primers used for the GRSPaV-SY FLC and some commonly used primers for GRSPaV diversity analysis</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Maximum likelihood GRSPaV distribution in <em>Vitis</em> using primers RSP35-RSP36</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Maximum likelihood distribution of RSP35-SY-RSP36 clones from 14 <em>Vitis</em> sources</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Maximum likelihood distribution in Syrah with red canopy using primers RSP35-RSP36</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Maximum likelihood distribution in Syrah without red canopy using RSP35-RSP36</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Maximum likelihood distribution in hybrid varieties using primers RSP35-RSP36</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>MegAlign alignment of RSP35-RSP36 clones from hybrids</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Sequence alignment of the 35S CaMV promoter with that of pHST40</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Sequence alignment of the F2a sequence with nucleotides 1-2127 of GRSPaV-SY sequence</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Sequence alignment of the F2b sequence in pGEM-F2b and nucleotides 1659-4410 of the GRSPaV-SY sequence</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Sequence alignment of the F3 sequence in pGEM-F3 and nucleotides 4399-7592 of the GRSPaV-SY sequence</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Sequence alignment of the F4 sequence in pGEM-F4 and nucleotides 7581-8742 of the GRSPaV-SY sequence</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Sequence alignment of the HDVnos sequence in pHST40 with the HDVnos sequence in pBS</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Alignment of predicted aa sequence of GRSPaV-JH ORF1 with GRSPaV-SY ORF1</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Alignment of predicted aa sequence of GRSPaV-JH ORF2 with GRSPaV-SY ORF2</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Alignment of predicted aa sequence of GRSPaV-JH ORF3 with GRSPaV-SY ORF3</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Alignment of predicted aa sequence of GRSPaV-JH ORF4 with GRSPaV-SY ORF4</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Alignment of predicted aa sequence of GRSPaV-JH ORF5 with GRSPaV-SY ORF5</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>Agarose gel electrophoresis</td>
<td></td>
</tr>
<tr>
<td>AlkB</td>
<td>Alkylation B domain</td>
<td></td>
</tr>
<tr>
<td>CB</td>
<td>Corky bark</td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>Capsid protein/coat protein</td>
<td></td>
</tr>
<tr>
<td>GLD</td>
<td>Grapevine leafroll disease</td>
<td></td>
</tr>
<tr>
<td>HEL</td>
<td>Helicase</td>
<td></td>
</tr>
<tr>
<td>HVR</td>
<td>Highly variable region</td>
<td></td>
</tr>
<tr>
<td>KSG</td>
<td>Kober stem grooving</td>
<td></td>
</tr>
<tr>
<td>LNSG</td>
<td>LN-33 stem grooving</td>
<td></td>
</tr>
<tr>
<td>MTR</td>
<td>Methyltransferase</td>
<td></td>
</tr>
<tr>
<td>O-Pro</td>
<td>Ovarian tumour-like protease</td>
<td></td>
</tr>
<tr>
<td>P-Pro</td>
<td>Papain-like cysteine protease</td>
<td></td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
<td></td>
</tr>
<tr>
<td>RdRp</td>
<td>RNA dependent RNA polymerase</td>
<td></td>
</tr>
<tr>
<td>RSP</td>
<td>Rupestris stem pitting</td>
<td></td>
</tr>
<tr>
<td>RW</td>
<td>Rugose wood</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>Syrah Decline</td>
<td></td>
</tr>
<tr>
<td>SDM</td>
<td>Site-directed mutagenesis</td>
<td></td>
</tr>
<tr>
<td>sgRNA</td>
<td>Subgenomic RNA</td>
<td></td>
</tr>
<tr>
<td>ssRNA</td>
<td>single-stranded RNA</td>
<td></td>
</tr>
<tr>
<td>TGB</td>
<td>Triple Gene Block</td>
<td></td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
<td></td>
</tr>
<tr>
<td>VIGS</td>
<td>Virus induced gene silencing</td>
<td></td>
</tr>
</tbody>
</table>
LIST OF VIRUS ABBREVIATIONS

ASPV  Apple stem pitting virus
CaMV  Cauliflower mosaic virus
CLBV  Citrus leaf blotch virus
GFLV  Grapevine fanleaf virus
GLRaV  Grapevine leafroll associated virus
GRSPaV  Grapevine rupestris stem pitting-associated virus
GRVFV  Grapevine rupestris vein feathering virus
GSyV-1  Grapevine Syrah virus-1
GVA  Grapevine virus A
GVB  Grapevine virus B
HDV  Hepatitis delta virus
OBDV  Oat blue dwarf virus
PVM  Potato virus M
PVX  Potato virus X
PVY  Potato virus Y
RhPV  Rhopalosiphum padi virus
CHAPTER 1: INTRODUCTION

1.1 Grapevine history: Cultivation and exchange of viral disease

Viticulture has been practiced worldwide for thousands of years as a result of the cultural and historical importance of wine. Thus, grapevines have been established as one of the most ancient and extensively cultivated fruit crops today (This et al., 2006; Reisch & Pratt, 1996; Meng & Gonsalves, 2007). Viticulture gained economic importance as the wine and juice industries grew, making mass cultivation of grapes in vineyards a profitable yet scenic phenomenon (This et al., 2006). Countries all across the globe have established vineyards including the most popular wine-producing grapevine species, *Vitis vinifera*, and hybrids of *V. vinifera* and North American grapevine species to implement unique flavours, colours, aromas, and other characteristics that are difficult, if not impossible, to recreate.

The health of a vineyard can reflect directly on its fruit production and wine quality. Unfortunately today grapevines are subject to nearly 70 different viral species, with a commonality for multiple infections in one grapevine (Meng et al., 2006; Martelli & Boudon-Padieu, 2006; Coetzee et al., 2010). Some grapevine viruses may have coevolved with the crop as certain variants are only found in certain grape cultivars, and this coevolution allowed evolutionary masking mechanisms to arise and made observable symptoms minimal or absent (Meng et al., 2006; Meng et al., 2013; Terlizzi et al., 2010). It is also possible that scion propagation has limited the evolution of grapevines, which may have allowed viruses to evolve to their maximum fit. The asymptomatic nature of some viruses was overturned by the increased popularity of using grafting techniques in viticulture (Terlizzi et al., 2010). Grafting of grapevines was first implemented to combat the destruction of European vineyards, due to
unintentionally introduced phylloxera (*Daktulosphaira vitifoliae*), by grafting onto phylloxera-resistant rootstock cultivars from North America (*This et al.*, 2006; *Meng & Gonsalves*, 2007; *Meng et al.*, 2006). Grafting different grapevine species and cultivars onto one another has resulted in an increase in disease symptoms caused by grapevine viruses, as they become present in a species they were not native to. In addition, the spread of grapevine viruses is accelerated by global movement and exchange of plant materials between viticulturists. Because of these factors, routine indexing of grapevines became an important tool in disease prevention and management.

1.2 Grapevine viruses

Research on grapevine viruses has become extensive due to the range of viruses which infect this important crop. While there is still much to be learned about these viruses, there is a general consensus that mixed infections are likely the culprit of most detrimental diseases. Many of these viral species include two or more sequence variants or “strains” which differ in nucleotide sequence but belong to the same species. The overall genetic makeup, virion structure, and infectious properties constitute viruses of the same species, but nucleotide sequences have diverged to a point of distinction. When looking at these viral variants sequences phylogenetically, those which have related sequences, evolving from a common ancestor, form clusters or phylogroups. These viral variant groups which may have associations with different disease symptoms. Understanding grapevine viruses as complex individual species and as a population will benefit vineyard health and disease prevention.
1.2.1 Fanleaf degeneration

Fanleaf degeneration is a grapevine disease of vast importance. The reason for this problem, Grapevine fanleaf virus (GFLV) species (family Secoviridae, genus Nepovirus) causes what has been described as the most severe disease by a grapevine virus and has been found in all vine-growing regions worldwide (Martelli & Savino, 1990; Vigne et al., 2004; Jez Krebelj et al., 2015; Sanfaçon et al., 2009). Fanleaf degeneration induced by GFLV results in malformed berries, canes, and leaves, reduced yield and quality of fruit, and reduced grapevine lifespan (Martelli & Savino, 1990; Martelli, 1993; Sanfaçon et al., 2009). Like many viruses, GFLV has multiple sequence variants which attribute to a range of symptoms including fanleaf, chlorotic ringspots and lines, yellow mosaic, banding within veins, and leaf mottle (Martelli & Savino, 1990; Vigne et al., 2004). GFLV is transmitted by a nematode vector, Xiphinema index, as well as grafting propagation, making spread within and between vineyards rapid (Sanfaçon et al., 2009; Jez Krebelj et al., 2015). A genetic diversity study of the helicase (HEL) and polymerase (POL) regions of six GFLV isolates from the Czech Republic and Italy revealed high conservation between viral isolates from different grapevine cultivars, and identical mutations were found between some isolates from Italy and the Czech Republic (Eichmeier et al., 2011). While this may be the result of infected grapevine propagating materials from the same source used in both vineyards, it may also be an indication of conservation of a potentially beneficial mutation in GFLV variants. Phylogenetic analysis of the POL gene region (365 nt) by the Czech Republic group revealed two distinct clusters when analyzed using neighbour-joining method with 100 bootstrap replicates. Eichmeier et al. (2011) describe the range of symptoms observed on known grapevine cultivars infected with different POL gene sequences from GFLV variants. The isolates from California clustered with a reference sequence from a French GFLV isolate.
forming one group, while variants from Italy, the United States (non-Californian sequences), the Czech Republic and New Zealand formed the other (Eichmeier et al., 2011).

A recent GFLV genetic diversity study conducted in China indicated some level of geographic segregation of sequence isolates. When analyzing the data set of 142 clones phylogenetically, all Chinese isolates grouped together in one group. However, the sequence isolates from Chile, America, Italy, France, Sprain, Iran, Slovenia, and South Africa were dispersed among one another. GFLV isolates in this study clustered into two major phylogroups, with the group of Chinese isolates forming a third divergent group (Zhou et al., 2015).

1.2.2 Grapevine leafroll disease complex

A well-studied complex of grapevine viruses is the grapevine leafroll disease (GLD) complex of viruses. As the name suggests, grapevines with GLD show a downward rolling of leaf margins, and additionally a reduction in vine vigor, quality, and yield of berries (Maree et al., 2013; Poojari et al., 2012). Notably different symptoms are seen between red and white cultivars with GLD. Red cultivars with GLD display symptoms of red leaves with green veins, while white cultivars often show chlorosis in the areas between veins in addition to the downward rolling of leaf blades (Maree et al., 2013). The GLD complex is highly associated with the viral Closteroviridae family of positive sense ssRNA viruses. Virions of Closteroviridae are very long, flexuous, non-enveloped filaments. Closteroviridae is made up of four genera; Crinivirus, Ampelovirus, Closterovirus, and Velarivirus. The GLD complex viral species are spread among three of these genera; Ampelovirus, Closterovirus, and Velarivirus (Martelli et al., 2012). The primary causal agent of GLD, Grapevine leafroll-associated virus-3 (GLRaV-3) (genus Ampelovirus), is known due to its worldwide distribution in grapevines with GLD (Maree
et al., 2013). GLRaV-3 is transmitted by a variety of insect vectors which include mealy bugs and soft scale insects (Maree et al., 2013). GLRaV-3 shows notably different symptoms in different grapevine scions and rootstocks, in addition to symptomless vines, influenced by season and climate (Maree et al., 2013). Phylogenetic analysis of GLRaV-3 CP sequence variants gave six well-supported phylogroups, indicating evolutionary divergence of this species (Maree et al., 2013). In addition to GLRaV-3, the GLD complex consists of several other GLRaV species including GRLaV-1 through GLRaV-4 and GLRaV-7, each likely also composed of many sequence variant groups (Martelli et al., 2014). Different GLRaVs involved in GLD was first documented by Gugerli et al. (1984), when observations of two distinctly different Closteroviridae virions, now named GLRaV-1 and GLRaV-2, were obtained from symptomatic grapevine leaves (Maree et al., 2013; Gugerli et al. 1984). GLRaV-7 differs from the other GLRaVs as it is the only representative of the Velarivirus genus and a transmission vector for this virus is still unknown (Maree et al., 2013; Al Rwahnih et al., 2011; Jelkmann et al., 2012). More recently, sequence variants of GLRaV-4 (genus Ampelovirus) designated GLRaV-Pr, GLRaV-Dr, and GLRaV-Car, have been identified (Maree et al., 2013). Interestingly, GLRaV-2 (genus Closterovirus) includes many sequence variants, but two; GLRaV-2-PN and GLRaV-2-RG, are known to induce fatal graft incompatibility when present in certain combinations of rootstock and scion varieties (Maree et al., 2013). V. vinifera scion cultivars grafted onto Kober 5BB, Couderc 1616, and Teleki 5C rootstocks were some of the most sensitive to these GLRaV-2 variants (Maree et al., 2013; Bertazzon et al., 2010; Alkowni et al., 2011). It is important to understand that different sequence variants of GLRaV-2 induce different symptoms, including vine death, when present in different grapevine varieties. This may also be the case for other grapevine viruses. Transmission of GLRaV-2 by insect vectors is yet unknown and there is no
data supporting transmission of GLRaV-2 within vineyards (Maree et al., 2013; Poojari et al., 2012).

1.3 GRSPaV implications in disease

A detrimental disease complex within viticulture is the rugose wood (RW) complex – a group of damaging diseases transmissible via grafting of grapevines (Terlizzi et al., 2010; Martelli, 1993). The RW complex includes corky bark (CB) disease, Kober stem grooving (KSG), LN stem grooving (LNSG), and rupestris stem pitting (RSP) (Figure 1A-D) (Martelli, 1993; Garau et al., 1994). These diseases disrupt the flow of nutrients and water through grapevine vascular tissue, resulting in malnourished regions of the plant (Gribaudo et al., 2006). KSG symptoms, long depressions along the stem, is the result of infection of cultivar ‘Kober 5BB’ with a highly prevalent grapevine virus called Grapevine virus A (GVA) (family Betaflexiviridae; genus Vitivirus) (Garau et al., 1994). Kober 5BB rootstocks are commonly used by viticulturists, and have been used as GVA indicator plants prior to molecular detection (Komínek & Komínková, 2008). GVA is primarily spread via grapevine propagation and several mealybug insect vectors, including Pseudococcus longispinus (Rosciglione et al., 1988). Three major phylogroups of GVA isolates have been described in the literature from South Africa and the Czech Republic (Goszczynski & Jooste, 2003; Kominek & Komínková, 2008). However, Kominek and Komínková (2008) reported the GVA isolate they studied from the Czech Republic is genetically divergent from other reported GVA isolates and does not cause KGS on the Müller Thurgau clone 25/7 grafted Kober 5BB rootstocks. GVA is also sensitive to grapevine thermotherapy, experimentally shown to reduce GVA titre by approximately 70% (Panattoni & Triolo, 2010). Thermotherapy, as used by Panattoni and Triolo (2010) involved growing virus-
inoculated grapevines in a controlled growth chamber at 37°C +/- 0.5°C for 48 days in an effort to eradicate common grapevine viruses.

RSP is one of the most commonly observed diseases of the RW complex, alongside KSG (Garau et al., 1994). First documented in 1970 in California, grapevines with RSP display characteristic pit formation along stems, delayed growth, stunting, and sometimes vine decline (Meng & Gonsalves, 2007; Goheen, 1988). These pits are the result of xylem modifications linked to viral infection (Martelli et al., 2007). Grapevines exhibiting these disease symptoms produce poor yield and quality of grapes. Grapevine rupestris stem pitting-associated virus (GRSPaV) is a key virus highly associated with RSP and is likely the most prevalent virus found in grapevines (Meng et al., 2006; Terlizzi, 2011). RSP has been closely associated with this virus as the vast majority of plants with RSP have tested positive for GRSPaV (Meng et al., 1999; Zhang et al., 1999; Terlizzi et al., 2011). The exact etiology of RSP has not yet been established due to the lack of fulfillment of Koch’s postulates (Meng & Gonsalves, 2007; Meng et al., 2013; Meng et al., 1998). GRSPaV was recently established as the putative causal agent for Grapevine vein necrosis (GVN) (Figure 1F) (Bouyahia et al., 2005; Bouyahia et al., 2006). GVN is characterized by necrosis of veinlets on leaf blades and in severe cases, near complete termination of growth and death of the vine (Bouyahia et al., 2006). In all European grapevine cultivars and most Vitis spp. and hybrids from North America GVN is latent, with the exception of rootstock 110R (V. rupestris x V. berlandieri) (Bouyahia et al., 2006). Thus, 110R has been the biological indicator for GVN prior to molecular-based detection. GRSPaV has been proposed to be the causal agent of GVN and there is no known natural vector; graft transmission is the primary method of spread (Bouyahia et al., 2006).
The detection of GRSPaV in asymptomatic grapevines suggests that biotic or abiotic factors may act in combination with GRSPaV to cause disease symptoms (Lima et al., 2009). These may include grapevine rootstock and scion cultivar, GRSPaV variants present, other microbial pathogens present, soil salinity and moisture, and seasonal changes. Grafting with infected materials or infected propagating tools is the primary method for the spread of GRSPaV, however it remains unclear how this virus spreads in the wild (Meng & Gonsalves, 2007).

Research has suggested that pollen and seeds from infected grapevines may account for a small fraction of GRSPaV transmission (Rowhani et al., 2000; Stewart & Nassuth, 2001). Further, it was determined experimentally that 0.4% of seedlings from infected ‘Cabernet Sauvignon’ were GRSPaV-positive (Lima et al., 2006b). While pollen and seed transmission do not appear to be viable means of GRSPaV transmission, the existence of a yet unknown insect vector for GRSPaV may be a method for the majority of the natural spread of infection (Meng & Gonsalves, 2007). Current knowledge indicates that grapevines are the only natural host for GRSPaV; however herbaceous hosts have been used in the laboratory as an experimental system (Meng & Gonsalves, 2007; Meng et al., 2013).

The implications of GRSPaV in Syrah decline (SD) has been the subject of debate among grapevine virologists. SD – a disease which causes swelling and cracking at the graft union and early leaf reddening – affects all grapevine rootstock and scion clones, and results in unripe berries with poor colour and sugar content (Figure 1F) (Renault-Spilmont et al., 2003). SD has devastated young vineyards in many countries, which include France, Australia, the United States of America, and Canada (Meng et al., 2013; Habili et al., 2006; Lima et al., 2009; Renault-Spilmont et al., 2003). Curiously, non-grafted Syrah grapevines in phylloxera-free soils in Chile and Argentina have also experienced SD, suggesting it may not require grafting for
Figure 1: Grapevine disease symptoms associated with *Betaflexiviridae*. (A) Corky bark pit- or cork-like symptoms near the graft union. (B) A Kober 5BB rootstock displaying Kober stem grooving symptoms (top) in comparison to a healthy control (bottom). (C) LN33 rootstock with LN33 stem grooving symptoms upon peeling away of the bark. (D) Rupestris stem pitting pit symptoms (left) compared to health controls (right). (E) Vein necrosis symptoms on the underside of a leaf. (F) Syrah decline red canopy symptoms on a Syrah vine next two healthy vines. (WSU Viticulture & Enology Online; Martelli, 1993; Zorloni, 2011)
symptoms to develop (Renault-Spilmont et al., 2003). SD is of great concern within viticulture as it severely affects grape production and is poorly understood. It is believed that Syrah cultivars may be most prone to infection and exhibit the most conspicuous symptoms (Battany et al., 2004; Lima et al. 2006; Al Rwahnih et al. 2009). Symptoms that arise on Syrah cultivars include uniform reddening and/or scorching of leaves, swelling and cracking of the graft union, stem necrosis, and in severe cases eventual death of the vine (Battany et al., 2004; Stamp, 2004). Interestingly, grapevines with only grooves and graft union swelling can survive for many years, whereas grapevines displaying grooving and red leaves often die within a few years following the initial early reddening of the leaves (Beuve et al., 2012). It has been suggested that the symptoms observed on ‘Syrah’ are the result of genetic incompatibility, viral influence, or a combination of both (Golino, 1993). Previous etiological studies on SD suggested that Syrah clones have different sensitivities to SD symptoms, making them more prone to detrimental symptoms (Renault-Spilmont et al., 2007).

A definitive causal pathogen for SD has been difficult to identify as a result of mixed infection by multiple viruses, overlap of symptoms between diseases, and seasonal variability in symptoms (Al Rwahnih et al., 2009). Reports from California, France and Australia have suggested a strong correlation between SD and GRSPaV (Al Rwahnih et al., 2009; Lima et al., 2006a; Habili et al. 2006). Deep sequencing analysis of total RNA from two Syrah vines, one with and one without decline symptoms, revealed the significant majority (46,029 of 57,365 hits) of viral sequences to be derived from GRSPaV (Al Rwahnih et al., 2009). This study did not conclude GRSPaV was responsible for SD, but may be involved in mixed infection. GRSPaV is the most commonly detected grapevine virus (Meng & Gonsalves, 2003; Meng et al., 2006; Habili et al., 2006). The second largest number of hits (9,697) in the Al Rwahnih et al., (2009)
Syrah deep sequencing study were derived from Grapevine rupestris vein-feathering virus (GRVFV; family Tymoviridae, tentative genus Marafivirus). GRSPaV and GRVFV have been known as mild or asymptomatic when present as a single agent, however disease symptoms may be caused or modified by interactions between other infectious species (Meng et al., 2005; Komar et al., 2007; Al Rwahnih et al., 2009). This is very likely the case for SD. A small subset of sequence hits (1,527) from the deep sequencing results led to the first description of Grapevine Syrah virus-1 (GSyV-1), which shares a similar genome organization to marafiviruses (order Tymovirales) (Al Rwahnih et al., 2009). GSyV-1 may be a key virus in SD symptoms as it was discovered in the diseased Syrah clone, however this vine had significantly more viral or viroid sequences than the asymptomatic plant (Al Rwahnih et al., 2009).

A study conducted in France by Beuve et al. (2012) detected GRSPaV in 100% (22/22) of Syrah clones with varying SD symptoms. Of the 271 clones sequenced from 22 Syrah cultivars, 60% clustered with the GRSPaV-SY lineage (Beuve et al., 2012). However their data did not support the suggestion that GRSPaV is responsible for SD. The findings by Beuve et al. (2012) support the previous research conducted by Goszczynski (2010) that concluded GRSPaV, including divergent –SY variants, were present in ‘Syrah’ clones with and without SD symptoms, and are not responsible for SD. These conclusions may be hindered by the staggered initial onset of SD symptoms. Symptoms may take a few years to arise, making determination of non-declining Syrah difficult to interpret. Additionally, these conclusions were made based on findings from grapevines selections in France and South Africa, regions with very different grapevine growing climates with much less cold damage and stress. A greater understanding of this disease is necessary as SD is devastating vineyards each year.
1.4 GRSPaV taxonomy, genome and virion structure

Both GRSPaV and GVA are members of the *Betaflexiviridae* family. This family is of the order *Tymovirales* and is made up of two sub-families; *Quinvirinae* and *Tririnae*, and some unassigned species (ICTV, 2017). *Quinvirinae* is made up of three genera; *Carlavirus*, *Foveavirus*, and the newly established *Robigovirus*. *Tririnae* is made up of eight genera; *Capillovirus*, *Chordovirus*, *Citrivirus*, *Divavirus*, *Prunevirus*, *Tepovirus*, *Trichovirus*, and *Vitivirus*. Members of *Betaflexiviridae* genera show slight differences in genome structure which may give some insight to their evolutionary paths (Figure 2). *Betaflexiviridae* viruses are positive-sense, single-stranded RNA (ssRNA) viruses which infect plants, mostly perennials (Martelli et al., 2007). The genome of *Betaflexiviridae* members include sequences that encode highly conserved methyl transferase (MTR), a superfamily-1 (SF-1) RNA helicase (HEL), and RNA dependent RNA polymerase (RdRp) domains, as well as contain a cap structure at the 5’ end and a polyA tail at the 3’ end (Martelli et al., 2007). Some other protein domains are found in some but not all members of *Betaflexiviridae*, which include the papain-like protease (P-Pro), ovarian tumor protease (O-Pro), and alkylation B (AlkB) domains. Additionally, the movement proteins in members of *Betaflexiviridae* are either in the form of the triple gene block (TGB) movement proteins or the p30-like superfamily of movement proteins (Martelli et al., 2007). The TGB is a genetic feature encoding movement proteins common to certain genera of *Alpha- and Betaflexiviridae* viruses (Meng & Gonsalves, 2007). The TGB MPs have been classified into two categories on the basis of phylogeny and viral movement mechanisms. Viruses with potex-like TGBs form filamentous virions with monopartie RNA genomes and require CP-dependent cell-to-cell movement. Viruses of the *Potexvirus* and *Allexivirus* genera of the *Alphaflexiviridae* family, and *Carlavirus* and *Foveavirus* of the *Betaflexiviridae* have potex-like TGBs (Park et al.,
Viruses with hordei-like TGBs are rod-shaped with multipartie RNA genomes and are not CP-dependent for cell-to-cell movement. Included among hordei-like TGB viruses are *Hordeivirus* and *Pomovirus* of the family *Virgaviridae* (Park et al., 2014; Adams et al., 2004; Morozov & Solovyev, 2003).

Virions of *Betaflexiviridae* are filamentous, ranging from ~470 to 1000 nm and 10 to 15 nm in diameter with variation of flexibility (Meng & Gonsalves, 2007; Martelli et al., 2007). *Betaflexiviridae* virions are non-enveloped with a capsid made up of a repeated single capsid protein (CP) (Martelli et al., 2007). Within the same order *Tymovirales*, *Alphaflexiviridae* is a family of viruses which primarily infect fungi and plants, such as citrus and potatoes. Like *Betaflexiviridae*, virions are also non-enveloped, flexuous, filamentous virions. Genomes of *Alphaflexiviridae* are also positive-sense, linear ssRNA. The mode of transmission of most *Alphaflexiviridae* genera is unknown, however species from *Allexivirus* and *Potexvirus* are known to be transmitted by mites and insects, respectively (ICTV, 2017; Hulo et al., 2011).

GRSPaV is a *Betaflexiviridae* virus belonging the *Foveavirus* genus (King et al., 2011a; Martelli & Jelkmann, 1998; Petrovic et al., 2003). Comparisons with partial or complete GRSPaV genomic nucleotide (nt) and amino acid (aa) sequence revealed closest relationships to ASPV (genus *Foveavirus*) and more distantly, Potato virus M (PVM; genus *Carlavirus*) (Meng et al., 1998). Virions of GRSPaV are 723 nm long flexuous filaments which contain a genome of 8,725 nucleotides (Zhang et al., 1998; Meng et al., 1998; Petrovic et al., 2003). The GRSPaV genome contains 5 open reading frames (ORFs) and is believed to be capped at the 5’ end and have a polyA tail at the 3’ end (Figure 3) (Meng & Gonsalves, 2007). The genome has a 5’ non-coding region of 60 nt and a 3’ non-coding region of 176 nt. ORF1 contains the elements characteristic of the Alphavirus-like superfamily of RNA viruses, to which GRSPaV belongs.
ORF1 encodes a polyprotein with domains including a MTR, HEL, and RdRp. Uncharacteristic of Alphavirus-like superfamily, but common for several genera of Betaflexiviridae are the presence O-Pro and P-Pro proteases (Gorbalenya et al., 1988; Meng et al., 2013). The P-Pro of GRSPaV has a cysteine-histidine catalytic diad found in other plant RNA viruses, which supports the hypothesis that the polyprotein encoded by ORF1 undergoes proteolytic processing for downstream function of polypeptides (Meng & Gonsalves, 2007; Russo et al., 2006; Udaskin, 2015). Products of auto-cleavage likely function together as replication machinery (Meng & Gonsalves, 2007). Subcellular localization of the GRSPaV replicase polyprotein in N. benthamiana and tobacco protoplasts revealed cytoplasmic punctate structures, potentially representing viral replication complexes (Prosser et al., 2015). The N-terminal region of the replicase polyprotein, the MTR, associated with an undetermined cellular membrane (Prosser et al., 2015). The GRSPaV genome also encodes the AlkB domain, of which the function is yet to be determined. Bratlie and Drabløs (2005) suggested the AlkB domain may have a role in combating the host RNA silencing machinery by reverting host-induced methylation of viral RNA. ORF2 to 4 collectively make up the TGB (Zhang et al., 1998; Meng et al., 1998), encoding respectively TGBp1, TGBp2, and TGBp3. TGBp1 is predicted to be a 24.4 kDa protein belonging to the SF-1 RNA helicase family, capable of both cytosolic and nuclear localization (Meng & Gonsalves, 2007; Rebelo et al., 2008). TGBp1 is classified as an SF-1 helicase on the basis of the 7 conserved SF-1 motifs present, genetically divergent from the SF-1 HEL in ORF1 (Meng & Gonsalves, 2007; Morozov & Solovyev, 2003). The helicase activity of TGBp1 may be involved in breaking down secondary structures in viral RNAs prior to translocation, as well as host RNA silencing suppression and movement of other proteins.
**Quinvirinae**

- **Carlavirus**
  - MTR
  - AlkB
  - ORf
  - Pro
  - p-Pro
  - HEL
  - RdRp
  - TGBp1
  - CP
  - NB
  - An

- **Foveovirus**
  - MTR
  - AlkB
  - ORf
  - Pro
  - p-Pro
  - HEL
  - RdRp
  - TGBp1
  - CP
  - An

- **Robigovirus**
  - MTR
  - AlkB
  - ORf
  - Pro
  - p-Pro
  - HEL
  - RdRp
  - TGBp1
  - CP
  - An

**Trivirinae**

- **Citrivirus**
  - MTR
  - AlkB
  - ORf
  - Pro
  - p-Pro
  - HEL
  - RdRp
  - p30-like MP
  - CP
  - An

- **Prunevirus**
  - MTR
  - AlkB
  - ORf
  - Pro
  - p-Pro
  - HEL
  - RdRp
  - p30-like MP
  - CP
  - NB
  - An

- **Vitivirus**
  - MTR
  - ORf
  - Pro
  - HEL
  - RdRp
  - 20k
  - p30-like MP
  - CP
  - NB
  - An

- **Divavirus**
  - MTR
  - ORf
  - HEL
  - RdRp
  - p30-like MP
  - An

- **Capillovirus**
  - MTR
  - AlkB
  - ORf
  - Pro
  - HEL
  - RdRp
  - p30-like MP
  - CP
  - An

- **Tepovirus**
  - MTR
  - ORf
  - HEL
  - RdRp
  - CP
  - NB
  - An

- **Trichovirus**
  - MTR
  - AlkB
  - ORf
  - Pro
  - HEL
  - RdRp
  - p30-like MP
  - CP
  - NB
  - An

**Chordvirus**

Figure 2: Genome organization of ten Betaflexiviridae genera. The ORFs encoding movement proteins (blue) differ between members of *Quinvirinae* and *Trivirinae*. *Quinvirinae* genomes contain the TGB MPs, *Trivirinae* genomes contain the p30-like MP. *Trichovirus* may contain an unknown region (?) between the replicase and p30-like MP. Some members from both *Quinvirinae* and *Trivirinae* contain a nucleic acid binding sequence (NB). *Chordvirus* genome information was not available through NCBI or other sources and a diagram is not included. An represents a polyA tail. See text for explanation of the abbreviations.
**Figure 3**: GRSPaV genome organization. ORF1 (yellow) is from nt 61-6546 and contains the MTR, HVR, AlkB, O-Pro, P-Pro, HEL, and RdRp. ORF2-4 (blue) represent TGBp1-3 movement proteins. ORF2 is from nt 6577-7242, followed by ORF3 from nt 7244-7597, and ORF4 from nt 7581-7760 which in-part overlaps ORF3. ORF5 (green) encoding the CP gene is from nt 7770-8549. The 60 nt 5’ NCR and 176 nt 3’ NCR are not shown. The polyA tail is represented by An. See text for explanation of the abbreviations.
away from the plasmodesmata (Meng & Gonsalves, 2007). It is also predicted to be involved with the binding and movement of ssRNA viral ribonucleoprotein (RNP) complexes into a neighbouring cell by increasing the plasmodesmata size exclusion limit (Meng & Gonsalves, 2007; Howard et al., 2004; Morozov & Solovyev, 2003). TGBp2 is a 12.8 kDa protein and TGBp3 is an 8.4 kDa protein presumably, produced by ribosomal leaky scanning of a single sgRNA (Meng & Gonsalves, 2007). Together TGBp2 and TGBp3 may play a role in translocation of viral RNPs to the plasmodesmata, and subsequently to the neighbouring cell via associations with the cell membrane (Meng & Gonsalves, 2007). The product of ORF5 is the CP which is essential for virion production. The CP is 28 kDa in size and forms the interlocking pieces of the virion capsid. The presence of a conserved amino acid motif among filamentous viruses CP, “R/QX-XFDF” (where X is any residue), likely allows for salt bridge formation between CP polypeptides (Dolja et al., 1991; Meng & Gonsalves, 2007). The CP contains a nuclear localization signal not found in other members of the Foveavirus genus, and which was demonstrated to target the nucleus in tobacco protoplasts (Meng & Li, 2010). The presence of a nuclear localization signal encoded in the CP suggests some interaction of the viral CP with the cell nucleus, such as post transcriptional modifications (Meng et al., 2003; Meng et al., 2013; Petrovic et al., 2003).

1.5 Replication and life cycle

Little is known about the specifics of the GRSPaV life cycle. However, a basic understanding of the viral genomic replication and life cycle can be inferred from what is known about other positive-sense ssRNA plant viruses. While the transmission vector for GRSPaV is yet unknown, aphids are the most commonly known vector for plant viruses, and at least 100 different aphid species exist that are known as vectors for plant viruses (Whitfield et al., 2014). It
is estimated that 28% of known plant viruses are transmitted by aphids, however some related grapevine viruses with known vectors, like GVA, are transmitted by mealy bugs and soft-scale insects (Maree et al., 2013). Aphid transmission of GRSPaV cannot be assumed, but entry into the host cell is likely mediated by insect vectors feeding on host plants. Invasion of the host, replication of vRNA, cell-to-cell movement, long distance movement, and release from the host are all critical stages of viral infection. Host invasion is carried out through physical penetration by the aphid into the phloem cells of the host plant where the infection cycle begins (Figure 4, Step 1).

The order Tymovirales includes four families which share similarities in replication proteins, most of which infect plants (Martelli et al. 2007). Potexviruses have been fairly well studied and what is known about the Potexvirus life cycle can be used to predict the basic life cycle of GRSPaV. Genomic organization of members of Potexvirus (family Alphaflexiviridae) are very similar to the organization of those of the Foveavirus genus, and are known to be transmitted by aphids (Baulcombe et al., 1993). Like Foveaviruses, Potexviruses have five ORFs; the first encoding replication machinery including a RdRp for replication, overlapping ORFs 2-4 encode the TGB movement proteins, and ORF5 encodes a CP (Adams et al., 2004; Park et al., 2014). Once inside the cell, positive-sense single-stranded vRNA is released from the virion to the cytoplasm and host translation machinery is used for production of the viral replicase encoded in ORF1 (Figure 4, Step 2). Host factors, like lipid membranes, and the replicase form the viral replication complex (VRC). The VRC encloses and protects replication machinery during negative-sense RNA synthesis from the positive-sense vRNA. Newly synthesized negative-sense vRNA is used to synthesize new positive-sense vRNA (genomic replication) or positive-sense viral sgRNAs, such as those encoding the TGB MPs (movement)
(Figure 4, Step 3). TGBp1-3 and the CP are translated from the positive-sense viral sgRNAs using host machinery (Figure 4, Step 4). The repeated CPs encapsidate the progeny positive-sense vRNA in the process of new virion production. The TGB MPs move progeny from cytoplasm to plasmodesmata and onto neighbouring cells for spread of infection (Figure 4, Step 5). For potexviruses, cell-to-cell movement requires the three TGB MPs and the CP (Park et al., 2014; Forster et al., 1992, Morozov & Solovyev, 2003; Niehl & Heinlein, 2011; Schoelz et al., 2011; Solovyev et al., 2012). However, not all genera with TGB MPs require CP presence, such as the hordei-like TGB viruses described previously (Park et al., 2014; Adams et al., 2004; Morozov & Solovyev, 2003). Unfortunately the control of aphids is not as simple as other insect vectors of plant viruses, like coleopterans or lepidopterans which are susceptible to transgenic crops expressing insecticidal proteins isolated from *Bacillus thuringiensis* (Whitfield et al., 2014). These proteins, called Cry proteins, bind to midgut receptors of susceptible insects after feeding on the plant, but aphids do not appear to have the necessary midgut receptor, or the aphid midgut is unsuitable for function of the Cry proteins. Because of this, insecticides are the best current available option for control of aphids (Whitfield et al., 2014).
Figure 4: Adapted from Roossinck (2005). Viral replication cycle in a plant cell. 1. Injection into the cell by an aphid. Entry may also occur by movement through plasmodesmata from an infected neighbouring cell, or mechanical inoculation of viral RNA or virions under experimentation. 2. Early translation of the MTR (MTr), HEL (H), and RdRp (Pol) from genomic vRNA. 3. Synthesis of sgRNAs. The VRC is formed including MTR, HEL, RdRp and host factors (HF). Within, positive-sense RNA is copied to negative-sense RNA, and replication of positive-sense genomic RNA and subgenomic messenger RNAs are made from the negativesenses template via internal initiation. 4. MP and CP mRNAs are translated. 5. The CPs encapsidate the new genomic positive-sense RNA forming new virions. The MPs change the size-exclusion limit of the plasmodesmata and facilitate movement of progeny virions or viral RNA-protein complexes to the neighbouring cell.
1.6 GRSPaV genetic diversity and phylogeny

The interest in diversity of GRSPaV is stimulated by the reoccurring observation that sequence isolates form distinct phylogenetic groups, regardless of geographical location (Meng et al., 2006; Terlizzi et al., 2009). This clustering suggests a close relationship between isolates and may give insight to related biological properties like symptom phenotypes, as discussed in detail below. Today there are 15 fully sequenced isolates of GRSPaV. Included among these are GRSPaV-1 (NC001948) sequenced from mixed sources, GRSPaV (AF026278) first discovered in *V. vinifera* Cabernet Sauvignon, GRSPaV-SG1 (AY881626), the predominant isolate found in indicator *V. rupestris* St George, GRSPaV-BS (AY881627) which was discovered in the French-American hybrid Bertille Seyve 5563, and GRSPaV-PN (AY368172) found in *V. vinifera* Pinot Noir clone 23/3309 (Zhang et al., 1998; Meng et al., 1998; Lima et al., 2009; Meng et al., 2005). The remaining fully sequenced isolates are GRSPaV-SY (AY368590) from declining *V. vinifera* Syrah, GRSPaV-MG (FR691076) from *V. vinifera* Moscato Giallo, GRSPaV-GG (JQ922417) from *V. riparia* Grande Glabre, GRSPaV-PG (HE591388) from *V. vinifera* Pinot Gris, GRSPaV-3138-07 (JX559646) of an unknown *V. vinifera* cultivar, and GRSPaV-WA (KC427107) from *V. vinifera* Merlot and Cabernet Franc (Meng et al., 2013; Poojari et al., 2012; Lima et al., 2006a; Martelli et al., 2011; Giampetruzzi et al., 2012; Rott et al., 2012). Four fully sequenced isolates were added in 2015; GRSPaV-JF (KR054734) and GRSPaV-LSL (KR054735), both detected in China in 2015, GRSPaV-TannatRspav1 (KR528585) detected from *V. vinifera* in South Korea, and the near complete GRSPaV-VF1 (KT948710) first detected in Brazil (Hu et al., 2015; Cho et al., 2015; Fajardo & Nickel, 2015).

The natural vector for spread of GRSPaV remains unknown, posing the question of how grapevines in the genus *Vitis* became infected with multiple sequence variants. The theory that
the GRSPaV ancestor co-evolved with its host and diverged as *Vitis* diverged may have resulted in the GRSPaV-1 lineage adapted to *V. riperia* and the GRSPaV-SG1 lineage adapted to *V. rupestris* (Meng *et al.*, 2006). Whether *V. vinifera* became infected through co-evolution or through horizontal transmission by human interference via grafting remains unknown (Lima *et al.*, 2006a).

While the genome of GRSPaV isolates have been characterized to some extent, a challenge arises when studying this virus due to the fact that different viral sequence isolates differ by as much as 22.9%, with GRSPaV-SY being the most divergent (Meng & Gonsalves, 2007; Meng *et al.*, 2006; Habi *et al.*, 2006). A contributing factor for such great differences in nucleotide sequences between isolates is the lack of proofreading activity by the polymerase encoded in the viral genome, allowing a greater number of mutations to arise (Meng *et al.*, 2006; Terlizzi *et al.*, 2010). In addition, mixed viral infections resulting from grafting likely also play a role in larger changes in nucleotide sequences due to recombination events between viral isolates (Meng & Gonsalves, 2007; Meng *et al.*, 2006; Terlizzi *et al.*, 2010). Such great differences in sequences pose a technical challenge for detection of GRSPaV via reverse transcription-polymerase chain reaction (RT-PCR) (Terlizzi, 2011). Primers for broad-spectrum GRSPaV detection must target highly conserved regions within the genome, such as the HEL or CP. When looking for specific GRSPaV sequence variants of which the sequence is known, primers targeting non-conserved regions, such as the highly variable region (HVR) of ORF1 allow for variant-specific detection.

Sequence comparisons have been made between GRSPaV isolates, confirming genetic variability within the species. An investigation by Meng *et al.* (2005) of GRSPaV variant distribution in St George gave rise to GRSPaV-SG1, GRSPaV-SG2, GRSPaV-SG3 clusters. The
GRSPaV-BS variant was also first noted in this study, isolated from a ‘BS5563’ hybrid (Meng et al., 2005). The sequences from GRSPaV-SG1 and GRSPaV-BS were compared to the GRSPaV-1 sequence. The GRSPaV-BS sequenced differed most, having 83.9% and 84.3% nt similarities compared to GRSPaV-SG1 and GRSPaV-1, respectively, at the genome level. The genomic nt identity between GRSPaV-1 and GRSPaV-SG1 was found to be 87.3% (Meng et al., 2005).

Within ORF1, sequence similarity between the three sequence isolates ranged from 85.0-86.5% at the nt level and 91.6-92.7% at the aa level (Meng et al., 2005). These sequences gave some of the first clear evidence that GRSPaV is divided into different genetic lineages or groups. In a study in Italy, primer pairs RSP13 and RSP14 targeting a region of the HEL, and RSP2 and RSP21 targeting a region of the CP gene were used to investigate sequence variance in 17 Italian GRSPaV isolates. It was determined that RSP13 and RSP14 performed better for broad-spectrum GRSPaV detection. Sequence identity between isolates was found to range from 78.8% to 99.7%. At the amino acid level, these Italian CP sequences (259 aa) showed sequence identity ranging from 91.1 to 99.6%, suggesting CP function is likely conserved (Terlizzi et al., 2010).

These isolates clustered into four phylogenetic groups (phylogroups), or lineages. Nucleotide identity of isolates within phylogroups compared to their respective reference sequences are summarized in Table 1. Terlizzi et al. (2010) analyzed the nucleotide sequence identities between all isolates available in an online database. A summary of the ranges of nucleotide identity as a percent is included in Table 1. They found that phylogroups I (reference isolate GRSPaV-GG JQ922417) and II (reference isolate GRSPaV-SG1 AY881626) shared high nucleotide identity, suggesting these isolates are more closely related. Groups II and III were most distantly related with a minimum identity of 76.4% (Terlizzi et al., 2010). Group IV
(reference strain GRSPaV-SY AY368590) sequences had a range of 78.3-86% identity with a group I reference isolate, 78.1-87.2% identity with group II, and 80.4-84% identity to group III.

As reliable detection methods for GRSPaV become more available, genomic sequence comparisons become more informative. With the available information on the 15 fully sequenced GRSPaV isolate genomes, detected strains can be used in phylogenetic studies in order to determine which isolates are most related. While biological indexing with *V. rupestris* St George has been carried out for a long time, it is a process which takes years to detect GRSPaV and only detects those isolates that induce RSP on St George. In fact, it was determined that the majority of “RSP-negative” control St George clones were actually positive for certain variants of GRSPaV, as confirmed by RT-PCR and Western blotting (Meng *et al.*, 2000; Meng *et al.*, 2003; Habili *et al.*, 2006). The same findings were found in Italian St George clones (Meng *et al.*, 2005; Minafra *et al.*, 2000). With the availability of genomic sequence information, many primers have been designed to target broad-spectrum or isolate-specific GRSPaV sequences. Thus, RT-PCR can be used to obtain information about GRSPaV infection in grapevines significantly faster and more specific than biological indexing. Using sequence analysis studies, individual research groups have established that there are at least five groups of GRSPaV sequence variants, however it is likely that viral variants outside of these groups exist (Meng & Gonsalves, 2007; Martelli & Boudon-Padieu, 2006; Nolasco, *et al.*, 2006; Santos *et al.*, 2003; Casati *et al.*, 2003). Figure 5 shows the results of phylogenetic analysis of GRSPaV sequences from Italy, Canada, and the United States. They form distinct groups, and no geographical relationship. Group designation names differ in literature, so for the purpose of consistency, this thesis uses the most common and informative designations (alternate names in parentheses): -SG1 (2a, II), -PN (no previous designations), -BS (3, III), -SY (1, IV, -VS), -GG (2b, -1, I).
There is a lack of separation on the basis of geographical region and GRSPaV variants reported by any research group, suggesting potential evolutionary nucleotide changes in the genome likely did not result solely from environmental factors (Figure 5) (Meng et al., 2006, Meng & Gonsalves, 2007). However, the movement of GRSPaV worldwide has been aided by exchange of plant material between viticulturists. To uncover the existence of phylogenetic groups, different research groups have used broad-spectrum primers targeting conserved regions of the GRSPaV genome to obtain GRSPaV sequence clones for analysis. Meng et al. (2006) conducted analysis using primers targeting regions of both the CP gene and the helicase domain of GRSPaV from Niagara, Kober 5BB, Paulson 1103, and Millerdet 101-14 varieties. Primers RSP13-RSP14 amplify a 339 bp fragment within the HEL domain (ORF1; nt 4373-4711) (Meng et al., 1999; Meng et al., 2006). These primers detect a broad spectrum of GRSPaV variants and the HEL serves as a good target region for phylogenetic analysis (Terlizzi et al., 2009). Primers RSP21-RSP22 amplify a 441 bp fragment in the central CP gene (ORF5; nt 7917-8357) (Meng et al., 2003; Meng et al., 2006). The CP gene is well conserved among GRSPaV sequence variants and makes a good target region for broad spectrum primers. More recently, degenerate primers targeting a conserved region of the RdRp have been used, RSP35-RSP36. These primers amplify a 476 bp fragment ranging from nucleotide positions 5705-6180 of the GRSPaV genome. This region was identified for amplification due to high conservation between five different GRSPaV variants (GRSPaV-1, NC001948; -SG1, AY881626; -BS, AY881627; -SY, AY368590; -PN, AY368172) and the corresponding region of Apple stem-pitting virus (ASPV; genus Foveavirus). RSP35 corresponds to part of motif II of the RdRp, with conserved residues ‘AK’, and RSP36 corresponds to ‘AGDDMC’ within motif VI of the RdRp (Figure 6) (Terlizzi et al., 2011). Motifs II and VI are found within the “palm” and “fingers” respectively, if the
RdRp is imagined as a right hand, with a thumb, palm, and fingers (Ferrero et al., 2015). Phylogenetic GRSPaV analysis using RSP35-RSP36 was compared to analysis using RSP13-RSP14 by Terlizzi et al. (2011). Their results suggest RSP35-RSP36 is comparable, if not more efficient, in the detection of multiple GRSPaV variants than RSP13-RSP14. RSP35-RSP36 allowed for detection of two new GRSPaV variants, -MT from Mueller Thurgau and –ML from Merlot, and a partial sequence of GRSPaV-ML was obtained. GRSPaV-ML identified on a nucleotide level most highly with GRSPaV-BS (92.2%) at the 3’ end (Terlizzi et al., 2011). Similar to previous findings, isolates obtained from both scions and rootstocks clustered into groups regardless of primer pair used (Figure 7) (Meng et al., 2006). The newly identified GRSPaV-MT and –ML sequences clustered into what Terlizzi et al. (2011) tentatively called groups VI and VII respectively. The authors note that these two groups cluster closest with GRSPaV-PN, but differed on a nt level by 13.2-15.8%, suggesting distinction from –PN.

Meng et al. (2006) found that scion varieties had GRSPaV isolates from more than one of these variant groups while the rootstock varieties each contained GRSPaV variants from only one group. With this information, relationships may be uncovered regarding symptomology of viral sequence isolates and may give some insight to the evolution of GRSPaV. Similar work previously conducted in different geographical regions support the existence of multiple GRSPaV lineages. Santos (2003) identified four groups of GRSPaV sequence variants in Portuguese grapevines using broad spectrum primers RSP52 and RSP53 to target a 780 bp region of the CP (nt 7709-8613) (Terlizzi et al., 2010). Data presented by Casati (2003) suggests GRSPaV sequence variants from Italy and California cluster into three groups. Terlizzi & Credi(2003) identified three GRSPaV sequence groups using single-strand conformation polymorphism and also found three GRSPaV groups when sequences were compared.
Figure 5: GRSPaV distribution of isolates from Italy, Canada, and the United States using primers targeting the (B) HEL and (C) CP. There is no clear segregation of clones with respect to their origin, suggesting no geographical influence on different GRSPaV genotypes. Adapted from Meng et al. (2006). The GRSPaV-1 lineage is represented by I; the –SG1 lineage is represented by II; the –BS lineage is represented by III; and group IV represents the –SY lineage.
Figure 6: Target sequence for primers RSP35 and RSP36 across ASPV (nt positions 5761-5821; 6178-6238) and 5 GRSPaV sequence isolates (nt positions 5704-5764; 6121-6181) representing five GRSPaV lineages are covered; -GG (-1), -SG1, -BS, -SY, and -PN. Adapted from Terlizzi et al. (2011). Yellow boxed nucleotides represent identical matches. R = A + G; V = G + A + C; Y = C + T.
**Figure 7:** GRSPaV distribution of Italian isolates using (A) primers RSP13 and RSP14 targeting the helicase, and (B) RSP35 and RSP36 targeting the polymerase. Variants of the GRSPaV–SY (green) and –BS (sea foam) lineages are prominent in both trees. GRSPaV–SG1 and -1 groups were more highly detected using the helicase primers. The GRSPaV–ML and –MT lineages are represented with red labels. Only the polymerase primers detected the–ML lineage. In both trees, no clones cluster closely with the GRSPaV-PN lineage (purple). The bar represents 0.1 and 0.05 nt substitutions/site, for (A) and (B), respectively. Adapted from Terlizzi et al. (2011).
As knowledge of GRSPaV and disease relationships becomes more complex, it is important that GRSPaV is not looked at as a single species, but one that is also complex. These distinct variant groups, or lineages, may give key information to unraveling GRSPaV disease implications. Previous research correlated the GRSPaV-1 with RSP (St George) and GVN (110R), GRSPaV–BS with RSP (110R), and the -SG1 and –GG lineages with GVN (Rawhani et al., 2000; Meng et al., 2005; Borgo et al., 2009). The –SY lineage is highly divergent from the other groups which poses questions about what makes this group unique.

1.7 GRSPaV phylogroup disease associations

The different phylogroups may have differences in pathogenicity and symptomology (Meng & Gonsalves, 2007). Variation in disease and symptom expression in GRSPaV-positive grapevines may be the result of mixtures of GRSPaV variants present in different grapevine species. However, other biotic or abiotic factors cannot yet be ruled out (Lima et al., 2009). In the study by Meng et al. (2005) it was determined that GRSPaV-SG1 did not bring on RSP symptoms when graft-inoculated on V. rupestris St George vines, however, GRSPaV-1 caused mild RSP symptoms (Meng & Gonsalves, 2007; Meng et al., 2005). Further, a study by Borgo et al. (2009) found that GRSPaV-1 is highly correlated with vein necrosis while GRSPaV-BS is highly correlated with RSP symptoms on Richter 110R rootstocks. These reports support the notion that different GRSPaV sequence variants may induce different symptoms when graft inoculated on Vitis species. Habili et al. (2006) used GRSPaV-SY replicase-specific primers Sy9F/Sy8R on Australian grapevines from two different vineyards. The 3 of 15 sources that indexed negative for GRSPaV also tested negative when using the GRSPaV-SY-specific primers, but tested positive when using primers RS48/RS49, targeting the CP gene. In fact, all 15 sources in their study, including the “healthy” St George control, tested positive with the CP
primers (Halibi et al., 2006). This study also found that all Syrah sources with a range of SD symptoms were positive for GRSPaV using both the GRSPaV-SY-specific primers and the CP primers, indicating variants of the GRSPaV-SY lineage were present in all Syrah sources tested. However, nucleotide sequences of clones obtained using the GRSPaV-SY-specific primers were only up to 65% similar to American isolates, suggesting the sequences may be derived from a different species altogether, and that more sequence information is needed (Habili et al., 2006). In addition to different genetic variants likely being involved in different grapevine disease symptoms, mixed infections with different viral species and other grapevine pathogens cannot be ruled out. The combination of different virus species responsible for the variability in diseases and symptoms is a logical speculation, as a similar phenomenon was observed in a related virus of the Alphaflexiviridae family, Potato virus X (PVX, genus Potexvirus). Milne (1998) noted that PVX can establish infection resulting in disease in plants including potatoes, peppers, and tomatoes, however the outcomes were significantly worse when also other viruses were present, particularly Potato virus Y.

In an attempt to better understand the correlation between SD and GRSPaV-SY, a full-length infectious cDNA clone (FLC) of a GRSPaV-SY sequence variant is being constructed to produce infectious transcripts. Future grapevine tissue infiltration studies with this FLC construct will allow for more conclusive evidence of a relationship between the GRSPaV-SY viral variant and SD symptoms in Syrah.

1.8 Full-length infectious clones

The use of full-length clones (FLCs) has become an important method of studying viral replication, infection, and disease. An infectious clone can be a cDNA copy of a RNA genome and stimulate infection in a host (Boyer & Haenni, 1994). A ssRNA viral genome can be used as
Table 1: Nucleotide identity between Italian GRSPaV isolates. The GRSPaV-1 lineage is represented by I; the –SG1 lineage is represented by II; the –BS lineage is represented by III; and group IV represents the –SY lineage. Adapted from Terlizzi et al. (2010).

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of isolates in group</th>
<th>Ref. Strain</th>
<th>Identity with reference strain (%)</th>
<th>Nucleotide identity between groups (All isolates in GenBank) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>2</td>
<td>-1</td>
<td>89.4-96.4</td>
<td>Group I</td>
</tr>
<tr>
<td>II</td>
<td>3</td>
<td>-SG1</td>
<td>89.4-96.4</td>
<td>94.1-99.7</td>
</tr>
<tr>
<td>III</td>
<td>7</td>
<td>-BS</td>
<td>93.5-96.3</td>
<td>87.7-93.6</td>
</tr>
<tr>
<td>IV</td>
<td>4</td>
<td>-VS (-SY)</td>
<td>90.3-97.6</td>
<td>94.1-99.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>91.0-97.1</td>
<td>77.2-84.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>78.3-86.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>II</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89.1-99.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>76.4-82.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>78.1-87.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>III</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>86.4-98.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80.4-84.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IV</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>90.0-100.0</td>
</tr>
</tbody>
</table>
a template for cDNA synthesis, producing cDNA corresponding to the entire viral genome. This cDNA genome is cloned between the left and right flanking border sequences of a binary vector for integration of the cDNA sequence into the host plant nucleus and/or genome for transcription. Viral RNA promoters like the CaMV 35S promoter enable host machinery to transcribe the cDNA sequence to an infectious RNA transcript, which is then translated by host machinery, virions are formed, and infection ensues.

FLCs are incredibly useful for positive sense ssRNA viruses, which are often difficult to obtain in high titre or too unstable for excessive laboratory handling. For the particular case of nonretroviral, positive-sense ssRNA viruses, the lack of DNA intermediate during genome replication makes acquisition of a stable genomic sequence very difficult (Boyer & Haenni, 1994). The small nature of viral genomes allows for RT-PCR amplification of DNA fragments representing the entire viral genome. Natural restriction sites flanking or within these fragments allows for cloning of viral cDNA in sequential order without disrupting the sequence. A constitutive viral RNA promoter, in this case the 35S CaMV promoter, is cloned immediately upstream of the 5’ end of the viral cDNA, and nonviral nucleotides are deleted via site-directed mutagenesis (SDM). Alternatively, some shorter promoter sequences like T7 and SP6 promoters can be included in the 5’ genomic forward primer to insert the sequence immediately upstream of the first genomic nucleotide (Boyer & Haenni, 1994). It has been found that any nonviral nucleotides following the promoter sequence strongly reduce infectivity of viral transcripts. Another major limitation of FLC synthesis is obtaining the first full-length cDNA strand from the vRNA. This is likely in part due to strong vRNA secondary structures hindering the reverse transcriptase polymerization. It is noteworthy to mention that different viral variants or strains likely behave differently in regards to vRNA secondary structure, regardless, use of thermostable
reverse transcriptases and DNA polymerases are important for improved yield (Boyer & Haenni, 1994). In addition, the cloning vector and *E. coli* strain are factors to consider closely, as cDNA clones can be highly unstable in bacteria and may be subject to mutations or degradation. With the aforementioned under careful consideration, infectious cDNAs are incredibly advantageous in virology and the implications of this technique are limitless.

1.9 FLCs in literature

Meng *et al.* (2013) successfully created a FLC of GRSPaV-GG and a modified GFP-tagged variant. Using double-stranded RNA from *V. riparia* as a template, RT-PCR products corresponding to the 5’ and 3’ halves of the viral genome were separately cloned into the TA vector, pGEM-T. The 5’ fragment was digested out using unique restriction sites and cloned into pBluescript KS. Following this, the region correlating to the 3’ half of the viral genome was cloned into the resultant vector immediately after the 5’ fragment. The entire region corresponding to the full GRSPaV-GG genome was subcloned into pHST40 – a cloning vector that allows for the insertion of plant viral cDNAs between two restriction sites, cleavable by the HDV ribozyme sequence in the vector. The genome of HDV encodes both genomic and antigenomic sequences of this ribozyme, capable of self-cleave at the base of its complex pseudoknot catalytic fold. Flanking sequences up- and downstream of the ribozyme sequence self-pair into folded ends (Chadalavada *et al.*, 2007; Ferre-D'Amare *et al.*, 1998; Ke *et al.*, 2004). Mutagenesis PCR was used to remove non-viral sequences upstream of the GRSPaV-GG corresponding region and amplify the FLC DNA, which was subsequently subcloned into the binary vector, pCAMBIA. This vector clone can replicate in both *Escherichia coli* and *Agrobacterium tumefaciens*, and can be used in agro-infection of grapevine plantlets and the herbaceous model plant *Nicotiana benthamiana* (Meng *et al.*, 2013). These constructs were
proven to be infectious in *N. benthamiana*, supporting the notion that FLCs are a reasonable method for the study of GRSPaV (Meng et al., 2013).

As infectious clones become a more informative method of studying RNA viruses, their implications on GRSPaV do not fall short. cDNA infectious clones play a major role in etiological studies, as they can be used to infiltrate healthy host plants to induce or observe viral infection. Edwards and Weiland (2010) created a full-length infectious cDNA clone of Oat blue dwarf virus (OBDV) (genus *Marafivirus*). The resultant cDNA clone proved to be infectious in both maize and oat, inducing typical symptoms of OBDV infection in oat (Edwards & Weiland, 2010). A system in which expression of a ssRNA plant virus can be induced, observed and monitored is an irreplaceable tool to exploit for the purposes of virology. Subsequently, observations of plant disease symptoms can be monitored, and correlations can be made between viral isolates and host symptoms. Flatken et al. (2008) constructed a FLC of PVM (family *Betaflexiviridae*, genus *Carlavirus*) by ligating RT-PCR fragments of 5.5 kB and 3.2 kB corresponding to the entire PVM genome just downstream of the CaMV 35S promoter sequence in a binary vector. This group used the PVM FLC to replicate systemic infection in different host cultivars, indistinguishable from the wildtype PVM infection (Flatken et al., 2008). Vives et al. (2008) used agro-inoculation of citrus plants, *N. benthamiana*, and *N. occidentalis* with a FLC of Citrus leaf blotch virus (CLBV; family *Betaflexiviridae*, genus *Citrivirus*). Their results show that infection resultant of the CLBV FLC mimicked that of wildtype replication, movement, and pathogenicity. This FLC was used to confirm that CLBV causes Dweet mottle disease in citrus, and rule out that this virus is the causal agent of bud union crease syndrome in citrus (Vives et al., 2008). In another study, Boyapalle et al. (2008) constructed a FLC of Rhopalosiphum padi virus (RhPV; family *Dicistroviridae*, genus *Cripavirus*). RhPV infects at least seven species of
aphid, insects which are an important factor in grapevine virus transmission. The FLC was constructed in hopes to eventually understand unique genetic moieties like internal ribosomal entry sites found in the RhPV genome, determining how Discistroviridae infect hosts including honey bees, and to be used as potential aphid pest control. The infections initiated by the RhPV FLC exhibited the same infectious properties as infection with wildtype RhPV infection in aphids (Boyapalle et al., 2008). It is important that FLCs are tested in their native host before making any major conclusions about symptomology. Spetz et al. (2008) created a FLC of Poinsettia mosaic virus (family Tymoviridae, unassigned genus) which was found to be infectious in N. benthamiana but not in the natural host, poinsettia.

While infection with cDNA clones alone does not necessarily fulfill Koch’s postulates in the strict sense, it gives useful insight to diseases associated with viruses and their interactions with host plants. The use of a cDNA clone of GRSPaV would make a model system for the Betaflexiviridae family which is of interest as the understanding of RNA viruses which infect woody hosts is minimal (Rebelo et al., 2008). As described, the GRSPaV genome encodes most of the elements represented in members of Betaflexiviridae plus a number of domains absent in most other RNA viruses, such as the P-Pro, O-Pro, and AlkB domains which are of interest due to their uncertain functions. GRSPaV has also been highly associated with graft incompatibility, which is clearly a severe issue among many hosts of Betaflexiviridae viruses (King et al., 2011a). Further, cDNA clones have great implications in viral gene probing studies. Forment et al. (2005) describe a method of using cDNA clones in a microarray in order to conduct genomic probing studies. While these are not genome-length infectious cDNA clones of an RNA virus, it is an innovative example of the use of cDNA clones in genomic probing and potentially playing a role in diagnosis of grapevine viruses.
1.10 Rationale

Grapevines maintain great economic importance, but unfortunately experience many diseases caused or influenced by a large number of grapevine viruses. Many of these diseases have become widespread throughout the world as a result of contaminated grafting tools and the exchange of grapevine materials between viticulturists and grafting between scions and rootstocks. GRSPaV has been highly associated with RSP, GVN, and SD; however the exact roles of this virus in the diseases are still poorly understood. SD has a particularly high association with GRSPaV as studies have suggested all symptomatic and many non-symptomatic grapevines are positive for GRSPaV. This poses the idea that the genetic variants of GRSPaV may have differences in pathogenicity. The GRSPaV-SY lineage is genetically divergent from other lineages and may be a key factor in the onset of SD symptoms. GRSPaV is a key member of the Betaflexiviridae family of positive-sense, ssRNA viruses, most of which infect woody fruit crops. A greater understanding of the life cycle and pathogenesis of GRSPaV will aid in disease control by members of Flexiviridae.

The purpose of this study was to investigate the genetic diversity of GRSPaV in Ontarian vineyards and to identify potential correlations between GRSPaV sequence variants and SD symptoms. Broad-spectrum degenerate primers RSP35-RSP36 target a 475 bp region of the GRSPaV RdRp. As described previously, these primers target a conserved region of ORF1, and detect a wide range of GRSPaV variants in different phylogenetic groups (Terlizzi et al., 2011). The resulting 476 bp amplicon is a small enough region to be easily amplified, cloned, and sequenced, but is large enough to give distinction between phylogroups. Previous studies used small conserved gene regions (ORF1 RSP13-RSP14 339bp; ORF5 RSP21-RSP22 441 bp) for GRSPaV diversity and found distinctive phylogroups (Meng et al., 2006; Terlizzi et al., 2010).
The unique and severe symptoms experienced by Syrah cultivars with SD and the high instance of GRSPaV presence in SD plants poses a question about the link between the two. The GRSPaV-SY lineage is highly divergent and detected in the vast majority of declining Syrah vines. It is believed that different GRSPaV sequence variants are involved in different diseases and symptoms. The work presented here investigates the possible correlation between SD red canopy in declining Syrah cultivars and GRSPaV.

1.1 Hypotheses

It was hypothesized that GRSPaV variants in Ontario are diverse, composed of at least four distinct genetic lineages. Further, it was hypothesized that GRSPaV-SY is correlates to for SD symptoms in Syrah either alone or in combination with other agents.

1.2 Objectives

To investigate this, three objectives were set; (1) to determine the genetic diversity of GRSPaV in Ontarian vineyards; (2) to find correlations between SD and GRSPaV sequence variants; and (3) to construct a full-length infectious cDNA clone of a GRSPaV-SY sequence variant for downstream confirmation of SD and GRSPaV-SY correlations.

Samples from a spectrum of grapevine varieties and cultivars with varying grapevine disease symptoms (GLR, GVN, SD, no observable symptoms) were collected by members of our lab between May 2015 and July 2015 for this study. Tissue from grapevine leaves, stems, petioles, and fruits were collected from different vineyards in the Niagara region of Ontario for viral analysis in our lab but for this study only leaf tissue was used. Objective 1 involved RNA extractions from source leaf tissue, followed by RT-PCR, cloning and sequencing of RSP35-
RSP36 amplicons and phylogenetic analysis of resultant GRSPaV sequence data. Data for objective 2 was a subset of the total data collected in objective 1. The red canopy was the primary symptom used to designate severely declining Syrah. The comparative group, “Syrah without red canopy”, cannot be designated “SD-negative” because of the late onset of symptoms and the lack of individual vine data on stem pitting symptoms. We were not able to remove bark near the graft union to look for the presence of pitting. To choose grapevine source material for synthesis of the GRSPaV-SY cDNA clone, data from objective 1 was used to find a source which produced a relatively high percentage of GRSPaV-SY variants with primers pairs RSP35-RSP36, and RSP35-SY – a GRSPaV-SY specific-RSP35 primer, targeting the exact same region – and RSP36 (Table 6, Appendix I). A cultivar with sequences highly identical (94-98%) to the GRSPaV-SY reference sequence (AY368590) was used as source material, to aid specificity to FLC primers, designed according to this sequence. To initiate transcription at the true 5’ nucleotide of the GRSPaV viral genome once inside the plant nucleus, the constitutive Cauliflower mosaic virus (CaMV) 35S promoter was chosen to drive transcription of viral RNA. The nopaline synthase (nos) polyA signal sequence at the 3’ end of the viral genome was chosen for polyadenylation of the viral transcript to terminate transcription. The antigenomic Hepatitis delta virus (HDV) ribozyme (Rz) sequence allows for cleavage of the viral transcript ends to resemble the 3’ end of an RNA virus.
CHAPTER 2: MATERIALS AND METHODS

2.1 GRSPaV Diversity

The GRSPaV diversity study was completed by repeating a few simple protocols for each grapevine source. First, total plant RNA was extracted, followed by RNA purification. Reverse transcriptase (RT) polymerase chain reaction (PCR) was used to amplify target regions within the RdRp and the CP genes. Resultant amplicons were A-tailed and ligated into a TA-cloning vector, pGEM-T-Easy. E. coli DH5α competent cells were used for transformation of resulting plasmid constructs. After purification, these plasmid constructs containing the desired region of interest were sequenced and sequence and phylogenetic analysis was carried out as described below.

2.1.1 Plant Total RNA extraction

Total plant RNA was extracted using Spectrum™ Plant Total RNA kit (Sigma-Aldrich, St. Louis, Missouri, USA) following the manufacturer’s instructions with slight modification to the lysis buffer. The recommended β-mercaptoethanol was included in addition to 2.5% PVP-40 and vortexed thoroughly. 800 µL of this solution was added to 50 µg of grapevine leaf tissue and combined using a mortar and pestle. This mixture was transferred to a microfuge tube and heated between 65-80°C for 5 minutes. The mixture was then centrifuged to pellet the plant debris and the kit was used according to manufacturer’s instructions. Total RNA concentrations were checked using a Nanodrop™ 1000 (Thermo Scientific, Waltham, Massachusetts, USA).

2.1.2 RT-PCR of RdRp fragments for sequencing

Reverse transcriptase (RT) cDNA was obtained using up to 2 µg of total RNA and 10 µMol random primers in Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT)
(Promega, Fitchburg, Wisconsin, USA) reactions. The protocol was carried out according to manufacturer’s instructions with an extension incubation at 37°C for 2 hours. RT products were stored at -20°C.

The target fragment was amplified via PCR using random primer RT cDNA as a template and primers RSP35 and RSP36. These primers target genomic nucleotide positions 5705-6180 providing a resultant fragment of 475 bp. PCR was carried out using 1 U Taq polymerase in a total reaction volume of 25 µL. PCR cycling conditions were as follows; 95°C for 5 minutes for initial denaturation, followed by 95°C for 30 seconds, 50°C for 30 seconds to allow primer annealing, and 72°C for 45 seconds for elongation, repeated for 30 cycles, followed by a final elongation at 72°C for 7 minutes.

2.1.3 Cloning of RdRp PCR products into pGEM-T-Easy

PCR products were cloned directly into the pGEM-T-Easy vector (Promega) according to manufacturer’s instructions. Ligation reaction mixtures were incubated overnight at 4°C. 2 µL of each ligation reaction was used in transformation of competent E. coli DH5α using the heat shock method. Competent cells were thawed on ice and 50 µL were mixed gently with the ligation reaction product and incubated on ice for 20-40 minutes. Heat shocking of cells was performed in a 42°C water bath for 48 seconds and cells were immediately returned to ice for 2 minutes. 950 µL of Luria-Bertani (LB) broth was added to each tube of cell mixture and incubated at 37°C at 150 rpm for 1.5-2 hours. Cells were plated on LB agar containing 100 µg/mL ampicillin and spread with 100 µL of 100 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 20 µL of 50 mg/mL X-gal prior to incubation at 37°C overnight.
Resultant white colonies, indicative of a successful ligation, were screened using colony PCR for the presence of the RdRp fragment of interest. Using Taq polymerase in a total reaction volume of 25 µL, primers RSP35 and RSP36 were used for detection of the fragment of interest in colonies used as templates. PCR products were run at 100 V on a 1.2% agarose gel via agarose gel electrophoresis (AGE) and detection of DNA bands was carried out by soaking the gel for 15-20 minutes in 0.5 µg/mL ethidium bromide solution and visualizing using UV light.

Positive colonies were then grown overnight at 37°C and 250 rpm in 5 mL LB broth containing 100 µg/mL ampicillin. 1.5 mL of resulting overnight culture was used for plasmid extraction using the BioBasic® Plasmid Miniprep kit (BioBasic, Toronto, Ontario, Canada) according to the manufacturer’s protocol.

2.1.4 Sanger sequencing of plasmid DNA

Approximately 10 positive clones per grapevine source were selected for sequencing. 130 ng of plasmid DNA diluted in a total volume of 3 µL were combined with 1 µL of BigDye® Terminator v3.1 Ready Reaction Mix (Thermo Fisher Scientific), 2 µL of 5X BigDye® Sequencing Buffer (Thermo Fisher Scientific), 1 µL 10 mM M13F primer, and 8 µL of water for a total reaction volume of 15 µL. Reaction vessels underwent thermocycling with an initial denaturation at 95°C for 2 minutes, 30 cycles of 95°C for 30 seconds, 50°C for 15 seconds for annealing, 70°C for 45 seconds elongation, followed by a final elongation step at 70°C for 5 minutes. Reaction products were then sent to the University of Guelph’s Advanced Analysis Centre Genomics Facility to determine sequence outcome.
2.1.5 Phylogenetic analysis of sequenced fragments

Initial sequence outcomes were subject to NCBI’s Basic Local Alignment Search Tool (BLASTn) to determine that nucleotide sequences were indeed derived from GRSPaV and to give some indication of which GRSPaV variant was detected. Using DNASTAR bioinformatics software, each sequence was oriented and aligned by MegAlign ClustalW and the data was presented in a tree format for preliminary results. Final phylogenetic analysis was carried out kindly by Clayton Moore. Sequences were aligned with ClustalW using Molecular Evolutionary Genetics Analysis version 7.0 (MEGA) software using both maximum likelihood and neighbour joining algorithms with 1000 bootstrapping replicates (Kumar et al., 2016; Saitou & Nei, 1987). Reference sequences were from genome positions 5704-6180 nt for each of the following GRSPaV variants from the NCBI gene bank; GRSPaV (AF026278), GRSPaV-1 (NC_001948), GRSPaV-PN (AY368172), GRSPaV-BS (AY881627), GRSPaV-SG1 (AY881626), GRSPaV-SY (AY368590), GRSPaV-MG (FR691076), GRSPaV-GG (JQ922417), GRSPaV-PG (HE591388), GRSPaV-3138-07 (JX559646), GRSPaV-JF (KR054734), GRSPaV-VF1 (KT948710), GRSPaV-TannatRspav1 (KR528585), GRSPaV-LSL (KR054735), and GRSPaV-WA (KC427107). The corresponding region from ASPV was included as an outgroup from within the same genus, Foveavirus. Phylogenetic trees were formatted using Dendroscope v3.5.7 [January 2016].

2.1.6 CP diversity

In order to give more significance to the groups found in the phylogenetic trees, a small-scale diversity study targeting a region of the CP was conducted. For this study, selections of sources used in the RdRp diversity analysis were chosen for repeat experimentation, only using
primers targeting the CP. The data obtained from the RdRp diversity study was used to find suitable source material for the CP diversity study. Primers designed to specifically target the very beginning of the CP gene (nt 7770) in six GRSPaV subgroups were combined in a mixture with a single reverse primer. Subgroups targeted were GRSPaV-SY, -PN, -BS, -SG1, -GG and –MG. The reverse primer, RSP8634R, was designed previously in our lab and targets a highly conserved region near the end of the CP gene. The individual forward primers paired with the sole reverse primer were initially tested to find a temperature at which each set of primers could successfully produce amplicons, and an annealing temperature of 50°C was most chosen and the extension time was increased to 1 minute at 72°C. PCR amplicons were separated via AGE with 1% and 0.8% agarose. The band at 864 bp was excised, purified using the Wizard SV Gel and PCR Clean-Up System (Promega) and the purified product was ligated to the pGEM-T-Easy cloning vector in an approximate 1:1 insert:vector ratio using 5 U T4 DNA ligase (Promega) and incubated at 4°C overnight. E. coli DH5α competent cells were transformed using the heat-shock method as described above, and grown on LB agar with 100 mg/mL ampicillin, 100 µL of 100 mM IPTG and 20 µL of 50 mg/mL X-gal at 37°C for 20 hours. White colonies picked and cultured in sterile 5 mL liquid LB with 100 mg/mL ampicillin at 37°C and shaking at 250 rpm. Plasmid DNA was purified using the BioBasic Miniprep Kit (BioBasic). Plasmids were screened by PCR using M13F and M13R, which flank the pGEM-T-Easy MCS, to screen for the correct insert size. This was in lieu of using the CP primer mix for pre-sequencing screening, which would likely result in a lot of non-specific amplification due to the high conservation of this region. A total of 12 clones were sequenced using M13F and M13R as described previously.
2.2 Full-Length Clone

2.2.1 Primers

In order to clone the CaMV 35S promoter sequence into the pBluescript KS II+ vector, primers were designed to target the 35S sequence in the pHST40 vector sequence. Restriction sites were added to the 5’ end (NotI) of the forward primer and 3’ end (XbaI) of the reverse primer to allow for downstream cloning via restriction digest without disrupting the promoter sequence. The forward primer was designated NotI35S-F and the reverse designated Xba35S-R (Table 6).

Primers were designed targeting natural unique restriction sites within the GRSPaV-SY (AY368590) sequence, that were also available in sequential order in both pBluescript II KS+ and pCB301.3. The original design for obtaining a full length clone from GRSPaV-SY genome fragments is depicted in Figure 8. Fragments were designed as follows: F1 1-445 (445 bp), F2.1 433-4410 (3977 bp), F3 4399-7584 (3193 bp), and F4 7581-8725 (1191 bp). In the final design, F2 contains the entire F1 fragment, using the F1 forward primer (Figure 9). The predicted fragments correspond with the following nucleotide positions; F2a 1-2127 (expected size 2127 bp); F2b 1659-4410 (2751 bp); F3a 4399-6074 (1675 bp); F3b 5938-7592 (1654 bp); and F4 7581-8725 (1191 bp).

2.2.2 PCR Amplification and Cloning of CaMV 35S Promoter

Primers NotI35S-F and Xba35S-R were used to amplify the 318 bp 35S CaMV promoter sequence. PCR was carried out in a total reaction volume of 50 µL with KOD 10x Hot Start Buffer, 10 µM of each primer, 2 mM dNTPs, 1 U KOD Hot Start Polymerase (Merrick
Millipore, Billerica, Massachusetts, USA), and 20 ng pHST40 plasmid as the template. Thermocycler conditions were as follows; 95°C for 3 min initial denaturation, followed by 35 cycles of 95°C for 30 seconds, 61°C for 30 seconds, 70°C for 30 seconds, and a final elongation of 70°C for 7 min. PCR products were run on a 0.8% agarose gel at 100 V for approximately 50 minutes. The expected band of 318 bp was excised from the gel and purified using the Wizard SV Gel and PCR Clean-Up System (Promega). The resultant purified PCR product was subjected to digestion by NotI and XbaI in a 50 µL reaction containing 20 U of each restriction enzyme and 1333 ng of DNA and NEBuffer 3.1 (New England Biolabs). The reaction was incubated at 37°C for 1.5 hours and analyzed on a 0.8% agarose gel run at 100 V for approximately an hour before staining with 0.05 µg/mL EtBr for 15 minutes. The expected band was further excised from the gel and purified as described previously. 2 µg of the pBluescript II KS+ vector was digested with 10 U each of XbaI and NotI and incubated at 37°C for 2 hours. The digested vector product was analyzed on a 0.8% agarose gel and excised and purified as the 35S CaMV fragment. The digested vector sequence was then dephosphorylated using 5 U Antarctic phosphatase with buffer, and incubated in a 37°C water bath for 25 minutes, followed by enzyme deactivation on a 70°C heating block for 10 minutes.

The NotI-XbaI digested vector and insert fragments were ligated together in a 3:1 ratio (insert:vector) in a 15 µL reaction using 5 U T4 DNA ligase (Thermo Fisher Scientific) at 4°C for 20 hours. From the ligation reaction, 2 µL were used to transform E. coli DH5α competent cells using the heat-shock method. Cells were plated on LB Agar with 100 µg/mL ampicillin and grown for 14 hours at 37°C. Colony PCR was used to confirm the presence of the 35S promoter in six resultant colonies. A colony was chosen for overnight culturing in liquid LB and the plasmid DNA was extracted using the BioBasic Miniprep Kit (BioBasic).
Figure 8: GRSPaV-SY FLC Restriction Map – Original Plan. F1 (0.4 kb), F2.1 (4 kb), F3 (3.2 kb), and F4 (1.1 kb) fragments are depicted below the genome diagram from 5’ to 3’ using dotted lines. The complete FLC including promoter and terminator sequences (9.4 kb) is flanked by NotI35S-F and KpnNos-R, as depicted at the bottom of the figure.
Figure 9: GRSPaV-SY FLC Restriction Map – Final Design. F2a (2.1 kb), F2b (2.7 kb), F3a (1.6 kb), F3b (1.6 kb), and F4 (1.1 kb) fragments are depicted below the genome diagram, from 5’ to 3’ using dotted lines. The full-length F2 (4.4 kb) and F3 (3.2 kb) fragments are represented below their respective overlapping fragments. Finally, the entire FLC genome including the 35S CaMV promoter and the HDVnos terminator sequences are represented at the bottom.
2.2.3 RNA Extraction and RT-PCR of FLC Fragments

Plant material from which to isolate RNA template for downstream RT-PCR was chosen based on findings from the GRSPaV diversity survey. Chosen was a Cabernet Franc cultivar (vine 50-10) of unknown scion and rootstock, expressing red blotch and leafroll symptoms. This vine, a source for the GRSPaV diversity study, was originally chosen as the source material for the FLC due to its relatively high instance of detection of GRSPaV-SY variants with both RSP35 and RSP36 and RSP35-SY and RSP36. Sequences obtained from the survey of 50-10 resulted in clones up to 99% identical to the GenBank sequence, using RSP35-SY specific forward primer, and up to 97% using RSP35 non-specific primer. Preliminary RT-PCR troubleshooting was carried out with 50-10 RNA. Vine 93-5, a Syrah clone 1 expressing red canopy was chosen for the final FLC source material. More successful PCR products were obtained from this source during FLC PCR troubleshooting (Chapter 4). GRSPaV-SY clones obtained using RSP35 and RSP36 from vine 93-5 were up to 94% identical to the GRSPaV-SY GenBank sequence.

Plant total RNA was extracted using the Spectrum™ Plant Total RNA kit (Sigma-Aldrich) with modified PVP buffer as described. The incubation step following the lysis was 80°C for 7 minutes to allow for full denaturation of RNA. RT reactions were carried out using the SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) protocol. Reverse primers for each FLC fragment were used in individual RT reactions with 93-5 total RNA template. RT reactions were incubated at 55°C for 3 hours to allow for extension of long cDNAs, followed by incubation at 70°C for 15 mins.

PCR was carried out using the KOD Hot Start DNA polymerase kit (Merck Millipore, Billerica, Massachusetts, USA). KOD Hot Start 10x Buffer, 1 mM dNTPs, 10 µM of each
primer, and 1 U KOD polymerase was combined with water to a total volume of 25 µL. 1 µL each RT reaction was used as a template. Denaturation occurred at 95°C and extension at 70°C for all reactions. Optimal annealing temperature and expected amplicon length are described for each GRSPaV-SY genome primer pair in Table 2.

2.2.4 Cloning FLC fragments into pGEM-T-Easy

Purified PCR products representing each fragment; F2a, F2b, F3a, F3b, F4, were eluted in nuclease-free water and spun with heating to evaporate any residual ethanol. These products were then subjected to polyA-tailing by Taq polymerase. In a total of 15 µL, Taq buffer, 2 mM dATPs, 1 U Taq polymerase and purified PCR product were mixed and incubated at 72°C for 30 minutes. The resulting A-tailed products were ligated to pGEM-T-Easy vector (Promega) in a 1:1 or 3:1 insert:vector ratio using 5 U T4 DNA ligase (Promega) and incubated overnight at 4°C. Up to 3 µL of the ligation reaction were used to transform competent *E. coli* DH5α using the heat-shock method. Cells were incubated at 37°C with shaking at 250 rpm for 2 hours. Cells were gently pelleted at 3000 rpm for 6 minutes, resuspended in 100 µL of LB and plated on 1% LB agar plates with 100 mg/mL ampicillin and 100 µL of 100 mM IPTG and 20 µL of 50 mg/mL X-gal, and incubated at 37°C for approximately 20 hours. Resultant white colonies were screened via colony PCR and positive clones were used to inoculate 6 mL LB with 100 mg/ mL ampicillin and incubated at 37°C and shaking at 250 rpm for 12-14 hours. The EZ-10 Spin Column Plasmid DNA Miniprep Kit (BioBasic, Toronto, Ontario, Canada) or the Presto™ Mini Plasmid Kit (Geneaid, New Taipei City, Taiwan) was used to extract plasmid DNA from 1.5 mL of cell culture. Clones here are designated pGEM-F2a (nt 1-2127), pGEM-F2b (nt 1659-4410), pGEM-F3a (nt 4399-6074), pGEM-F3b (nt 5938-7592), pGEM-F4 (nt 7584-8725).
2.2.5 Overlap PCR of F2 and F3

Due to the large fragment sizes of the originally planned F2 and F3, 4410 nt and 3193 nt respectively, a decision was made to divide both fragments into two overlapping halves and use the overlapping region as a megaprimers for overlap PCR (Figure 8, Figure 9). For this method, amplification of a large genome fragment can be obtained by designing primers targeting a small region in the middle of the fragment. This is attainable as the genome sequence of each fragment is known. A forward primer targeting the 5’ end of the total fragment is paired with the reverse primer just downstream of the middle of the total fragment. Next, the forward primer just upstream of the middle region is paired with the reverse primer for the total fragment. These reactions create the two megaprimers template fragments, and span the length of the total desired fragment. In the overlap PCR, only the forward and reverse primers at the very 3’ and 5’ of the total fragment are used. Figure 10 is a depiction of the F2 overlap PCR. After annealing and extension from the megaprimers overlapping region, the forward and reverse primers can bind the 5’ and 3’ ends of the fragment (Higuchi et al., 1988).

Primers RSP-SY2414F and RSP-SY2481R were designed to produce an overlapping region of 67 nt (Figure 22). The successful overlap fragments for F2 used HVR primers RSP-SY1659F and RSP-SY2127R, creating an overlapping region of 468 bp. Primers RSP-SY5938F and RSP-SY6074R created an overlapping region of 136 bp for F3 (Figure 11). These new primers allowed for amplification of shorter fragments; F2a 2127 nt and F2b 2751 nt; F3a 1675 nt and F3b 1654 nt. F2a and F2b act as megaprimers templates for overlap PCR of F2, and likewise for the F3 overlap. The PCR products for each were run on a 0.8% agarose gel and fragments were excised and purified using the Wizard SV Gel Clean-Up Kit (Promega) as described previously. The F2a and F2b fragments were used as templates for the overlap PCR of
Table 2: GRSPaV-SY genome FLC fragment primer pairs and optimal annealing temperatures, and expected fragment lengths. See Table 6 for primer sequences.

<table>
<thead>
<tr>
<th>Fragment Designation</th>
<th>GRSPaV-SY Genome Region (nt)</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>Annealing Temp. (°C)</th>
<th>Expected Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>35S</td>
<td>35S CaMV Promoter</td>
<td>NotI35S-F</td>
<td>Xba35S-R</td>
<td>61</td>
<td>318</td>
</tr>
<tr>
<td>F2a</td>
<td>1-2127</td>
<td>RSP-SY1F(Xba)</td>
<td>RSP-SY2127R</td>
<td>54</td>
<td>2127</td>
</tr>
<tr>
<td>F2b</td>
<td>1659-4410</td>
<td>RSP-SY1659F</td>
<td>RSP-SY4410R(Cla)</td>
<td>54</td>
<td>2751</td>
</tr>
<tr>
<td>F2</td>
<td>1-4410</td>
<td>RSP-SY1F(Xba)</td>
<td>RSP-SY4410(Cla)</td>
<td>54</td>
<td>4410</td>
</tr>
<tr>
<td>F3a</td>
<td>4399-6074</td>
<td>RSP-SY4399F(Cla)</td>
<td>RSP-SY6074R</td>
<td>52</td>
<td>1675</td>
</tr>
<tr>
<td>F3b</td>
<td>5938-7592</td>
<td>RSP-SY5938F</td>
<td>RSP-SY7592R(Sal)</td>
<td>52</td>
<td>1654</td>
</tr>
<tr>
<td>F3</td>
<td>4399-7592</td>
<td>RSP-SY4399F(Cla)</td>
<td>RSP-SY7592R(Sal)</td>
<td>52</td>
<td>3193</td>
</tr>
<tr>
<td>F4</td>
<td>7584-8725</td>
<td>RSP-SY7584F(Sal)</td>
<td>RSP-SY8725R(Kpn)</td>
<td>50</td>
<td>1191</td>
</tr>
</tbody>
</table>
the entire F2 fragment. Primers RSP-SY1F(Xba) and RSP-SY4410R(Cla) were used to target only full-length F2 overlap fragment of 4410 bp. The PCR product was run on an agarose gel and purified again using warm nuclease-free elution water, followed by A-tailing by Taq polymerase and cloning into pGEM-T-Easy to create pGEM-F2. F3a and F3b fragments were run on an agarose gel and purified as before. Primers RSP-SY4399F(Cla) and RSP-SY7592R(Sal) were used to amplify the entire 3193 bp F3 fragment via overlap of F3a and F3b. The F3 fragment was then A-tailed using Taq polymerase and cloned into pGEM-T-Easy to produce pGEM-F3. The 3’ genomic fragment, F4, was amplified using RT-PCR to produce the full 1191 bp fragment. Due to the smaller size of this fragment, no overlap PCR was required. The F4 fragment was A-tailed and cloned into pGEM-T-Easy to produce pGEM-F4.

2.2.6 Sanger sequencing of FLC fragments in pGEM-T-Easy

Sequences of the four FLC fragment pGEM-T-Easy constructs needed to be obtained to ensure they were most closely related to GRSPaV-SY (AY368590). Between two and four different clones of each construct were sequenced using Sanger sequencing described in section 2.1.4, with plasmid concentration was adjusted to accommodate the general guideline of 28 ng DNA per 1 kb plasmid. Primer annealing temperature was maintained at 50°C and extension time was increased to 2 minutes. Initial sequencing of these fragments using M13F and M13R provided confident sequences for about 700-800 nt each, leaving the sequence of the internal regions of these fragments unknown. One clone per construct was selected for downstream work, chosen based on closest identity to GRSPaV-SY and cover query. A complete sequence needed to be obtained for each pGEM-T-Easy construct; -F2a, -F2b, -F3, and -F4. The primers for sequencing the internal regions of F2a, F2b, F3a and F3b were designed based on the sequencing results.
1. PCR amplification of F2a and F2b from pGEM-T-Easy constructs using KOD polymerase. Products are purified and used as the templates for the F2 overlap PCR.

2. In the overlap PCR mix ds F2a and F2b denature, and ss fragments pair at the overlap region at the annealing step.

3. KOD polymerization from 3' ends of megaprimer to produce full-length F2 template strands.

4. Full-length F2 progeny are used as templates for extension from forward and reverse F2 primers.

**Figure 10:** Overlap PCR of the F2 fragment. F2a and F2b are amplified from pGEM-F2a and pGEM-F2b, respectively, and PCR products are purified. The overlap region (green box) acts as a megaprimer for extension of the full-length F2 templates. F2 flanking primers, RSP-SY1F(Xba) and RSP-SY4410-R(Cla) are used as forward and reverse primers, and KOD allows for long distance polymerization.
Figure 11: Primer map including primers used for internal overlap regions in F2; RSP-SY1659-F, RSP-SY2127-R and F3; RSP-SY5938-F, RSP-SY6074. The F1 fragment reverse primer, RSP-SY445R, is not included because the large F2 fragment was revised to use the very 5’ primer RSP-SY1F(Xba) as the forward primer, covering the small F1 region. Each of the newly designed internal F2 and F3 primers map to the corresponding regions within both the GRSPaV-SY sequence (Appendix I, Figure 21, Table 6). The newly obtained internal sequences were combined with the sequences obtained using M13F and M13R for each of the pGEM-T-Easy constructs (approximately 800 nt in each direction), producing the entirety of all fragment sequences prior to downstream cloning. The assembled genome has now been denoted GRSPaV-JH and is awaiting submission into NCBI GenBank.
2.2.7 Cloning F4 and F3 into pBS35S

The pGEM-F4 construct was digested with SalI and KpnI restriction enzymes to cut the 1191 bp fragment from the vector with the appropriate nucleotide overhangs for subcloning into pBS35S. The digest reaction product was analyzed via AGE and gel purified as described previously. The pBS35S backbone was also cut with SalI and KpnI and gel purified. The purified backbone product was dephosphorylated by Antarctic phosphatase (New England) in a 40 µL solution containing Antarctic phosphatase buffer incubated at 37°C for 35 minutes and the enzyme was inactivated by incubation at 70°C for 10 minutes. In an approximate 3:1 and 1:1 insert:backbone ratio, the F4 fragment was ligated to pBS35S by T4 DNA ligase (Promega) in a 20 µL reaction incubated at 4°C overnight. E. coli DH5α cells were transformed with 5 µL of the ligation product using the heat-shock method, plated on LB agar containing 100 mg/mL ampicillin, 100 µL of 100 mM IPTG, and 20 µL of 50 mg/mL X-gal, and incubated for 20 hours at 37°C. Using similar methods to the F4 cloning strategy, the pGEM-F3 construct was digested with ClaI and SalI restriction enzymes to cut the F3 fragment at natural restriction sites, to clone into the pBS-35S construct cut with the same enzymes.
CHAPTER 3: RESULTS

3.1 RdRp and CP diversity

Clones for diversity analysis were subjected to ClustalW alignment using MEGA and were subjected to phylogenetic analysis by both maximum likelihood and neighbour-joining algorithms. By both methods all tree data sets showed the same distribution or clustering patterns. The neighbour-joining trees are illustrated here, and all maximum likelihood trees are presented in Appendix II.

3.1.1 Genetic diversity of GRSPaV from *Vitis* varieties

Among all grapevines surveyed, four distinct clusters of GRSPaV lineages are apparent, corresponding with GRSPaV-SY, -PN, -SG1, and -GG lineages (Figure 12). Table 3 summarizes the distribution of clones in Figure 12. The –PN lineage is represented by two apparent subgroups, which share a common ancestor with the –PN reference isolate. The majority of clones (53 of 169) clustered with one of the –PN subgroups, having 93-96% nucleotide sequence similarity with the GRSPaV-VF1 reference sequence. Isolates within the –VF1 subgroup were 85-86% similar to the GRSPaV-PN reference sequence. Isolates of this subgroup were also 92-94% similar to the GRSPaV-ML variant identified by Terlizzi *et al.* (2011). The second –PN subgroup, with GRSPaV–JF as a reference isolate, has 35 of 169 clones. Isolates within the –JF subgroup were between 96-98% similar to the GRSPaV-JF reference sequence and 87-88% similar to the GRSPaV-PN reference sequence. Isolates within the –JF subgroup also identified closely with GRSPaV-RSP47-4 (EF105294.1), with 96-98% nt similarity. However, two clones, 107.9F and 107.9G from vine 93-5 Syrah clone 1 expressing red canopy, were only 91% similar to both the –JF and –RSP47-4 reference isolates, and cluster separately from the rest of the
group. In total, approximately half of all clones (88 of 169) clustered with the –JF and –VF1 subgroups. These clones were from Syrah clone 1 (vines 66-6, 93-4, 93-5; cDNA 79.6, 107.8, 107.9), Syrah clone 1 on 3309 rootstocks (vines 30-1, 66-11; cDNA 32.1, 79.11), Syrah clone 100 (vines 93-21, 93-22, 93-26, 93-28; cDNA 107.15, 107.16, 128.20, 107.20), Merlot clone 184 on 3309 rootstocks (vines 66-16, 66-17; cDNA 76.22, 76.23), Chardonnay (vine 49-9; cDNA 56.6), Cabernet Franc (vines 50-2, 50-9; cDNA 76.2, 76.7), Riesling (vines 49-15, 84-15; cDNA 56.8, 128.4).

The second largest cluster was the –SG1 lineage (39 of 169). Clones within the –SG1 lineage were 93-95% similar to the GRSPaV-SG1 reference isolate. The majority of the clones in the –SG1 lineage group clustered closely with the GRSPaV-MG reference isolate and were 96-97% similar to this reference sequence. Vines that were positive for variants of the –SG1 lineage were Syrah clone 1 (vines 66-6, 93-4; cDNA 79.6, 107.8) and clone 1 on a 3309 rootstock (vine 66-11; cDNA 79.11), Syrah clone 100 (vines 93-22, 93-26, 93-28; cDNA 107.16, 128.20, 107.20) and clone 100 on 3309 rootstocks (vine 30-1; cDNA 32.1), Merlot clone 184 on a 3309 rootstock (vine 66-18; cDNA 76.24), Cabernet Sauvignon (vine 50-20; cDNA 76.17), Cabernet Franc (vine 50-2; cDNA 76.2) and a Riesling (vine 84-15; cDNA 128.4).

The –SY lineage had a total of 28 of 169 clones. Clones from 11 of 21 Vitis sources had at least one clone cluster with the –SY lineage using primers RSP35 and RSP36. Included among these were Merlot clone 184 on 3309 rootstocks (vines 66-17, 66-18; cDNA 76.23, 76.24), Chardonnay (vine 49-9; cDNA 56.6), Cabernet Franc (vines 50-9, 50-10; cDNA 76.7, 76.8), Cabernet Sauvignon (vine 50-20; cDNA 76.17), Syrah clone 1 on 3309 rootstock (vine 66-11; cDNA 79.11), Syrah clone 1 with unknown rootstocks (vines 66-6, 93-2, 93-5; cDNA 79.6, 107.6, 107.9), and a Syrah clone 100 of unknown rootstock (vine 93-22; cDNA 107.16). Clones
within the –SY lineage were 92-97% similar to the GRSPaV-SY reference sequence. Clone 79.11B from Syrah clone 1 (vine 66-11; cDNA 79.11) branches off earlier than the rest of the –SY group, and was 88% identical to the GRSPaV-SY reference sequence. Some of the clones within the –SY cluster were 92-96% identical to the GRSPaV-GR3 (JN683373.1).

The –GG group had 12 clones in total and nucleotide identities ranged from 97-99% with the GRSPaV-GG reference sequence. These clones were sourced from a Chardonnay (vine 49-9; cDNA 56.6), Cabernet Franc (vine 50-7; cDNA 56.1), and a Merlot clone 184 on a 3309 rootstock (vine 66-16; cDNA 76.22). Only one clone clustered with the GRSPaV-BS lineage, with 99% sequence identity, which was obtained from Syrah clone 100 (vine 93-21; cDNA 107.15).

Figure 12 (page 60): Distribution of RSP35-RSP36 clones from all Vitis sources. Phylogenetic analysis was conducted using neighbour-joining with 1000 bootstrap replicates. Bootstrap values are presented as a decimal, with 1 being 100% confidence. The bar indicates 0.01 nucleotide substitutions/site, or genetic distance over time. Numbers in brackets represent the number of identical sequence clones obtained from the source. Reference sequences for the RSP35-RSP36 region of 15 sequence isolates obtained from GenBank are included, each prefixed with “RSP35-36 GRSPaV-”. The corresponding region from ASPV is included as an outgroup. Each group is represented by coloured branches; -SY, red; -PN, dark blue; -VF1, medium blue; -JF, cyan; -GG, magenta; -SG1, green; - BS, yellow.
SG1
39 clones
23%

GG
12 clones
7.1%

BS 1 clone 0.6%

JF 35 clones
20.7%

PN
89 clones
52.7%

VF1 52 clones
30.7%

SY 28 clones
16.6%

Total 169 clones
Table 3: GRSPaV *Vitis* clone distribution table. Corresponding cDNA indicates the cDNA sample number used for sequencing and in phylogenetic trees.

<table>
<thead>
<tr>
<th>Source</th>
<th>Red Leaf</th>
<th>GRSPaV Subgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GG (I)</td>
</tr>
<tr>
<td><strong>30-1</strong></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>cDNA 32.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety: Syrah—Clone 100, Rootstock 3309</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>50-7</strong></td>
<td></td>
<td>Some</td>
</tr>
<tr>
<td>cDNA 56.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety: Cabernet Franc</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>49-9</strong></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>cDNA 56.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety: Chardonnay—Clone unknown, Rootstock unknown, planted 1974</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>49-15</strong></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>cDNA 56.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety: Riesling—Clone Weis, Rootstock unknown, planted 1975</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>50-2</strong></td>
<td></td>
<td>Some</td>
</tr>
<tr>
<td>cDNA 76.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety: Cabernet Franc</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>50-9</strong></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>cDNA 76.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety: Cabernet Franc</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>50-10</strong></td>
<td></td>
<td>Some</td>
</tr>
<tr>
<td>cDNA 76.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety: Cabernet Franc</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>50-20</strong></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>cDNA 76.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety: Cabernet Sauvignon</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>66-16</strong></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>cDNA 76.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>66-17</strong></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>cDNA 76.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>cDNA</td>
<td>Variety</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>66-18</td>
<td>76.24</td>
<td>Merlot</td>
</tr>
<tr>
<td>66-6</td>
<td>79.6</td>
<td>Syrah</td>
</tr>
<tr>
<td>66-11</td>
<td>79.11</td>
<td>Syrah</td>
</tr>
<tr>
<td>93-2</td>
<td>107.6</td>
<td>Syrah</td>
</tr>
<tr>
<td>93-4</td>
<td>107.8</td>
<td>Syrah</td>
</tr>
<tr>
<td>93-5</td>
<td>107.9</td>
<td>Syrah</td>
</tr>
<tr>
<td>93-21</td>
<td>107.15</td>
<td>Syrah</td>
</tr>
<tr>
<td>93-22</td>
<td>107.16</td>
<td>Syrah</td>
</tr>
<tr>
<td>93-28</td>
<td>107.20</td>
<td>Syrah</td>
</tr>
<tr>
<td>84-15</td>
<td>128.4</td>
<td>Riesling</td>
</tr>
<tr>
<td>93-21</td>
<td>128.17</td>
<td>Syrah</td>
</tr>
<tr>
<td>Sample</td>
<td>cDNA</td>
<td>Variety</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>93-26</td>
<td>128.20</td>
<td>Syrah—Clone 100, Rootstock unknown</td>
</tr>
<tr>
<td>93-27</td>
<td>176.11 (RSP35-SY/RSP36 only)</td>
<td>Syrah—Clone 100, Rootstock unknown</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.1.2 RSP35/RSP36 amino acid sequence alignment

Sequence isolates obtained using RSP35-RSP36 (RdRp region) were translated to their predicted amino acid sequence using Ugene Integrated Bioinformatics Tools and aligned within Clustal Omega in order to find any aa changes in the RSP35-RSP36 sequence (Figure 13). As expected, the sequences were highly conserved across the subgroups, with some small differences. In all sequences, the conserved ‘AK’ residues of motif II of the RdRp are present, and the ‘AGDDM’ conserved sequence is also present, however the cysteine residue immediately following is not visible in the sequence. Variants of the –SG1 lineage had few significant changes to aa sequence in comparison to the consensus. An A1939T substitution is apparent in one of the –SG1 isolates, and this mutation is also found in both representative GRSPaV-GG sequences. The –SY variants have two distinct differences from the other sequences; G1941N and A1964S. Few aa differences were observed in the –BS sequence but a D1977E change is evident. Within the –PN lineage, no clones identified highly with the –PN sequence but two clones clustered most closely and are represented by 107.9F in Figure 13. An F1883L change is found in one of the –GG sequences and in 107.9F. Variants of the –PN subgroups all had V1945I change, which was also found in one of the sequences representing the –SY lineage. Distinct differences were seen at some aa positions in the –SY sequences; A1964S and G1941N. Residue 1999 is represented by a serine in variants of the –SG1, -GG, and –BS lineages, and an alanine in the –SY and –PN subgroup sequences. Figure 13 shows the aa alignment of these sequences representing each of the GRSPaV phylogroups.
Figure 13: RSP35-RSP36 sequences were translated to their predicted amino acid sequence and aligned using ClustalOmega. A selection of sequences highly detected in each subgroup were chosen and depicted here. Two sequence isolates represent the GRSPaV-SG1 group; 79.6D and 76.24B. Two sequence isolates represent the GRSPaV-GG group; 56.6C and 76.22D. The only clone which clustered with the GRSPaV-BS lineage is represented; 107.15A. Five sequence isolates representing the GRSPaV-PN lineage were selected; two from the –VF1 subgroup, two from the –JF subgroup, and the sequence which clustered most closely to the –PN sequence isolate (107.9F). Three sequence isolates represent the GRSPaV-SY group; 76.24D, 107.16C, and 79.6B. The conserved ‘AK’ residues of RdRp motif II and ‘AGDDM’ of the ‘AGDDMC’ conserved sequence of RdRp motif VI. Red solid blocked residues represent those which are identical. Red lettering indicates minor residue changes and black lettering indicates significant residue changes, such as those with opposing polarity. X represents an unknown amino acid.
3.1.3 RSP35-SY/RSP36 to specifically target GRSPaV-SY

Primers RSP35-SY and RSP36 (Table 6) were used to target a 476 nt fragment specifically GRSPaV-SY in *Vitis* varieties to more accurately understand the distribution of the –SY lineage, as other variants may outcompete in PCR amplification with degenerate primers. Of the 53 clones sequenced, 50 clustered with the GRSPaV–SY lineage, with 94% bootstrapping support. Two of the 3 remaining clones, from Cabernet Franc (vine 50-10, cDNA 76.8) and Syrah clone 100 (vine 93-21; cDNA 128.17), clustered with the –VF1 lineage. The other remaining clone, from a Merlot clone 184 rootstock 3309 (vine 66-17; cDNA 76.23) source, clustered distantly with GRSPaV-VF1, and identified to have 95% nt similarity to GRSPaV-Hai1 (AB277787), isolated in Japan from a *V. lubruscana* cv. Kyoho (Nakaune & Nakano, 2006).

Variants of the –SY lineage were present in all 14 (100%) sources tested with the –SY-specific primers, which were Merlot (3 sources, 12 clones), Cabernet Franc (3 sources, 12 clones), Riesling (1 source, 3 clones), and Syrah (7 vines, 26 clones) cultivars (Figure 14, Figure 24).

**Figure 14 (page 67):** Distribution of RSP35-SY-RSP36 clones from 14 *Vitis* sources. Phylogenetic analysis was conducted using neighbour-joining with 1000 bootstrap replicates. Bootstrap values are presented as a decimal, with 1 being 100% confidence. The bar indicates 0.01 nucleotide substitutions/site, or genetic distance over time. As expected the vast majority of clones clustered with GRSPaV-SY, however two clones clustering with the –PN subgroups were found using these SY-specific RSP35-RSP36 primers. Numbers in brackets represent the number of identical sequence clones obtained from the source. Reference sequences for the RSP35-RSP36 region of 15 sequence isolates obtained from GenBank are included, each prefixed with “RSP35-36 GRSPaV-“. The corresponding region from ASPV is included as an outgroup. Each group is represented by coloured branches; -SY, red; -PN, dark blue; -VF1, medium blue; -JF, cyan; -GG, magenta; -SG1, green; - BS, yellow.
Total 53 clones

-SY
50 clones

-VF1
3 clones
3.1.4 Syrah cultivars with and without red canopy

Syrah with red canopy had 13 of 35 clones cluster with –SY, with 95% bootstrapping support. The four vines discussed in the Syrah with red canopy group were Syrah clone 1 with unknown rootstocks (vines 93-2, 93-4, 93-5; corresponding cDNA 107.6, 107.8, 107.9 respectively), and one Syrah clone 100 (vine 93-21; cDNA 107.15). In total, 9 of 35 clones from Syrah with red canopy clustered with the –VF1 subgroup. The –SG1 lineage had 7 clones, which cluster closely with the GRSPaV–MG reference sequence. Five cones clustered with the –JF subgroup of the GRSPaV–PN lineage. The only clone that clustered closely with the –BS lineage is observable in this tree (Figure 15). Clone 79.11B from Syrah clone 1 expressing red blotch and red leaf symptoms identified by only 88% with the GRSPaV–SY sequence. The presence of these red leaf symptoms indicates SD.

The four vines in the Syrah without red canopy group include Syrah of unknown clone and rootstock (vine 66-6; cDNA 79.6) and Syrah clone 100 with unknown rootstocks (vines 93-26, 93-22, 93-28; cDNA 128.20, 107.16, 107.20). Syrah without red canopy had 3 of 38 (7.9%) clones from four Syrah sources cluster with the GRSPaV–SY lineage. In Syrah without red canopy 28 of 38 clones clustered within the two –PN subgroups. Of the –PN lineage, 12 clones clustered with the –VF1 subgroup and 16 clones clustered with the –JF subgroup (Figure 16, Figure 26).

Variant distribution differed in the number of clones in the –SY cluster when comparing Syrah cultivars with and without the characteristic SD red canopy phenotype. Both groups had many clones cluster with –VF1 and –JF subgroups. In all Syrah sampled, including those with or without symptoms, almost half of the clones (28 of 64) clustered with the subgroups of the –PN lineage, while 16 of 64 clones clustered with the GRSPaV–SY lineage.
**Figure 15:** Distribution of GRSPaV sequence variants in four Syrah with red canopy (cDNA 107.6, 107.8, 107.9, 107.15). The analyzed sequence was that what amplified with primers RSP35-RSP36. Phylogenetic analysis was conducted using neighbour-joining with 1000 bootstrap replicates. Bootstrap values are presented as a decimal, with 1 being 100% confidence. The bar indicates 0.01 nucleotide substitutions/site, or genetic distance, over time. Clones representing all groups were found in this data set; GRSPaV-SY, -BS, -VF1, -JF, -SG1, and -GG. Reference sequences for the RSP35-RSP36 region of the 15 sequence isolates obtained from GenBank are included. The corresponding region from ASPV is included as an outgroup. Each group is represented by coloured branches; -SY, red; -PN, dark blue; -VF1, medium blue; -JF, cyan; -GG, magenta; -SG1, green; -BS, yellow.
Figure 16: Distribution of GRSPaV sequence variants in four Syrah without red canopy (cDNA 79.6, 128.20, 107.16, 107.20). The analyzed sequence was that what amplified with primers RSP35-RSP36. Phylogenetic analysis was conducted using the neighbour-joining method and 1000 bootstrap replicates. Bootstrap values are presented as a decimal, with 1 being 100% confidence. The bar indicates 0.01 nucleotide substitutions/site, or genetic distance, over time. Clones representing most groups were found in this study; GRSPaV-SY, -SG1, -VF1, -JF, and -GG. Reference sequences for the RSP35-RSP36 region of the 15 sequence isolates obtained from GenBank are included. The corresponding region from ASPV is included as an outgroup. Each group is represented by coloured branches; -SY, red; -PN, dark blue; -VF1, medium blue; -JF, cyan; -GG, magenta; -SG1, green; - BS, yellow.
3.1.5 Genetic diversity of hybrid varieties

In hybrid varieties, GRSPaV variants from only three lineages were found; GRSPaV-SY, GRSPaV-BS, and the GRSPaV-VF1 subgroup of the GRSPaV-PN lineage. A total of 19 clones from 8 hybrid sources were included in this study; Niagara (1 source), Baco (4 sources), and Vidal (3 sources) (Figure 17, Figure 27). Unfortunately no detailed information on source clone, age, and symptoms was collected, so no disease correlations can be made and the sample size may be too small to accurately represent all hybrid varieties. Before the removal of unfit sequences due to excessive sequence length, a total of 30 clones were included in this study. Number of sequences removed was proportional for all GRSPaV clusters, and the original data presented as processed by the MegAlign (DNASTar) ClustalW software is presented in Appendix II (Figure 28). One clone in particular from Vidal (vine 205-53; clone 51-1) that clustered with the GRSPaV-GG lineage was removed, and had a sequence most similar to the GRSPaV-PG sequence isolate of this lineage. This was the only other lineage represented in the hybrid varieties. The majority of the clones, 12 of 19, clustered with the GRSPaV-SY lineage. All three varieties tested – Baco, Niagara, and Vidal – had variants clustering with the –SY lineage and all clones from Niagara clustered with GRSPaV–SY. The second most commonly detected group was GRSPaV-BS, with 4 of 19 total clones. These clones were from two Baco sources, and all four were 99% similar to GRSPaV-BS. Clones in the –VF1 lineage were 92 or 96% similar to the GRSPaV-VF1 reference sequence, 86% similar to the GRSPaV-PN sequence, and 93-94% similar to the GRSPaV-ML. To our knowledge this is the first report of GRSPaV variant distribution in hybrid varieties in Ontario.
Figure 17: GRSPaV sequence variant distribution in hybrid varieties. The analyzed sequence was that what amplified with primers RSP35-RSP36. Phylogenetic analysis was conducted using neighbour-joining method and 1000 bootstrap replicates. Bootstrap values are presented as a decimal, with 1 being 100% confidence. The bar indicates 0.01 nucleotide substitutions/site, or genetic distance, over time. Clones representing only GRSPaV-SY, -BS, and –VF1 subgroups were found in this data set. Reference sequences for the RSP35-RSP36 region are included for 15 sequence isolates obtained from GenBank. The corresponding region from ASPV is included as an outgroup. Each group is represented by coloured branches; -SY, red; -PN, dark blue; -VF1, medium blue; -JF, cyan; -GG, magenta; -SG1, green; - BS, yellow.
Table 4: GRSPaV distribution in hybrid sources according to sequences from amplicons using primers RSP35-RSP36. The number of clones within each lineage or subgroup is indicated for each individual source.

<table>
<thead>
<tr>
<th>Hybrid Sources</th>
<th>GRSPaV Subgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GG</td>
</tr>
<tr>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>205-3</td>
<td></td>
</tr>
<tr>
<td>Vidal</td>
<td></td>
</tr>
<tr>
<td>205-17</td>
<td></td>
</tr>
<tr>
<td>Baco</td>
<td></td>
</tr>
<tr>
<td>205-19</td>
<td></td>
</tr>
<tr>
<td>Baco</td>
<td></td>
</tr>
<tr>
<td>205-21</td>
<td></td>
</tr>
<tr>
<td>Baco</td>
<td></td>
</tr>
<tr>
<td>205-25</td>
<td></td>
</tr>
<tr>
<td>Baco</td>
<td></td>
</tr>
<tr>
<td>205-26</td>
<td></td>
</tr>
<tr>
<td>Baco</td>
<td></td>
</tr>
<tr>
<td>205-30</td>
<td></td>
</tr>
<tr>
<td>Niagara</td>
<td></td>
</tr>
<tr>
<td>205-39</td>
<td></td>
</tr>
<tr>
<td>Vidal</td>
<td></td>
</tr>
<tr>
<td>205-52</td>
<td></td>
</tr>
<tr>
<td>Vidal</td>
<td></td>
</tr>
<tr>
<td>205-53</td>
<td></td>
</tr>
<tr>
<td>Vidal</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
</tr>
</tbody>
</table>
3.1.6 CP diversity sequencing outcomes

Each of the CP primer pairs were individually tested with the RSP8634R primer to be sure each could produce successful amplicons. All produced successful amplicons with an annealing temperature between 50-53°C, with the exception of RSP7770F-PN which was expected as no reports of the –PN variant were made in the RdRp diversity analysis. Interestingly, the RSP7770F-BS primer appeared to amplify much more efficiently than the others, however the template cDNA came from grapevine 93-21; the only Vitis source positive for GRSPaV-BS as determined by the RdRp diversity study. Unfortunately, the results of preliminary sequencing revealed 9 of 12 clones identified as 23S rRNA from Vitis chloroplast and only the remaining 3 were GRSPaV. The PN7770F and RSP8634R did not produce successful amplicons during any attempt, though it should be noted no sequences clustering closely with –PN were found in this study, so a negative result is expected. This supports the theory that the GRSPaV-PN sequence may be a chimera.

Attempts to optimize the CP RT-PCR reaction were overall unsuccessful. Increasing annealing temperatures, different source material, increasing the total forward primers, and increasing template cDNA were parameters tested. Random primers were used for cDNA synthesis, to keep conditions congruent with the RdRp diversity study, however this is likely the source of the issue. Six more sources were tested using the CP primers. The expected CP amplicon was 864 bp long and the 23S rRNA amplicon was 885 bp, making distinguishing between these two amplicons very difficult. For clone screening, one of the three CP clones that identified with GRSPaV was used as a positive control, and one of the clones that identified with 23S rRNA was used as a negative control. Unfortunately the resulting screens indicated that the majority of amplicons were likely from 23S rRNA. The CP diversity study was not continued.
from here. The notion of redesigning this aspect of the diversity study outweighed the amount of cloning and screening required to determine only weakly confident clones for sequencing. Future directions for this experiment are explained in Chapter 4.

3.2 Full-Length Clone

The final FLC fragment constructs in pGEM-T-Easy obtained in this work cover the entire GRSPaV-SY genome, and each fragment is confirmed to be cloned in pGEM-T-Easy. These constructs are pGEM-F2a, pGEM-F2b, pGEM-F3, and pGEM-F4. In addition to these constructs, the CaMV 35S promoter was cloned into the downstream cloning target vector, pBluescript II KS+, producing pBS35S. The HDVnos terminator sequences has been cloned into pBS35S but digested out in preparation for downstream cloning. To finish construction of a full length clone, the F2 overlap fragment must be cloned into pGEM-T-Easy. Full sequences must be obtained for the small internal regions of fragments F2a, F2b, F3a, and F3b. Finally, the F2, F3, and F4 genome fragments must be cloned into pBS35S in addition to the HDVnos terminator sequence. A detailed description of the steps required to finish the FLC are described in the future directions section of chapter 5.

3.2.1 Sequence analysis of CaMV 35S promoter and HDVnos terminator sequences

The CaMV35S promoter and the HDV ribozyme nos sequence are required for constitutive expression of the FLC viral genome upon infection and termination of transcription, respectively. The resultant 318 nt CaMV 35S promoter sequence in pBS35S had 100% nt identity to the CaMV 35S promoter sequence in the pHST40 template sequence, with the exception of the NotI and XbaI sites added into the 35S primers (Appendix III, Figure 29). BLASTn analysis resulted in 100% nt identity to the CaMV 35S sequence in pCAPE1 virus-
induced gene silencing (VIGS) cloning vector (HQ687213.1) and in the template vector pHST40, and 99% identity to the CaMV sequence (V00141.1). NotI and XbaI sites were present at the expected 5’ and 3’ positions.

The HDVnos sequence was 100% identical to the HDVRz and nos terminator sequence in many cloning vectors including the template pHST40. The alignment of these sequences can be found in Appendix III (Figure 34). An additional nos terminator sequence, lacking the 5’ HDV ribozyme sequence, was under construction alongside the HDVnos FLC construct. It was intended for use as a future study to investigate the effect of the HDV ribozyme sequence on FLCs.

3.2.2 Sequence analysis of pGEM-F2a and pGEM-F2b

The choice for FLC source material was based on the outcome of the GRSPaV diversity analysis study. The primers for the FLC were designed to target the GRSPaV-SY sequence isolate from GenBank, however it was expected there would be variation in the GRSPaV-SY sequence from Ontarian grapevines but the FLC genome fragments would fall within the GRSPaV-SY lineage. Three clones positive for pGEM-F2a were sequenced. All identified to GRSPaV-SY and nucleotide identity was 95% at the 5’ end (genomic nt position 1), and 90% at the 3’ end (genomic nt position 2127). The clone designated pGEM1-2127-8 (pGEM-F2a for simplicity from here on) was chosen for its slightly higher cover query and cleaner sequence data. Following sequencing of the internal region, the full F2a sequence was assembled, identifying by 93% at the nt level to the GRSPaV-SY (AY368590) sequence in GenBank. Nucleotide alignment of F2a with the corresponding region in the –SY variant can be found in Appendix III (Figure 30). The two subsequent results of a BLASTn search of the GenBank
indicate the F2a sequence is 80% similar to GRSPaV-TannatRspav1 (KR528585) and 79% similar to GRSPaV-MG. The F2a nt sequence was translated into the aa sequence using DNASTar’s EditSeq function. The sequence was subject to BLASTp and identified most closely with the GRSPaV-SY (AY368590), with 93% identity, followed next by 80% similarity at the aa level to GRSPaV-1050-02 (JX513892) identified in Canada (Rott et al., 2012).

Two clones positive for pGEM-F2b were successfully sequenced with M13F and M13R. Based on the sequencing outcome, clone pGEM-1659-4410-22 (pGEM-F2b for simplicity) was chosen for its slightly higher nt identity and cover query. Primers were designed from the pGEM-F2b sequence to sequence the internal region. Upon assembly the entire F2b region identified by 92% to the GRSPaV-SY sequence at the nt level, followed by 81% identity to GRSPaV-TannatRspav1 (KR528585). The sequence alignment of F2b with the corresponding region of GRSPaV-SY sequence can be found in Appendix III (Figure 31). BLASTp analysis of the F2b aa sequence identified closest to GRSPaV-SY (AY368590) with 90% identity, followed by 77% identity to GRSPaV-TannatRspav1 (KR528585).

3.2.3 Sequence analysis of pGEM-RSP13-36, pGEM-F3a and pGEM-F3b

Preliminary attempts to obtain the F3a region (genome nt positions 4399-6074) used broad spectrum primers RSP13 and RSP36, which target genomic nt positions 4373-6180 and is referred to as the RSP13-36 fragment. A clone positive for the RSP13-36 fragment was sequenced and the resultant sequence was 89% similar to both GRSPaV-VF1 and GRSPaV-JF at the 5’ end (genome position 4373) (771 nt of clean sequence from nt 4373-5143 using M13R) and 91% similar to GRSPaV-JF from the 3’ end (genome position 6180) (835 nt of clean sequence from nt 6180-5344 using M13F). This construct was not used in any downstream work.
because following this conditions for successful GRSPaV-SY amplification with the RSP-SY4399F and RSP-SY7592R primers were found, using fresh total RNA extractions.

Two clones positive for pGEM-F3a were sequenced with M13F and M13R. The resultant sequences were 93-94% identity to GRSPaV-SY. The clone pGEM4399-6074-1 (pGEM-F3a for simplicity) was chosen due to slightly higher cover query. Upon full sequence assembly the F3a sequence identified most closely with the GRSPaV-SY sequence, with 93% identity. The complete F3a sequence identified by 84% to GRSPaV-TannatRspav1 (KR528585) and by 82% with GRSPaV-MG (FR691076) as the two subsequently closest results from GenBank.

Four clones positive for pGEM-F3b were sequenced with M13F and M13R. The resultant sequences were 94-97% similar to the GRSPaV-SY sequence at the 5’ end (genomic nt position 5938) and 92-93% similar at the 3’ end of this fragment (genomic nt position 7592). Clone pGEM-5938-7592-6 (pGEM-F3b for simplicity) was chosen for its sequence was most similar of the four clones to the GRSPaV-SY (AY368590) sequence. Internal regions of F3b were sequenced using primers designed from the outcome of the M13F and M13R sequence, and the full sequence was assembled. The full F3b region spanning nt positions 5938-7592 identified at 94% with the GRSPaV-SY sequence. The next most closely related GRSPaV variants were GRSPaV-OB1 (AB277783) by 92%, and GRSPaV-Hai1 (AB277787) by 80% similarity at the nucleotide level.

The entire F3 fragment was assembled, identifying to be 94% identical to the GRSPaV-SY sequence. The alignment of F3 with the corresponding region of the GRSPaV-SY sequence can be found in Appendix III (Figure 32). The next most closely related variants were GRSPaV-OB1 with 93% similarity, and GRSPaV-TannatRspav1 with 81% similarity at the nt level. After
translation of the entire F3 nt sequence to the predicted aa sequence using EditSeq (DNAStar),
BLASTp analysis revealed 99% aa similarity to GRSPaV-SY (AY368590), followed by 96% aa
similarity to GRSPaV-TannatRspav1 (KR528585).

### 3.2.4 Sequence analysis of pGEM-F4

The total length of the F4 fragment in pGEM-F4 was determined by overlapping
sequences from the M13F and M13R reads. Three positive pGEM-F4 clones were successfully
sequenced and gave sequences that are 96-97% identical to the GRSPaV-SY sequence on the nt
level. The resultant fragment was 1191, due to the polyT sequence included in the primer. The
pGEM-7584-8725-11 construct (pGEM-F4 for simplicity) was chosen due to its cleaner, more
reliable sequence. This clone identified with the GRSPaV-SY (AY368590) sequence by 97%
and the alignment can be found in Appendix III (Figure 33). GRSPaV-OB1 (AB277783) and
GRSPaV-H2 (GQ478315) followed with 96% and 95% identity, respectively. The F4 sequence
was translated into the predicted aa sequence and BLASTp analysis revealed 98% aa identity to
GRSPaV-SY (AY368590), trumped only by 99% aa identity to the isolate GRSPaV-FBFH1b
(FJ943297) which was first detected in Washington state (Alabi et al., 2009)

### 3.2.5 Attempt to clone F4 into pBS-35S failed

Two ligations of SalI-ClaI digested F4 and pBS-35S were transformed into E. coli DH5α
competent cells. No transformed colonies grew on LB Amp100 mg/mL agar plates from neither
the 1:3 not 1:1 insert:vector ligations. All cells from the transformation reactions were plated.
The ligation reaction failed, as the positive control was moderately successful. Ethanol
contamination was likely present in the purified F4 and pBS-35S fragments, inhibiting the
reaction.
Figure 18: FLC fragment alignment with GRSPaV-SY AY368590 (top solid black arrow). Fragments in descending order are F2a, F2b, F3, F3a, F3b, and F4. Nucleotide positions are indicated across the top of the figure. All fragments are in the same orientation and correspond to the respective GRSPaV-SY regions as expected.
3.2.6 Attempt to clone F3 into pBS-35S failed

The ligation reactions between SalI-ClaI digested F3 and pBS-35S was successful with repetition, producing more than 30 colonies on some agar plates. It is important to note that all transformants should appear white in blue-white screening, as the 35S CaMV promoter was already ligated within the MCS. PCR screening with overlap region primers RSP-SY5938F and RSP-SY6074R indicated the full fragment was likely inserted. Digestion to confirm the presence of the full F3 fragment in pBS35S gave unexpected results. Digestion with ClaI cut once, giving a single band of approximately 6.3 kb, as expected, and digestion with SalI should have given the same result. Two fragments of approximately 3.2 kb and 3.0 kb were visible with SalI digestion. No unexpected SalI sites were found in the sequencing file. It was previously determined that the F3 fragment ligated into the pGEM-T-Easy vector with the F3 SalI site towards the 5’ end of the MCS, and the F3 ClaI site towards the 3’ end of the MCS, or “backwards” based on sequence outcome. This meant the F3 SalI site and the pGEM-T-Easy MCS SalI site were on opposite ends of the F3 fragment, with the 3.2 kb F2 fragment in between (Figure 19). The SalI digestion was first to be carried out, giving a 3.2 kb fragment. A complete digest by ClaI in theory should cleave the 26 nt of the pGEM-T-Easy MCS between the T-cloning region and downstream SalI site. The complete ClaI digestion would remove unwanted MCS sequence and cut the F3 genome fragment at the natural ClaI site, and not interfere with downstream cloning. Restriction mapping indicated the ClaI digestion was incomplete, and SalI-digested pGEM-F3 ligated to SalI-digested pBS35S (Figure 19).
Figure 19: F3 cloning outcome as a result of incomplete ClaI digest. SalI-digested F3 fragment from pGEM-F3 (top) subcloned into SalI-digested pBS35S-F3 (bottom). The 26 nt region of the pGEM-F3 MCS is indicated. This occurred because the pGEM-F3 SalI digestion was carried out first, producing a 3.2 kB fragment that was purified before digestion with ClaI, which would delete only 26 nt. pBS-35S was also cut with SalI and purified before digestion with ClaI, expected to cleave 8 nt. Incomplete ClaI digestion of SalI-digested pBS-35S allowed for some fragments to have only a single cut at the SalI site in the MCS. This went undetected as the ClaI digest would only delete 8 nt from SalI-digested pBS-35S. The SalI-digested F3 fragment cut from pGEM-F3 also had incomplete ClaI digestion, leaving 26 nt of the pGEM-T-Easy MCS and two SalI sites flanking the F3 fragment. These flanking SalI sites allowed for cloning of F3 and the 26 nt region of the MCS into pBS-35S cut with only SalI.
CHAPTER 4: DISCUSSION

4.1 GRSPaV variants in Ontario are genetically diverse

In this work distinct clustering of GRSPaV sequence variants found in a range of grafted Vitis is observed, in similar findings to previous groups in different geographical locations (Lima et al., 2009; Terlizzi et al., 2011; Habili et al., 2006). Five distinct lineages were found to be present in Ontario; GRSPaV-SY, -SG1, -GG, -BS, and –PN. The GRSPaV-BS lineage was detected in one clone with 99% sequence identity to the reference strain, but variants of this lineage were not detected more than once. Both maximum likelihood and neighbour-joining algorithms were used for phylogenetic analysis. Maximum likelihood was initially chosen due to its more robust methods which find the most probable distribution of variants based on parameters of the data set that maximize the likelihood of the resultant tree representing the actual distribution (Kumar et al., 2016). Neighbour-joining is a simpler algorithm which uses a distance matrix to find relationships between genetic sequences by using genetic evolutionary distances, or nucleotide changes to predict relationships. Rather than finding the most likely distribution of the data set, neighbour-joining compares neighbouring (or next in sequence) position by position to link related sequences to one another (Saitou & Nei, 1987). The fact that all data sets analyzed by both maximum likelihood and neighbour joining algorithms gives the same distribution of clones provides an additional layer of confidence that the data presented in the phylogenetic trees is represented as accurately as possible. Because GRSPaV sequence variants are still highly related on a species level, evolutionary nt changes as defined by the bars on the phylogenetic trees, indicates that these sequences are highly related, and that all variants are from this viral species. Distinctions are observed by long horizontal branch distances, which helps define phylogroups. Phylogenetically, the GRSPaV-BS lineage is distinct and variants of
this lineage exist in Ontario, but were not as frequently detected as other lineages. The aa
alignment including the translated sequence for this sequence isolate identified very highly with
the –BS sequence. Few aas differed from those of the other lineages, and the majority of the
sequence variation was at the nt level, and did not change the aa outcome significantly. At aa
position 1977 all other sequence isolates encode glutamic acid except for the –BS sequence,
which has an aspartic acid at this position. Due to the similar properties of these aas being the
two polar acidic residues, this single aa change is not likely to cause significant disruption to the
RdRp in –BS variants, but does indicate an area of unique distinction for –BS sequences.

In *Vitis*, variants of the –VF1 subgroup were most prevalent, with 52 of 169 clones (30%)
belonging this lineage. Variants of the –VF1 subgroup were present across nearly all grapevine
varieties tested; Syrah clone 100, Chardonnay, Riesling clone Weis, Cabernet Franc, and Merlot
clon 184 on 3309 rootstocks. The –VF1 group was widespread among sources, detected in 11 of
23 grapevine sources. The -VF1 reference isolate was first described in Brazil in 2015 and was
isolated from *Vitis flexuosa*. The –VF1 group is a subgroup of the –PN lineage, as it appears to
share ancestry with –PN. Clones from this group cluster closely to the –PN reference sequence
but had a distinctly divergent lineage. It is interesting that the most commonly detected group is
also present in Brazilian *Vitis*. It is very likely that grapevine grafting material in Canada comes
from some of the same originating sources (from the USA) as grafting material used in Brazil.
Little research is available on this subgroup, however its distribution and lack of disease
symptom correlation suggests it is not involved directly in symptom onset. The other prevalent
subgroup of the –PN lineage is the –JF subgroup, with 35 of 169 clones clustering with this
isolate. The RSP35-RSP36 target region of the GRSPaV-VF1 GenBank sequence is 91% similar
to the GRSPaV-JF sequence. The –PN reference sequence for this region is 88% similar to the
GRSPaV-JF sequence and 87% similar to the GRSPaV-RSP47-4 sequence according to BLASTn. The results in Figure 12 show two clones, 107.9F and 107.9G, which are 91% similar to both GRSPaV-JF and GRSPaV-RSP-47-4, and are the most closely related clones to the GRSPaV-PN sequence. These sequences were not more similar to any other isolates in GenBank and may represent a novel variant of GRSPaV within the –PN lineage. The aa alignments of –JF, -VF1 and the sequence most closely related to the –PN isolate were included in Figure 13. At aa residue 1999, the sequences from the –SG1, -GG and –BS lineages encode a serine, whereas the –PN lineage, including –VF1 and –JF, and the sequences from the –SY lineage all encode an alanine at this position. The sequence most closely related to the –PN sequence (107.9F) was studied for aa changes that made it distinct from the –VF1 and –JF groups; at aa residue 1989 variants of –JF and –VF1 encode an isoleucine whereas the 107.9F encodes a valine. This change would likely not cause any significant structural or functional changes but does allow for an area of distinction within the -PN subgroups. One major change noted in the 107.9F sequence is the aa residue 1883 has a leucine in place of the phenylalanine, changing a non-polar hydrophobic residue to a polar, uncharged residue. One sequence isolate representing the –GG group also had this mutation. The other major difference between sequences of the subgroups is found at residue 1883, at which the –PN related sequence encode a leucine and the –VF1 and –JF sequences, and all others encode a phenylalanine, with the exception of one –GG sequence variant (76.22D) which also encodes a leucine at this position.

The GRSPaV-SG1 lineage had the second largest cluster of clones, with 39 of 169 clones. Clones from this lineage were from 11 of 23 grapevine selections, including Syrah clones 1 and 100, Cabernet Franc, Cabernet Sauvignon, Merlot, and Riesling. High prevalence of GRSPaV-SG1 sequence variants is likely the result of grafting onto St George indicator plants,
on which GRSPaV-SG1 does not induce RSP symptoms, making a lot of “GRSPaV-free”
grafting material contaminated. Many of the vineyards from which samples were taken were
planted before the current diagnostic testing available for GRSPaV. No obvious aa substitutions
were seen in the GRSPaV-SG1 representative sequences in the RSP35-RSP36 aa alignment
(Figure 13). The high prevalence of variants of the –SG1 group can be in part explained by
previous misnotation that St George cultivars were free of GRSPaV, and were used as RSP
bioindicator plants for many years (Meng et al., 2000; Meng et al., 2003; Habili et al., 2006).
Variants of the SG1 group do not induce RSP symptoms when graft-inoculated on St George,
and prior to the more recent availability of molecular diagnostic testing grafting tissues and tools
which were in contact with these so-called “GRSPaV-free” vines contaminated newly grafted
vines (Meng et al., 2005).

Variants of the GRSPaV-GG lineage were only confirmed for 12 clones of the 169 from
Vitis. The –GG group appears to share ancestry most closely with variants of the –SG1 group,
and the distinction of the –GG group from the –SG1 group may be the result of viable
evolutionary nucleotide changes. The aa alignment of –GG and –SG1 sequence variants
indicates one area which may support this theory (Figure 13). Residue 1939 is an alanine across
the –SY, -PN, -BS sequences, and one –SG1 sequence variant (79.6D). One sequence variant
from the –SG1 group (76.24B) and both sequence variants from the –GG group (56.6C and
76.22D) have a threonine at this position.

Clones clustering with the GRSPaV-SY linage were widespread among grapevines,
found in 11 of 23 sources. Clone 79.11B from Syrah clone 1 expressing red blotch and red leaf
was only 88% identical to the GRSPaV–SY isolate (Figure 12). This clone was not included in
the Syrah with red canopy group, as the entire canopy was not yet red. It would be of interest to
monitor this vine for SD symptoms over the coming years. Sequence isolates repeatedly detected which cluster with the GRSPaV-SY lineage had two significant amino acid changes that were not found in sequences from any of the other lineages. Residue 1941 of the aa sequence is glycine across all sequences expect those representing the –SY lineage and the –JF subgroup. One of the two –JF sequences encodes glycine (56.6K), the other encodes aspartic acid, changing from a non-polar, hydrophobic residue to a polar charged residue. All representing sequences of the –SY lineage encode an asparagine at residue 1941, a change from a non-polar residue to a polar, basic residue may be a key to the viability of variants of this group. Residue 1964 is alanine across all groups except the sequences representing the –SY group, which has a serine at this position, changing the hydrophobic alanine to a non-hydrophobic serine. Because of the right hand-like conformation of the RdRp protein, significant changes from hydrophobic to hydrophilic residues would affect the final folding of the RdRp, however it is evident that variants of the –SY group have adapted in a way to persist in Syrah and other cultivars.

It is expected that sequences within groups will vary as the viral RdRp has no proofreading activity, and sequence mutations are not corrected (Meng et al., 2006; Terlizzi et al., 2010). Viable mutations resulting in translation products capable of successful replication and infection expand the genetic variability of GRSPaV genomes and are important attributes of genetic diversity and evolution. Within the climate and soil conditions of Southern Ontario, in addition to many other biotic factors, *Vitis* are heavily infected with variants of the GRSPaV-VF1, –JF, and –SG1 isolates, indicating these variants have genetic mutations or adaptations allowing their persistence in these vines. On the other hand, only one clone related to the GRSPaV-BS isolate was detected in *Vitis* and only 12 clones clustered closely with the –GG isolate. Variants of these sequence isolates appears to be less persistent than –VF1, -JF, and -SG1
variants. Overall, variants of the GRSPaV-SY lineage were moderately detected in *Vitis,* but distribution is –SY variants differed when looking specifically at Syrah with red canopy and Syrah without red canopy, as discussed below.

### 4.2 GRSPaV-SY is likely involved in SD in Syrah, but not the sole causal agent

As seen in previous studies, variants of the GRSPaV-SY lineage were present in both declining and symptomless Syrah (Lima *et al.*, 2006a; Habili *et al.*, 2006). However the results obtained in this study show a much higher instance of clones from the –SY lineage in Syrah with red canopy compared to Syrah without observable SD symptoms. It is findings such as this that suggest the issue may be more complex than anticipated. Ontarian climate differs greatly from that of California and Australia, and different growth conditions, including soil and climate, may have effect on onset of SD symptoms (Habili *et al.*, 2006). SD may be influenced by different climates, and symptoms very likely change depending on the climate and soil conditions. Vines in Ontario are annually weakened by harsh winters, and frost does serious damage to grapevine tissue that requires time and energy for repair. This study compares four red canopy Syrah with four Syrah without red canopy, and gives an indication that there is a relationship between GRSPaV-SY variants with SD. A greater number of samples in each category would give a more conclusive indication of the validity of this relationship. Previous studies on GRSPaV-SY variants in declining Syrah conducted in warmer climates found –SY variants in both declining and symptomless Syrah (Lima *et al.*, 2006a; Habili *et al.*, 2006). Their findings alongside the findings of this thesis suggest variants of the –SY isolate persist more successfully in Syrah cultivar. However, the harsh fluctuating Ontarian climates add an additional cold stress to vine survival which causes the plant be selective on production of certain hormones, metabolites, and plant defence mechanisms (Thakur & Nayyar, 2013). Annual dormancy and exposure to severe
cold (-20°C) and changes of metabolism may allow viral replication to persist while without showing symptoms until the plant has matured. Age of vine needs to be considered carefully in future SD studies in Syrah, as onset of symptoms may not arise for a number of years. Annual monitoring of grapevine viruses including GRSPaV in newly planted Syrah would be an asset to understanding SD onset.

4.3 The GRSPaV-PN sequence may be a chimera

The GRSPaV-PN sequence was solved using cDNA libraries of plant viral RNA from *V. vinifera* cv. Pinot noir clone 23 on 3309 rootstocks (Lima *et al.*, 2009). At the time of GRSPaV-PN (AY368172) sequencing, much less GRSPaV sequence information was available, and only five other GRSPaV sequence variant genomes had been fully sequenced; GRSPaV (AF026278), GRSPaV-1 (NC_001948), GRSPaV-BS (AY881627), GRSPaV-SG1 (AY881626), and GRSPaV-SY (AY368590) (Zhang *et al.*, 1998; Meng *et al.*, 1998; Lima *et al.*, 2006a). Two of these sequence variants, GRSPaV (AF026278) and GRSPaV-1 (NC001948), differed at the nucleotide level by only 2%, thus they cluster within the same lineage (Lima *et al.*, 2009; Meng *et al.*, 1998; Zhang *et al.*, 1998). However, Lima *et al.* (2009) compared the GRSPaV-PN sequence identity at the gene- and genome level, which differed by 76-78% from the other lineages’ genomes (Table 5). Today we know multiple sequence variants of GRSPaV are almost always found within a vineyard, and importantly, within a single grapevine source, confirmed in different geographical locations (Meng *et al.*, 2006; Goszczynski, 2010; Lima *et al.*, 2006a). Sequence construction using cDNA libraries can be prone to misalignment of sequences by computational or human error when multiple viral sequence variants, known or unknown, are present within a single source. In this thesis research, no clones from any cultivar clustered closely with the GRSPaV-PN sequence, but many clustered in subgroups of the –PN lineage.
Previous GRSPaV diversity studies were also negative for clones highly similar to the –PN reference sequence (AY368172) (Meng et al., 2006; Lima et al., 2009). Many others have detected subgroups within the –PN lineage, as seen here. In this work, two near identical clones from a single Syrah clone 1 source share a common ancestor with the GRSPaV-PN sequence (Figure 12). Nucleotide sequence analysis of these clones via BLASTn resulted in 91% sequence identity to GRSPaV-JF and 89% identity to GRSPaV-VF1. These clones were also 91% identical to GRSPaV-RSP47-4 (EF105294), a partially sequenced isolate from cambium scraping of V. vinifera in the USA, and 90% identical to GRSPaV-GR4 (JN683374), an Egyptian isolate (Meng & Gonsalves, 2006). This suggests the two clones are derived from an isolate that is different from the –JF and –PN variants, but cluster together.

In preliminary work for this thesis, primers specific to the HVR of GRSPaV-BS, GRSPaV-SY, and GRSPaV-PN were used to target these sequence variants. Nine sources were tested using RSPSY1659F-RSPSY2127R which targets a 468 nt region of the GRSPaV-SY HVR. Three sources gave clear amplification with these primers; 50-7 Cabernet Franc, 49-9 Chardonnay, and 49-15 Riesling clone Weis. Primers BS1672F-BS2109R and PN1701F-PN2187R target the HVR of GRSPaV-BS and GRSPaV-PN respectively. Expected amplicons were 437 bp for the –BS variant and 486 bp for the –PN variant. These primer pairs were used to target –BS and –PN variants in 20 sources including Cabernet Franc, Chardonnay, Riesling, Cabernet Sauvignon, Merlot clone 184 x 3309, Syrah clone 1 x 3309. Only 50-2 Cabernet Franc and 50-20 Cabernet Sauvignon were positive with the GRSPaV-BS HVR primers, with some very weak amplification in some other sources. No sources were positive when primers targeting the GRSPaV-PN HVR were used, however a positive control was unavailable and thus primers may not have functioned under conditions tested. It is worth noting that during this research in
our lab, Dr Huogen Xiao found a grapevine source (Plant ID LS64 Carignan, originally from Foundation Plant Services at UC-Davis) which tested positive with the GRSPaV-PN HVR primers, producing a 488 nt amplicon. BLASTn sequence analysis of this fragment confirmed a 97% sequence identity to GRSPaV-PN at the nucleotide level. No other GRSPaV-PN sequences were found in other sources or with other primers. If the –PN sequence is in fact a chimera of two or more GRSPaV sequence variant genomes, it may be such that the –PN HVR sequence is from one –PN subgroup and the RSP35-RSP36 region of the RdRp is from another variant, as the RdRp primers target a region approximately 4 kB downstream. Primers RSP35 and RSP36 have not been tested on this source material.

4.4 GRSPaV distribution in hybrid varieties differed from Vitis vinifera varieties

The majority of clones clustered with the GRSPaV-SY lineage, and the four clones which clustered with GRSPaV-BS were the second largest group, followed by GRSPaV-VF1. While no information regarding symptoms or clones of the hybrid varieties was obtained, this is to our knowledge the first report on GRSPaV diversity in hybrids. Hybrid sources appear to be infected with a single GRSPaV variant group. Preliminary sequencing data confirm the presence of isolates from two different groups in two of the eight sources, however after removal of unfit sequences for bioinformatics processing, “single” infections appear in all sources. The preliminary data can be found in Appendix II, Figure 28. The sequences, while from GRSPaV variants, were removed because they were longer than the expected 476 nt fragment, and would have caused considerable disruption to the algorithms used in neighbour-joining and maximum likelihood and the phylogenetic trees would be less reliable. The two hybrids that had clones from more than one group were both Vidal. The first, 205-39, had one clone cluster with –SY, and 4 with the –PN subgroups. The other Vidal source, 205-53 had one clone cluster with –GG
and one with –SY. From these data we can conclude that hybrid varieties are infected with one or two variants, and the majority of infections are with GRSPaV-SY variants. Terlizzi et al. (2011) state that the 3’ end of the newly identified –ML variant is 92% identical to the –BS variant, however here we find clone 39-I from Vidal which clusters with the –VF1 group, shares a high sequence identity of 92% and 94% with both –VF1 and –ML, respectively. While the RdRp is highly conserved between variants, the –ML RdRp sequence appears to be more closely related to the –VF1 subgroup (-PN subgroup). This may suggest crossover events between the genetic material from a –VF1 variant and a –BS variant.

4.5 The FLC sequences are most closely related to the GRSPaV-SY isolate

All fragments used to construct the FLC identified most closely to GRSPaV-SY (AY368590) isolate upon BLASTn analysis. Fragments F2a, F2b, and F3 were 93%, 92%, and 94% similar, respectively, to the GRSPaV-SY nt sequence, whereas the 3’ F4 fragment was 97% similar to the –SY nt sequence. This is expected as the majority of the variation between GRSPaV sequence isolates is found within ORF1, including the HVR. The 3’ end of the genome, as represented by F4 was most similar of all fragments to the GRSPaV-SY sequence, in addition to having high similarity to many other GRSPaV sequence variants. This is the result of high sequence conservation of the CP gene. Any mutations in the CP gene sequence resulting in aa changes and CP protein folding changes are usually not viable, as virion formation cannot successfully take place. Repeated CPs interlock with one another to encapsidate the viral genome, but disruption to interlocking interfaces would misshape the virion and provide insufficient genome protection. Fortunately, at the aa level, the F4 fragment was 98% similar to the corresponding region of GRSPaV-SY and CP viability is very likely, and may be better functionally.
Table 5: Comparative analysis of GRSPaV-PN with four sequence variants. Adapted from Lima et al., (2009). The nt and aa identities of the –PN sequence compared to four GRSPaV sequence variants representing four different lineages are indicated for the complete genome, individual ORFs, and UTRs.

Table 1. Comparative analysis of nucleotide (nt) and amino acid (aa) sequences of the five open reading frames (ORF) of the GRSPaV-PN genome and four full-length GRSPaV genome sequences available in the database. Comparison is shown as percentage identity. Analyses were performed with ClustalW from the European Bioinformatics Institute. Regions for the genome compared included: UTR=untranslated region; REP=replicase; TGB=triple gene block; CP=coat protein.

<table>
<thead>
<tr>
<th>Encoded Gene</th>
<th>GRSPaV(^1) nt</th>
<th>GRSPaV-BS nt</th>
<th>GRSPaV-SG1 nt</th>
<th>GRSPaV-SY nt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete Genome</td>
<td>76 -</td>
<td>77 -</td>
<td>78 -</td>
<td>77 -</td>
</tr>
<tr>
<td>5'UTR</td>
<td>90 -</td>
<td>91 -</td>
<td>93 -</td>
<td>94 -</td>
</tr>
<tr>
<td>REP</td>
<td>75 84.7</td>
<td>75.6 84</td>
<td>74.8 84</td>
<td>76 84</td>
</tr>
<tr>
<td>TGB1</td>
<td>77.6 87</td>
<td>82.7 90.5</td>
<td>77 86</td>
<td>76 84</td>
</tr>
<tr>
<td>TGB2</td>
<td>78.8 83.7</td>
<td>82.7 87</td>
<td>80 82</td>
<td>74 78</td>
</tr>
<tr>
<td>TGB3</td>
<td>84 87.5</td>
<td>86.8 92</td>
<td>82 83.7</td>
<td>80 78</td>
</tr>
<tr>
<td>CP</td>
<td>82 93</td>
<td>88 95</td>
<td>81 91.8</td>
<td>83 93</td>
</tr>
<tr>
<td>3'UTR</td>
<td>93 -</td>
<td>93 -</td>
<td>95 -</td>
<td>91 -</td>
</tr>
</tbody>
</table>

\(^1\)GRSPaV (Zhang et al., 1998); GRSPaV-BS and GRSPaV-SG1 (Meng et al., 1998); GRSPaV-SY (Lima et al., 2006).
In many of the fragment BLAST analyses, GRPSaV-TannatRspav1 was a common second for nt and aa sequence similarity. This sequence was deposited into the GenBank database by a research group in Seoul, Korea, by direct submission, rendering little information about the variant available (Cho et al., 2015). Phylogroup analysis in this work shows that this isolate clusters with the GRSPaV-GG lineage, sharing a close common ancestor with the –SG1 lineage. The speculation that the second most closely related GRSPaV sequence variant is -TannatRspav1 is interesting. It suggests the FLC sequence isolate from this work shares a common ancestor with the GRSPaV-TannatRspav1 and infectious properties associated with the –SG1 lineage should be investigated for vines positive for the –JH variant.

The nt sequences for F2a, F2b, F3 and F4 were translated to aa sequences using EditSeq (DNAStar) and predicted aa sequences were analyzed using BLASTp. The aa sequences of the F3 fragment, containing the 3’ third of ORF1, ORF2, and ORF3, was 99% similar to the corresponding GRSPaV-SY sequence. The RdRp, TGBp1 and TGBp2 are encoded by the F3 sequence. Due to the essentiality of these proteins in viral infection, the high sequence similarity to the –SY sequence is favorable for the FLC, as these proteins will closely mimic that of wildtype GRSPaV-SY, and more importantly represent variants within the –SY lineage. Additionally, the F4 fragment identified with 97% nt similarity and 98% aa similarity to GRSPaV-SY. F4 aa sequence identity 99% similar to GRSPaV-FBFH1b (FJ943297), an isolate first identified in the Pacific Northwest, submitted to GenBank as a direct submission in 2009. This sequence was one of 139 clones of a 780 nt long CP gene region deposited at the time of submission (Alabi et al., 2009).

The tentative GRSPaV-JH genome was assembled completely and ORFs were predicted using the corresponding ORF nt positions of known GRSPaV sequence isolates, in particular the
GRSPaV-SY sequence. The –JH 5’ (non-coding region) NCR (nt 1-60) had 100% sequence identity to the GRSPaV-SY 5’ NCR sequence. This region is the most highly conserved region of the GRSPaV genome, and may be important for transcription initiation (Meng & Gonsalves, 2007). ORF1 (nt 61-6546) had 93% nt sequence similarity to GRSPaV-SY, followed distantly by 82% nt identity to GRSPaV-TannatRspav1. The nt sequence of ORF1 was translated into the predicted aa sequence, which identified by 92% similarity to the GRSPaV SY aa sequence and by 83% similarity to GRSPaV-TannatRspav1. The alignment of ORF1 of GRSPaV-JH and GRSPaV-SY is found in Appendix III (Figure 35). ORF2 (nt 6577-7242) encoding TGBp1 identified at 91% nt similarity to GRSPaV-SY. This outcome was followed by only one other BLAST sequence hit of 90% nt similarity to GRSPaV-OB1, a sequence which groups with the –SY lineage as determined by the RdRp genetic diversity data in this thesis. The predicted aa sequence for ORF2 had 92% similarity to GRSPaV-SY and 94% aa sequence similarity to GRSPaV-OB1. The alignment of ORF2 of GRSPaV-JH and GRSPaV-SY is found in Appendix III (Figure 36). It is interesting that this sequence did not align with more than two isolates in the GenBank database however the gene encoding the TGB sequences have not been the target for GRSPaV diversity studies. Many partial GRSPaV sequences are deposited following diversity analysis and/or screening using broad-spectrum primers, and to do this primers must target highly conserved genes like the CP or RdRp genes. Thus, many GRSPaV isolates have been partially sequenced and the sequences are available in GenBank. ORF3 (nt 7244-7597) encoding TGBp2 identified at 95% nt similarity to GRSPaV-SY, again followed by 94% nt similarity to GRSPaV-OB1. The predicted aa sequence encoded in ORF3 aligned with 99% similarity to GRSPaV-SY and 98% aa sequence similarity to GRSPaV-OB1. The alignment of ORF3 of GRSPaV-JH and GRSPaV-SY is found in Appendix III (Figure 37). ORF4 (nt 7581-7760)
encoding TGBp3 identified at 97% nt similarity with both GRSPaV-SY and GRSPaV-OB1. The predicted aa sequence for ORF4 of the –JH sequence identified to have 97% sequence similarity to both the –SY variant and the –OB1 variant. The alignment of ORF4 of GRSPaV-JH and GRSPaV-SY is found in Appendix III (Figure 38). ORF5 (nt 7770-8549) encoding the CP had 96% nt similarity to GRSPaV-SY and 98% nt identity to GRSPaV-LN-NHHL (KF731967). The predicted aa sequence for ORF5 identified by 98% sequence similarity to GRSPaV-SY and 99% aa sequence similarity to GRSPaV-FBFH1b (FJ943297). The alignment of ORF5 of GRSPaV-JH and GRSPaV-SY is found in Appendix III (Figure 39). The –LN-NHHL isolate was directly deposited into GenBank by a group in China in 2013. The group does not specify which grapevine cultivar this isolate was obtained from. As stated, the GRSPaV-FBFH1b isolate was first noted in the Pacific North West USA. Grapevine propagating material used in the USA and Ontario are likely from the same sources and suppliers, which may explain the presence of this isolate in Ontario. The beginning of the 3’ NCR is variable across four major GRSPaV groups; GRSPaV-1, -SG1, -BS, and –SY (Meng & Gonsalves, 2007). The 3’ NCR excluding the polyA sequence (nt 8550-8724) had 97% nt similarity to GRSPaV-SY. The high sequence similarity to the –SY variant further indicates the –JH sequence is very likely of the –SY lineage.

The entire assembled GRSPaV-JH genome is 8,724 nt long excluding the polyA tail and has 94% nt sequence similarity and 100% cover query with GRSPaV-SY. This indicates the GRSPaV-JH sequence variant is distinct from the GRSPaV-SY variant, and shares a common ancestor with GRSPaV-SY. The next BLASTn outcome identified full –JH genome to have 81% nt similarity to GRSPaV-TannatRspav1 followed by 80% nt similarity to GRSPaV-MG. These identities are significantly lower than the GRSPaV-SY identity, further indicating the GRSPaV-JH sequence is of the –SY lineage.
4.6 A GRSPaV-SY-related infectious clone is being constructed

Upon completion of the infectious clone, infiltrations of *N. benthamiana* and grapevine tissue culture will be carried out to confirm infectivity and use in a wide range of downstream studies. Grapevine micro-shoot tissue culture is a method in which a single plant is regenerated from a tiny shoot of approximately 0.5 mm. Viruses and other plant pathogens are eliminated from the host as a result of meristem regeneration outweighing the rate at which viral replication and infectivity can occur. Heat treatment is a method employed by clean plant programs for production of pathogen-free plants (Golino, UC Davis Online). Pathogen-free grapevine tissue culture will allow for *in vivo* studies observable in the native host. Localization of viral proteins and complexes using FLC constructs in grapevine micro-shoot tissue culture will allow for new information about the GRSPaV replication cycle. The GRSPaV-SY clone will allow observation of differences in virion number and protein expression from two distantly related GRSPaV sequence variants. The GRSPaV-SY infectious clone will be compared to a GRSPaV-GG infectious clone previously constructed in our lab, pRSP28-2(Cam), which will act as a positive control for infiltrations (Meng *et al.*, 2009). So far, pRSP28-2(Cam) has not been proven to move within *N. benthamiana*, a non-native host (Meng *et al.*, 2009; our unpublished data). It is therefore unlikely that based on the previous statement, movement within a model plant by the GRSPaV-SY infectious clone would provide new information about GRSPaV movement.

A grapevine tissue culture system is currently being developed in our lab. This will be used for downstream work from this thesis to infect virus-free grapevine tissue with the GRSPaV-SY infectious clone. Many experiments can be designed from these basic grapevine tissue infiltration methods. For example, using Real-time PCR, grapevine viral response gene
activation/deactivation could be determined by observing changes in grapevine gene expression at increasing time points post infection.

4.7 Troubleshooting

4.7.1 Sequencing issues

The requirement for sequencing a large number of clones lead to the decision to perform the Sanger sequencing reactions in the lab, and send the product to the AAC Genomics Facility at the University of Guelph. The protocol used in the Genomics Facility was generously taught to myself and Sunny Li by Jeff Gross and is described in the Methods section. Preliminary trials of small scale sequencing provided many failures.

Four clones were sequenced initially; two from this study and two GLRaV-3 clones prepared by Sunny Li. Fluorescence signal detected by the sequencer (Applied Biosystems 3730 DNA analyzer) was very weak, but expected sequences were obtained. It was later determined that the Biorad MyCycler Thermocycler used for the PCR reaction was malfunctioning or used improperly. The second attempt at sequencing involved a total of 8 clones and reactions were cycled in the AAC thermocycler (Applied Biosystems 2720). Fluorescence signal improved, however three sequences failed, giving only a short string of Ns. It was later determined that the short strip PCR tubes used in the reaction were incompatible with the AAC thermocycler, as tubes were too short to properly fit the machine. On the next attempt, Jeff Gross kindly offered to carry out the reaction as I watched, took notes, and asked questions. Three clones were sequenced and all were successful, and it was at this time that the issue with the short tubes was uncovered. Next, five clones were sent to BioBasics for sequencing by an outside source; only two clones were successful. This suggested that in this case the source of the problem was with
the material being sequenced, not the sequencing reaction machinery or reaction vessel. Using long tubes and the AAC thermocycler, two of three total clones were successful but gave a weak fluorescence signal, one clone gave the string of Ns. At this point the drying step in set up of the PCR reaction was in question. Another four clones were sequenced, without the drying step, and none were successful. Had either of these reactions been successful, an attempt at a larger number of sequences was planned to follow, however more testing was required. The protocol recommended 28 ng/kb however approximately 135 ng of DNA was used per clone for the GRSPaV diversity study, as Jeff advised more DNA would not be an issue. It was thought that the error may be in pipetting, as the stock plasmid concentration of the clones was relatively high, requiring pipetting of less than 0.5 µL for some samples. In order to make pipetting more accurate, plasmid DNA for sequencing was diluted to a concentration of 45 ng/µL. Pipetting 3 µL at this dilution gave a more accurate means of obtaining clean DNA at the appropriate concentration. Success was seen after this change was made and following successful trials, a whole plate (96 samples) came back with only successful samples. Following this, half a plate was sent (48 samples), and unfortunately none of the samples were successful; however the AAC Genomics facility’s positive control worked. Jeff suggested that a run of Gs and Cs just upstream of the MCS may cause a hairpin that stops the polymerase from extending the sequence all together. Controls were then to test the exact problem with all of our samples. This included; using samples with different vectors pBluescript II KS and pGEM-T Easy, drying primers and plasmid DNA for 3 minutes at 96°C compared to not having this drying step, plasmids that have previously been successful from the same plant as plasmids that were unsuccessful, and old and new sequencing buffer. Each of the conditions were duplicated; one group with drying and the second group without drying. Inconclusively, every sample in this test came back with great
sequencing signal. To conclude, it is believed that the dilution of template DNA to approximately 45 ng/µL was the biggest asset to improving sequencing quality, in addition to using the correct tubes, fresh reagents, and clean template DNA.

4.7.2 Full-length cDNAs for FLC

Obtaining high concentration, pure, large fragments for the FLC by RT-PCR proved to have a number of technical obstacles. First, cDNA from total plant RNA extractions was obtained using the GRSPaV-SY FLC genome reverse primers to target all regions upstream of these primer binding sites. This was also to ensure the entirety of the GRSPaV-SY variant genome was reverse transcribed to cDNA for PCR amplification. Small fragments F1 and F4 were easily amplified by PCR and were subsequently cloned into pGEM-T-Easy and sequenced to confirm GRSPaV-SY identity. When initial difficulties obtaining larger GRSPaV genome fragments, F2 and F3, continued to reoccur it was believed that the reverse transcriptase enzyme used M-MLV RT, was not suitable for long distance reverse transcription. SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) was chosen for RT reactions because it lacks RNase H activity and has superior viability at 50°C. RNase H degrades the RNA strand in RNA-DNA hybrids, which in the case for RT of viral RNA genomes, the RNA template is essential for high quality cDNA. Superscript III is highly thermostable, resulting in more efficient RT at higher temperatures, and room to increase temperature for primer specificity (Thermo website). RT reactions with SuperScript III were carried out for a maximum of 3 hours in order to obtain sufficient template cDNA for the FLC genome. Resultant cDNA was tested with GRSPaV primer pairs targeting small regions upstream of each GRSPaV-SY reverse primer.
4.7.3 Difficulties in obtaining large cDNA fragments and issues with ethanol

Large fragments, F2 and F3, were still unattainable so the decision was made to design primers for overlap PCR, and the PCR amplicon map and cloning strategy was adapted. Fragment F1 was no longer used, as the most successful primers for F2 overlap included RSP-SY1F(Xba), covering the entirety of the F1 region. Fragment F2 (nt 1-4410) was divided into F2a (nt 1-2127) and F2b (nt 1659-4410), with an overlapping region of 468 nt. F3 was divided into F3a (nt 4399-6074) and F3b (5938-7592) with an overlapping region of 136 nt. The overlapping region functions as a megaprimer for extension and amplification of the complete fragment. After determining the optimal conditions for the new internal primers, T-cloning of megaprimer fragment F3b was successful, however F2a, F2b, and F3a were difficult to obtain in high enough and pure enough concentrations for successful A-tailing and TA-cloning. When purifying the highly concentrated gel-cut PCR fragments resultant DNA would be at very low concentration, insufficient for downstream cloning. To combat this, warm nuclease-free water was used to elute DNA, instead of room temperature elution water. When tested with the NanoDrop and analyzed through AGE, DNA appeared highly concentrated and pure. However resulting clones continued to be negative for desired fragments. New competent DH5α and JM109 E. coli cells were prepared for testing with both cell types. Neither of two JM109 stocks proved to be successful in uptake of ligated DNA, and only a few colonies of the plasmid control were observed, suggesting loss of competency of these cells. Using fresh DH5α, proven to be competent according to tests with ligations and plasmid, pGEM FLC constructs were unsuccessful, resulting in only a few colonies. All resultant colonies for any attempt at transformation with these constructs were tested using cPCR and/or PCR from extracted plasmid, but no clear positives were given. The issue did not appear to be the competency of the
cells, so another look at the template DNA being subjected to A-tailing and TA-cloning by AGE was next. At this stage it was noted that the template DNA was floating out of the wells, indicating ethanol contamination. Spinning in a SpeedVac for 5-10 minutes on medium heat evaporated any ethanol in the template DNA, as confirmed by subsequent sinking into the gel wells. The increased temperature of the elution water is likely the cause of ethanol contamination, however it did increase concentration of eluted DNA, thus may still be necessary. This finding meant that the many previous attempts at A-tailing and ligation of these fragments were inhibited by ethanol contamination.

4.7.4 Overlap PCR conditions

The final challenge in the FLC work was finding appropriate overlap PCR conditions. The issue with ethanol contamination affected the overlap PCR reactions, as agarose gel imaging of template fragments appeared to be pure and highly concentrated. Prior to knowing of the ethanol contamination issue, an initial attempt at the F2 overlap PCR reaction was highly successful, but unfortunately the reactions could not be repeated. The successful F2 fragment was purified with warm elution water and cloning attempts were repeatedly unsuccessful, likely for the same ethanol contamination problem. However, this made it clear that the overlap reaction was indeed possible, and able to produce highly concentrated, large, pure DNA fragments. After determining the cause of inhibition of these reactions, all samples were subjected to SpeedVac heating to remove ethanol and were checked again by AGE. At this stage alterations to the overlap PCR conditions were tested. The addition of 3 min initial annealing and 7 min elongation steps were added to allow for full annealing of the megaprimer region and elongation by KOD polymerase. As cycling conditions for the F2 fragment were previously successful, the 54°C annealing temperature was maintained through the majority of F2 overlap
PCR attempts. Comparisons between the addition of flanking primers at the start of the PCR reaction (conventional) and adding primers after the initial denaturation, annealing and elongation steps showed no difference in producing weakly successful F2 bands. With the generous help of Sunny Li it was determined that a significant increase in primers in the overlap PCR reactions gave successful amplification. The addition of more primers meant that primers were not depleted by the amplification of non-specific bands, seen in most overlap reaction attempts. Primers were still available to amplify the large overlap fragments, which would take longer to synthesize. The FLC research in this thesis work ends with the successful amplification and TA-cloning of F2a, F2b, F3, and F4, encompassing the entire GRSPaV-SY genome, in addition to the pBS35S, pBS35SHDV and pBS35SNos (lacking the HDV sequence).

4.7.5 Failed ClaI digestion during F3-pBS35S cloning

The pBS35S-F3 ligation was believed to be successful based on the large number of transformants and success of the positive ligation control. ClaI digestion of clones that screened positive for the F3 overlap region gave the expected band of approximately 6.2 kB. The SalI digest gave two bands of approximately 3.2 and 3.0 kB, indicating the presence of an unexpected SalI site. During the design phase for the FLC, consideration was made for the presence of a SalI site in the pGEM-T-Easy MCS. The sequencing outcome for pGEM-F3 indicated its orientation was backward, as is common in TA cloning, however during sequence analysis, an unexpected KpnI site was found within the F3 fragment, which cuts at nt 6131 (Figure 19). The SalI digest outcome meant the unexpected SalI site was approximately flanking the F3 fragment, as this fragment is 3.2 kb. It was here the idea that the ClaI digest may not have been complete arose. SalI digestion of pGEM-F3 and pBS35S was carried out first, followed by AGE and purification of the digested products. The SalI digest products for pGEM-F3 gave an expected band at 3.2
kB, as well as 3.0 kB for the vector backbone. Purified F3 and pBS35S products were then digested with ClaI. The expected result was still a band of 3.2 kB for the F3 fragment, as only a 26 nt fragment of the pGEM MCS between the F3 ClaI site and the pGEM MCS SalI site. The error occurred during the purification step, since both the ClaI-SalI and SalI-digested F3 differ by only 26 nt and the difference in band size would be undetectable so that both products would end up in the purified product. A similar error occurred in the pBS35S digest products. The ClaI digest was incomplete, and ClaI-SalI digested pBS35S were purified with SalI-digested pBS35S products that only differ by 8 nt. During the ligation reactions, SalI-digested pBS35S ligated to SalI-digested F3 including a 26 nt region of the pGEM MCS (Figure 19). While the desired ClaI-SalI fragments in theory were present in the ligation mixture, all resultant clones that were confirmed F3-positive also contained the undesired 26 nt MCS region. A complete ClaI digestion is essential for subcloning of the F3 fragment. The purified F3 used in the ligation reaction was unknowingly composed of both SalI-ClaI-digested F3 and SalI-only F3 containing the unwanted 26 nt region. This sample should be subjected to complete ClaI digestion to remove the 26 nt region and provide ends compatible for subcloning with ClaI. It is also important to make sure there is no residual ethanol in the solution that would inhibit digestion or downstream cloning.

4.8 Future directions

4.8.1 CP diversity

The initial design of the 7770F primers was to produce a larger amplicon to be cloned and sequenced. Many past GRSPaV diversity studies use a relatively small target region for phylogeny. Small amplicons are much easier to clone, can be sequenced from only one direction, and make bioinformatic processing much simpler. The 864 nt amplicon could be shorted by
approximately 100 nt to allow for sequencing from one direction, saving a lot of time and money. Future attempts at repeating the CP diversity study should use an oligoT primer for RT from the viral polyA tail, as the CP target region is near the 3’ end of the genome and the oligo dT primer would not target ribosomal RNAs. Additionally, a higher RT incubation temperature should be used to avoid any non-specific binding of primers. New primers for this study need to be designed.

4.8.2 Steps for completing the GRSPaV-SY FLC

To complete the GRSPaV-SY FLC a number of straightforward steps need to be taken. As the complete sequence for the FLC has been obtained by sequencing each of the pGEM-T-Easy fragment constructs, mapping of any downstream subcloning obstacles is critical at this point. Any and all interfering mutations must be clearly mapped. It is currently known that there is an unexpected KpnI site (5’ GGTACC 3’) cutting at nt 3006 of the F3 sequence, 183 nt from the SalI cut site. SDM of this KpnI sequence to a match that of the corresponding GRSPaV-SY genome (5’ GGCACC 3’) should be carried out in the pGEM-F3 plasmid before cloning into the pBS vector construct. Restriction site mapping for the F2a, F3b, F3, and F4 fragments determined no other detrimental cut sites, like those required for downstream subcloning, were present in the rest of the fragment sequences.

The FLC construction first takes place in pBluescript II KS+. The 35S CaMV promoter has been cloned into this vector to create pBS-35S (Figure 20, Step 1). The F4 fragment must be cloned in next by digesting both pGEM-F4 and pBS-35S with SalI and KpnI (Figure 20, Step 2). The F4 fragment includes the 3’ end of the TGBp3 gene, the CP gene, and the 3’ NCR. Following the ligation of the F4 fragment to pBS-35S, the HDVnos sequence must be cloned in via KpnI. This sequence contains the HDV Rz sequence and the nos terminator sequence. KpnI
sites flank both ends of the HDVnos sequence, so sequence orientation must be carried out at this step (Figure 20, Step 3). The RSP-SY7581-F primer paired with the HDVnosR primer would be suitable here, producing a fragment of 1549 including F4 and the HDVnos sequence. After confirmation of the correct HDVnos sequence orientation, a mutation needs to be made to the KpnI site linking the 3’ end of the F4 fragment with the 5’ end of the HDVnos sequence (Figure 20, Step 4). The KpnI site should be mutated to all adenosine, as only the polyA tail follows this part of the genome. This will prevent removal of the KpnI-flanked HDVnos sequence during later subcloning. The F3 fragment must next be subcloned to pBS-35SF4Δ3KpnHDVnos construct via digestions with ClaI and SalI (Figure 20, Step 5). Setting up an overnight ligation with a relatively high concentration of F3 digest product is required as ligation of large fragments requires sufficient time for a high yield of products. Next, the full-length F2 fragment needs to be subcloned into pBS35S-HDVnosF3F4Δ3Kpn construct by digestions with XbaI and ClaI (Figure 20, Step 6). Again, due to the large fragment size of F2 (4.4 kB) and the target vector (8 kB), overnight ligations with highly-concentrated, ethanol-free digest products is essential for successful ligation. At this step, the non-viral sequence at the 5’ end of F2 must be deleted (Figure 20, Step 7). This short sequence of “TTTTTCTAGA” in the RSP-SY1F(Xba) primer includes the XbaI sequence and an additional thymine string to help RE binding and is no longer needed for downstream subcloning. At this step, the entire GRSPaV-SY genome should be ligated together with the 35S CaMV promoter and the HDVnos terminator sequence in place. Using NotI and KpnI the entire 35S-GRSPaVJH-HDVnos construct can be subcloned into the final target vector, pCB301.3 (Figure 20, Step 8) (Udaskin, 2015). Alternatively, if cloning the full FLC genome and flanking sequences via XbaI and KpnI digestions proves to be too difficult, the construct could be subcloned in two pieces. This is possible because the restriction sites
which have been taken advantage of thus far, particularly the ClaI site at nt 4403, are in the same XbaI-ClaI-KpnI sequence in the target vector, pCB301.3. Figure 20 shows a diagram of this cloning strategy.

4.8.3 Potential future experiments involving the GRSPaV-SY FLC

Upon completion of the infectious clone, infiltrations of *N. benthamiana* and grapevine tissue culture will be carried out to confirm infectivity and use in a wide range of downstream studies. The FLC lacking the HDV ribozyme sequence at the 5’ end of the *nos* terminator was intended for comparison to the FLC construct including the HDV ribozyme sequence. In the study by Vives *et al.* (2008) two pBIN19-SfiI L2.7Afl II modified binary vector-based FLCs of CLBV under the CaMV 35S promoter, differing in the presence or absence of the HDV ribozyme sequence were constructed for comparison of infectivity in both *N. benthamiana* and *N. occidentalis*. Their findings show both constructs could establish systemic infection however *N. benthamiana* had much greater CLBV RNA compared to *N. occidentalis* at the same time point. Their study also concluded that the construct with the ribozyme sequence gave a stronger Northern blot hybridization signal than the construct lacking this sequence. The ribozyme antigenomic sequence cleaves viral transcript RNA ends to more closely resemble the 3’ end of RNA viruses. Both CLBV FLC constructs were tested in the presence of the p19 suppressor in order to suppress host defence mechanisms.

So far, the GRSPaV-GG FLC pRSP28-2(Cam), has not been proven to move systemically within *N. benthamiana*, a non-native host (Meng *et al.*, 2009; our unpublished data). While it seems unlikely based on the previous statement, movement within a model plant by the GRSPaV-SY infectious clone would provide new information about GRSPaV movement. The
1. pBluescript MCS with CaMV 35S promoter previously cloned in via NotI and KpnI

2. Cloning F4 via SalI and KpnI

3. HDVnos terminator sequence cloned in via KpnI. Orientation confirmation must be done to confirm the sequence is in the correct orientation

4. SDM of F4 5’ KpnI site to polyA sequence

5. Cloning F3 via Clal and SalI, after the mutagenesis of the unexpected F3 KpnI site

6. Cloning of F2 via XbaI and ClaI

7. Deletion of 5’ non-viral sequence via SDM

8. Cloning of entire genome with promoter and terminator sequences into pCB301.3 via NotI and KpnI

**Figure 20:** FLC cloning strategy. Cloning F4 into pBS-35S is done before the cloning of the HDVnos terminator sequences to avoid downstream removal of the HDVRz and nos sequences which it includes. SDM of the KpnI site between F4 and the HDVnos sequence is critical. F3 is cloned next. F2 is cloned in last, followed by SDM of 5’ non-viral sequences. The final target binary vector, pCB301.3, has the same MCS sequence as pBluescript and is not shown. The 35S CaMV promoter is represented by ‘P’, the HDVRz and nos terminator sequences are represented by “HDV”. Left and right boarder sequences in the binary vector are represented by LB and RB (grey), respectively.
GRSPaV-SY FLC provides a complete genome sequence from which genetic regions can be recombined with regions of the GRSPaV-GG FLC. An example study could involve replacing the CP of the GRSPaV-GG FLC with the CP of the GRSPaV-SY FLC. Additionally, a grapevine tissue culture system is currently being developed in our lab. This will be used for downstream work from this thesis to infect virus-free grapevine tissue with the GRSPaV-SY infectious clone. Many experiments can be designed from these basic grapevine tissue infiltration methods. For example, grapevine viral response gene activation/deactivation could be determined by observing changes in grapevine gene expression via real-time PCR at increasing time points post infection. The FLC will also be useful in studies involving replacing GRSPaV genes with genes from another virus, such as the TGB movement proteins with the p30-like movement protein. Further, the FLC could be tagged with a fluorescent protein and observed when coinfectected with fluorescently-tagged clones of other foveaviruses and/or pRSP-GFP2(Cam), a GRSPaV-GG FLC. Most importantly, the FLC of a GRSPaV-SY variant could be used to fulfil the role that GRSPaV-SY plays in SD.

4.9 Concluding remarks

It is evident that GRSPaV is widespread in Ontarian grapevines and at least five distinct lineages are present. Syrah with red canopy had a much higher instance of GRSPaV-SY variants than Syrah without red canopy, however environmental and biological factors must be considered. More evidence is required to make a clear relationship between SD and GRSPaV. A FLC of a GRSPaV-SY variant is under construction, with the full genome currently sequenced and cloned into pGEM-T-Easy vectors. Two subgroups of the –PN lineage are highly prevalent, however the GRSPaV-PN reference sequence may be a chimera sequence of human error. Only three groups were detected in hybrid varieties; GRSPaV-SY, -BS, and –VF1. The first
distribution of GRSPaV in hybrid grapevine varieties indicates that GRSPaV-SY is highly prevalent in hybrids, followed distantly by GRSPaV-BS.

For the future it will be very important for grape growers to monitor their vines for SD symptoms and to be careful when it comes to replanting. The suboptimal grapevine growing conditions are likely a strong contributing factor to the decline of Syrah in Ontario. Ontarian grapevines endure harsh winters which severely stress the plant and cold survival mechanisms become priority. With weakened vines resultant of cold winters, host factors to suppress viral activity are likely less readily synthesized. It may be advisable to discontinue growing Syrah in Ontario if SD symptoms do not subside. Finding grapevine cultivars which better adapt to cold and changing climates may need to become the Ontarian signature wine varieties.

Clean plant programs are an incredibly valuable tool for agriculturists. Commercially available crops are bred from grapevines rigorously tested for grapevine pathogens. The National Clean Plant Program provides clean grapevine buds and rootstocks to growers. Annual testing for over 30 grapevine pathogen is done in both the laboratory and the field and therapies are used to eliminate viruses using microshoot tissue culture methods (NCPN, http://ncpngrapes.org). Cost benefits of starting a vineyard with clean vines in regions like Northern California are estimated at $50 million per year. Regions such as the Finger Lakes of New York, greatly affected by grapevine viral diseases, is estimated to raise net returns by $9,000-16,000 per acre over a 25 year vineyard lifespan (NCPN, http://ncpngrapes.org). When vine health is optimal, the berries produced will follow, and replanting as a result of vine disease will be few and far between.
REFERENCES


RT-PCR based detection of Rupestris stem pitting associated virus within field-grown grapevines throughout the year. Plant Dis., 85, 617-620.


**APPENDIX I: Primers**

**Table 6:** Primer table. GRSPaV target position and accession numbers for each primer is listed unless specified for another organism. For all primers, F indicates a forward primer, R indicates a reverse primer.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5’ → 3’</th>
<th>GRSPaV target position</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSP35</td>
<td>AGRYTTAGRGTGCTAARGC</td>
<td>5705-5724</td>
<td>Target RdRp F, conserved across four lineages</td>
</tr>
<tr>
<td>RSP35-SY</td>
<td>AGGTTAAGGTTAGCAAGG</td>
<td>5705-5724</td>
<td>RSP35 primer specific to GRSPaV-SY AY368590</td>
</tr>
<tr>
<td>RSP36</td>
<td>CACATRTCATVCYGGCAAA</td>
<td>6180-6161</td>
<td>Target RdRp R, conserved across four lineages</td>
</tr>
<tr>
<td>RSP13</td>
<td>GATGAGGTCCAGTTGTTTCC</td>
<td>4373-4392</td>
<td>Broad-spectrum GRSPaV F primer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Conserved among 9 clones</td>
</tr>
<tr>
<td>SY1659F</td>
<td>TAAGATGGCCTTGGGTGTTG</td>
<td>1659-1679</td>
<td>F primer targeting HVR of GRSPaV-SY AY368590</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>F primer for F2b fragment</td>
</tr>
<tr>
<td>SY2127R</td>
<td>ATTTATGGGATGGGCACATG</td>
<td>2127-2107</td>
<td>R primer targeting HVR of GRSPaV-SY AY368590</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R primer for F2a fragment</td>
</tr>
<tr>
<td>PN1701F</td>
<td>CTTCTTGGTGAAACAGCACC</td>
<td>1701-1720</td>
<td>F primer targeting HVR of GRSPaV-PN AY368172</td>
</tr>
<tr>
<td>PN2187R</td>
<td>AACAAATTGCCTCACAAGCC</td>
<td>2187-2167</td>
<td>R primer targeting HVR of GRSPaV-PN AY368172</td>
</tr>
<tr>
<td>BS1672F</td>
<td>AAAGCGACCGGACGTCTCAG</td>
<td>1672-1692</td>
<td>F primer targeting HVR of GRSPaV-BS AY881627</td>
</tr>
<tr>
<td>BS2109R</td>
<td>CGAAATCCCTCAGGGCTT</td>
<td>2109-2089</td>
<td>R primer targeting HVR of GRSPaV-BS AY881627</td>
</tr>
<tr>
<td>NotI35S-F</td>
<td>TTTTGGGCGCCCGGGCAAGAAAGCCCTCCTG</td>
<td>1-19</td>
<td>F primer for 35S CaMV promoter in pHST40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Introduces a 5’ NotI site</td>
</tr>
<tr>
<td>Xba35S-R</td>
<td>TTTTCTAGACCTCTCCAAGTGAAATGAGAAGGACTTCC</td>
<td>318-294</td>
<td>R primer for 35S CaMV promoter in pHST40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Introduces a 3’ XbaI site</td>
</tr>
<tr>
<td>Primer Code</td>
<td>Sequence</td>
<td>Length</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>KpnNos-F</td>
<td>TTTTGGTACCGAGCTGAAT TCCCCGATCG</td>
<td>1-21</td>
<td>F primer for nos terminator lacking 5’ HDV sequence</td>
</tr>
<tr>
<td></td>
<td>TTTTGGTACCGGTCCGGATGTATCAG</td>
<td>1-15</td>
<td>F primer for HDVnos terminator tandem sequences in pHST40</td>
</tr>
<tr>
<td>KpnNos-R</td>
<td>TTTTGGTACCGCGATCTAG TAACATAGATGACACC</td>
<td>354-328 for nos only</td>
<td></td>
</tr>
<tr>
<td>R primer for nos and HDVnos sequences in pHST40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TTTTCTAGAGATAAACATA ACAACAGAAATGCA</td>
<td>1-25</td>
<td>F primer for F1, F2a and F2</td>
</tr>
<tr>
<td>RSP-SY1F(Xba)</td>
<td>TTTTCTAGAGATAAACATA ACAACAGAAATGCA</td>
<td>1-25</td>
<td>Introduces a 3’ XbaI site</td>
</tr>
<tr>
<td></td>
<td>GTTACTAGTAAAGATGTAACG GGTATTAGCAG</td>
<td>433-462</td>
<td>F primer for FLC fragment 433-4410, originally planned to be F2</td>
</tr>
<tr>
<td>RSP-SY433F(Spe)</td>
<td>CTTCATAGTAAACAATCTA TTGAGTACAGA</td>
<td>445-415</td>
<td>R primer for F1</td>
</tr>
<tr>
<td>RSP-SY445R(Spe)</td>
<td>GAAATCCTTGCTTCACTGGA TGG</td>
<td>2414-2437</td>
<td>Used as F primer for initial F2b fragment</td>
</tr>
<tr>
<td>RSP-SY2414F</td>
<td>CTACATCGATTATGTCTTCTTGTTCATT</td>
<td>4399-4415</td>
<td>F primer for F3 and F3a</td>
</tr>
<tr>
<td>RSP-SY4399F(Cla)</td>
<td>TACATCGATTATGTCTTCTTTGTTCATT</td>
<td>4399-4415</td>
<td>Natural ClaI site in GRSPaV-SY genome (nt 4410)</td>
</tr>
<tr>
<td>RSP-SY4410R(Cla)</td>
<td>TAAATCGATGTAACCAGGG GG</td>
<td>4410-4389</td>
<td>R primer for F2 and F2b</td>
</tr>
<tr>
<td>RSP-SY7581F(Sal)</td>
<td>GCTGTCGACATGTGCAAGG</td>
<td>7581-7591</td>
<td>F primer for F4</td>
</tr>
<tr>
<td>Primer Name</td>
<td>Sequence</td>
<td>Location</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>RSP-SY5938F</td>
<td>CACTTCATCTTGTCATTTGA GTTGCG</td>
<td>5938-5963</td>
<td>F primer for F3b fragment</td>
</tr>
<tr>
<td>RSP-SY6074R</td>
<td>TCACCCGTAATCTCATTAT GGC</td>
<td>6074-6051</td>
<td>R primer for F3a fragment</td>
</tr>
<tr>
<td>RSP-SY7592R(Sal)</td>
<td>GCTGTCGACATGTGCAAGG</td>
<td>7592-7573</td>
<td>R primer for F3 and F3b</td>
</tr>
<tr>
<td>RSP-SY8725R(Kpn)</td>
<td>GGTACCTTTTTTTTTTTTTTTTTT TTGCGCGAAAAC</td>
<td>8725-8713</td>
<td>R primer for F4</td>
</tr>
</tbody>
</table>

Introduces KpnI site at 3'end of GRSPaV genome, to be mutated later.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
<th>Location</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSP7770F.BS</td>
<td>ATGGCAAGTCAAGTTGGAA AGTTTG</td>
<td>7770-7794</td>
<td>5' GRSPaV-BS CP F primer</td>
</tr>
<tr>
<td>RSP7770F.GG</td>
<td>ATGGCAAGTCAAAATTGGGA AACT</td>
<td>7770-7793</td>
<td>5' GRSPaV-GG CP F primer</td>
</tr>
<tr>
<td>RSP7770F.MG</td>
<td>ATGGCGAGTCAAGTTGGTAA GCT</td>
<td>7770-7793</td>
<td>5' GRSPaV-MG-CP F primer</td>
</tr>
<tr>
<td>RSP7770F.PN</td>
<td>ATGGCGAGCCAAATTGGGA AACT</td>
<td>7770-7789</td>
<td>5' GRSPaV-PN-CP F primer</td>
</tr>
<tr>
<td>RSP7770F.SG</td>
<td>ATGGCAAGTCAGATTGGTAA GCTC</td>
<td>7770-7794</td>
<td>5' GRSPaV-SG1-CP F primer</td>
</tr>
<tr>
<td>RSP7770F-SY</td>
<td>ATGGGAAGCCCACCAGGA</td>
<td>7770-7788</td>
<td>5'GRSPaV-SY-CP F primer</td>
</tr>
<tr>
<td>RSP8634R</td>
<td>GTACGGGTATTCAGCGAACA GGC</td>
<td>8634-8611</td>
<td>Broad-spectrum GRSPaV R primer</td>
</tr>
</tbody>
</table>

Targets highly conserved region of 3’ UTR.
**Figure 21**: Alignment map of GRSPaV-SY FLC fragments and internal primers with the GRSPaV-SY AY368590 (top arrow) reference sequence. Sequenced regions of fragments are represented by dashed (M13R) or solid black (M13F) arrows. Newly designed primers for sequencing of internal regions are depicted by small triangles. In descending order: GRSPaV-SY AY368590, F2a M13F (solid arrow x2), primer RSP-JH681F primer, F2a M13R (dashed arrow x2), F2b M13F (solid arrow), F2b M13R (dashed arrow), F2b M13F (solid arrow, longer sequence), RSP-JH2326F primer, F3a M13F (black arrow), F2b M13R (dashed arrow, longer sequence), F3a M13R (dashed arrow), RSP-JH3712R primer, F3b M13F (solid arrow), RSP-JH4997F primer, F3b M13R (dashed arrow), F3a M13R (dashed arrow, longer sequence), F4 M13R (solid arrow, sequenced in reverse orientation), F3 M13R (dashed arrow), RSP-JH6454F primer, F4 M13R (solid arrow, longer sequence), F4 M13F (dashed arrow, longer sequence).
Figure 22: Primer map including primers used for the GRSPaV-SY FLC and some commonly used primers for GRSPaV diversity analysis. Grey dashed lines represent positions and primers used in the FLC constructs. Yellow dashed lines represent positions and primers not used in the FLC construction but commonly used or discussed in this thesis.
APPENDIX II: GRSPaV diversity: Supplemental phylogenetic trees

Figure 23 (page 127): Distribution of RSP35-RSP36 clones from all *Vitis* sources. Phylogenetic analysis was conducted using maximum likelihood with 1000 bootstrap replicates. The bar indicates 0.01 nucleotide substitutions/site, or genetic distance over time. Numbers in brackets represent the number of identical sequence clones obtained from the source. Reference sequences for the RSP35-RSP36 region of 15 sequence isolates obtained from GenBank are included, each prefixed with “RSP35-36 GRSPaV-”. The corresponding region from ASPV is included as an outgroup. Each group is represented by coloured branches; -SY, red; -PN, dark blue; -VF1, medium blue; -JF, cyan; -GG, magenta; -SG1, green; - BS, yellow.
Figure 24 (page 129): Distribution of RSP35-SY-RSP36 clones from 14 *Vitis* sources. Phylogenetic analysis was conducted using maximum likelihood with 1000 bootstrap replicates. The bar indicates 0.1 nucleotide substitutions/site, or genetic distance over time. As expected the vast majority of clones clustered with GRSPaV-SY, however two clones clustering with the –PN subgroups were found using these SY-specific RSP35-RSP36 primers. Numbers in brackets represent the number of identical sequence clones obtained from the source. Reference sequences for the RSP35-RSP36 region of 15 sequence isolates obtained from GenBank are included, each prefixed with “RSP35-36 GRSPaV-”. The corresponding region from ASPV is included as an outgroup. Each group is represented by coloured branches; -SY, red; -PN, dark blue; -VF1, medium blue; -JF, cyan; -GG, magenta; -SG1, green; - BS, yellow.
Figure 25: Distribution of GRSPaV sequence variants in Syrah with red canopy. The analyzed sequence was that amplified with primers RSP35-RSP36. Phylogenetic analysis was conducted using maximum likelihood with 1000 bootstrap replicates. The bar indicates 0.01 nucleotide substitutions/site, or genetic distance, over time. Clones representing all groups were found in this data set; GRSPaV-SY, -BS, -VF1, -JF, -SG1, and -GG. Reference sequences for the RSP35-RSP36 region of the 15 sequence isolates obtained from GenBank are included, each prefixed with “RSP35-36 GRSPaV-“. The corresponding region from ASPV is included as an outgroup. Each group is represented by coloured branches; -SY, red; -PN, dark blue; -VF1, medium blue; -JF, cyan; -GG, magenta; -SG1, green; - BS, yellow.
Figure 26: GRSPaV sequence variant distribution in Syrah without red canopy. The analyzed sequence was that what amplified with primers RSP35-RSP36. Phylogenetic analysis was conducted using maximum likelihood and 1000 bootstrap replicates. The bar indicates 0.01 nucleotide substitutions/site, or genetic distance, over time. Clones representing most groups were found in this study; GRSPaV-SY, -SG1, -VF1, -JF, and -GG. Reference sequences for the RSP35-RSP36 region are included for 15 sequence isolates obtained from GenBank, each prefixed with “RSP35-36 GRSPaV-“. The corresponding region from ASPV is included as an outgroup. Each group is represented by coloured branches; -SY, red; -PN, dark blue; -VF1, medium blue; -JF, cyan; -GG, magenta; -SG1, green; - BS, yellow.
Figure 27: GRSPaV sequence variant distribution in hybrid varieties. The analyzed sequence was that what amplified with primers RSP35-RSP36. Phylogenetic analysis was conducted using maximum likelihood and 1000 bootstrap replicates. The bar indicates 0.01 nucleotide substitutions/site, or genetic distance, over time. Clones representing only GRSPaV-SY, -BS, and –VF1 subgroups were found in this data set. Reference sequences for the RSP35-RSP36 region are included for 15 sequence isolates obtained from GenBank, each prefixed with “RSP35-36 GRSPaV-“. The corresponding region from ASPV is included as an outgroup. Each group is represented by coloured branches; -SY, red; -PN, dark blue; -VF1, medium blue; -JF, cyan; -GG, magenta; -SG1, green; - BS, yellow.
Figure 28: RSP35-RSP36 clones from hybrid varieties before removal of unfit sequences shows highly similar distribution to the refined data trees (Figure 17, Figure 27). Alignment was done using MegAlign (DNAStar) ClustalW alignment viewed as a phylogenetic tree, and does not reflect true phylogenetic analysis. The bar across the bottom represents the number of nucleotide substitutions per 100 residues. The GRSpaV-SY lineage is represented by ‘SY’, the -PN lineage is represented by ‘VF1’, the -BS lineage is represented by ‘BS’, and the –SG1 lineage is represented by ‘PG’. The numbers in brackets represent the total number of clones within the respective cluster. Reference sequences for the RSP35-RSP36 region are included for 15 sequence isolates obtained from GenBank, each prefixed with “RSP35-36 GRSpaV-”. The corresponding region of ASPV was included as an outgroup.
APPENDIX III: Nucleotide alignment of FLC fragments

Figure 29: 35S CaMV sequence in the PBS-35S construct aligned with the 35S CaMV sequence in pHST40. The alignment was carried out using MegAlign (DNASTar). Nucleotides differing between the two sequences are indicated by black shading.
Figure 30: Sequence alignment of the F2a sequence in pGEM-F2a and nucleotides 1-2127 of the GRSPaV-SY AY368590 sequence. The alignment was carried out using MegAlign (DNASTar). Nucleotides differing between the two sequences are indicated by black shading.
Figure 31 (page 137): Sequence alignment of the F2b sequence in pGEM-F2b and nucleotides 1659–4410 of the GRSPaV-SY AY368590 sequence. The alignment was carried out using MegAlign (DNAStar). Nucleotides differing between the two sequences are indicated by black shading.
Figure 32 (pages 138-139): Sequence alignment of the F3 sequence in pGEM-F3 and nucleotides 4399-7592 of the GRSpaV-SY AY368590 sequence. The alignment was carried out using MegAlign (DNAStar). Nucleotides differing between the two sequences are indicated by black shading.
Figure 33: Sequence alignment of the F4 sequence in pGEM-F4 and nucleotides 7581-8742 of the GRSpaV-SY AY368590 sequence (includes the polyA sequence). The alignment was carried out using MegAlign (DNASTar). Nucleotides differing between the two sequences are indicated by black shading.
**Figure 34**: Sequence alignment of the HDVnos sequence in pBS with the HDVnos sequence in pHST40. The alignment was carried out using MegAlign (DNAStar). Nucleotides differing between the two sequences are indicated by black shading, here, the restriction enzyme cut sites.
Figure 35: ORF1 alignment of the GRSPaV-JH (query) predicted aa sequence with ORF1 of GRSPaV-SY (subject), the consensus is between the two.
24.4 kDa hypothetical protein [Grapevine rupestris stem pitting-associated virus]

Sequence ID: BAF34850.1  Length: 221  Number of Matches: 1

Range 1: 2 to 221  GenPept  Graphics

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect Method</th>
<th>Identities</th>
<th>Positives</th>
<th>Gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>429 bits(1103) 8e-152 Compositional matrix adjust.</td>
<td>206/220(94%)</td>
<td>214/220(97%)</td>
<td>0/220(0%)</td>
<td></td>
</tr>
</tbody>
</table>

Query 1  NNLVKALSAFEFVGVYSVLRFPVVHSVPGSGKSSLIRELETEDFTASFATASVPDTPLN NNLVKALS FEFVG+YSVLRFPVVHSVPGSGKSSLIREL+TEDESF AFTASVPDTPLN
Sbjct 2  NNLVKALSTFEFFVGIYSVLRFPVVHSVPGSGKSSLIRELTEFIAFTASVPDTPLN

Query 61  SGRYIKPYSPCGAVGRINILDELYSVDDISGFDVLFSHPYQNPISVPREAHFKSITCRF SGRYIKPYSPCGAV-G+INILDELYSV DISGFDVLFSHPYQNPISVPREAHFKSITCRF
Sbjct 62  SGRYIKPYSPCGAVGKINILDELYSVSDISGFDVLFSHPYQNPISVPREAHFKSITCRF

Query 121  GENTCKYLLSFGNFNHSDGLDEVI08GSPFELDVEGVLICFGKGEAVDLAVAHDHNSFKLPCE
gentckyllsfgfn-+hsd  devi08gspfeldvegvlificfgkgeavdla ah+s+fklpce
Sbjct 122  GENTCKYLLSFGFNLSKDRSDEVI08GSPFELDVEGVLICFGKGEAVDRAAHNSDFKLPCE

Query 181  VRGATFDFVYTLKSRDPTS-DKHNFYIAAARTHRKLIIMQ  VRGATFDFVYTLKSRDPTS-DKHNFYIAAARTHRKLIIMQ
Sbjct 182  VRGATFDFVYTLKSRDPTS-DKHNFYIAAARTHRKLIIMQ

**Figure 36:** ORF2 alignment of the GRSPaV-JH (query) predicted aa sequence with ORF2 of GRSPaV-SY (subject), the consensus is between the two.
<table>
<thead>
<tr>
<th>Range 1: 2 to 117</th>
<th>Score</th>
<th>Expect</th>
<th>Method</th>
<th>Identities</th>
<th>Positives</th>
<th>Gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>239 bits(609)</td>
<td>6e-80</td>
<td>Compositional matrix adjust.</td>
<td>115/116(99%)</td>
<td>115/116(99%)</td>
<td>0/116(0%)</td>
</tr>
<tr>
<td>Query 1</td>
<td>PFQPANHAKTVPTLTLIGGVHLRKLSPWLPYSGDNIHOFPGHRGGRDRDGKSITYCG</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sbjct 2</td>
<td>PFQPANHAKTVPTLTLIGGVHLRKLSPWLPYSGDNIHOFPGHRGGRDRDGKSITYCG</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Query 61</td>
<td>PRQSFPSGGLFGQSENFPLIIIVLTLASIVLSVWNSGSRRCNPNPCTCQQ</td>
<td>116</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sbjct 62</td>
<td>PRQSFPSGGLFGQSENFPLIIIVLTLASIVLSVWNSGSRRCNPNPCTCQQ</td>
<td>117</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 37:** ORF3 alignment of the GRSPaV-JH (query) predicted aa sequence with ORF3 of GRSPaV-SY (subject), the consensus is between the two.
**Figure 38:** ORF4 alignment of the GRSPaV-JH (query) predicted aa sequence with ORF4 of GRSPaV-SY (subject), the consensus is between the two.
coat protein [Grapevine rupestris stem pitting-associated virus]

Sequence ID: ADA79765.1  Length: 259  Number of Matches: 1

Figure 39: ORF5 alignment of the GRSPaV-JH (query) predicted aa sequence with ORF5 of GRSPaV-SY (subject), the consensus is between the two.